云图投影到一个面上。如图 10.11 是旋转后的立方体，选中投影到 z-y 平面。

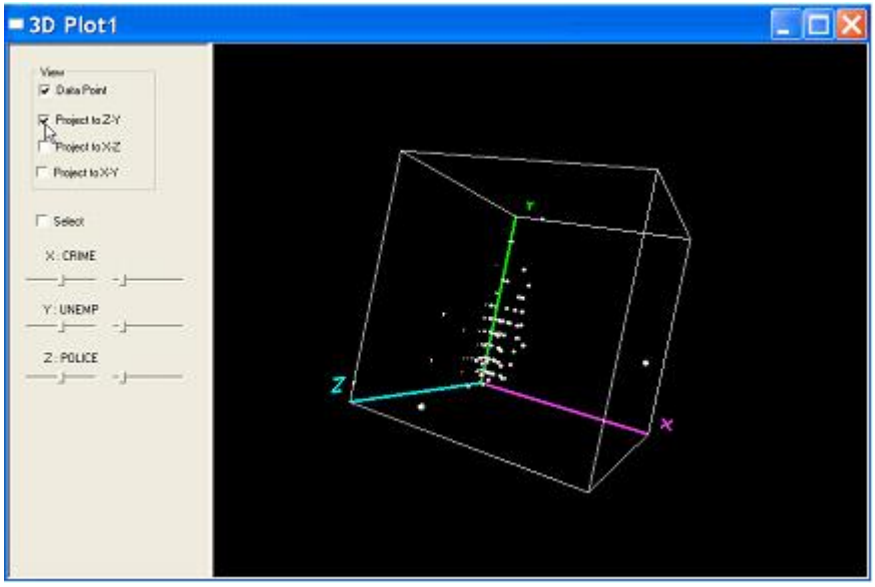


图 10.11 投放后的 3D 散点图及投影到 zy 平面上的 2D 图

左边面板中底部的选项定义了选择形状和控制刷。选中 Select 选项，图中将会出现一个红色立方体。移动变量名右边和下边的滑块，沿这一轴改变选择框的大小。如图 10.12，选择框会随着滑块向右移动沿 X 维(CRIME)扩大。操作每一个变量的滑块，直到选择框有一个大的范围。可以旋转立方体，更好的体会你的选择框在 3D 空间的位置。

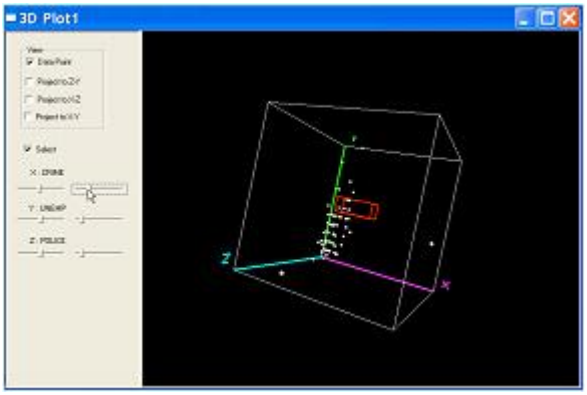


图 10.12 设置 3D 图形状选项

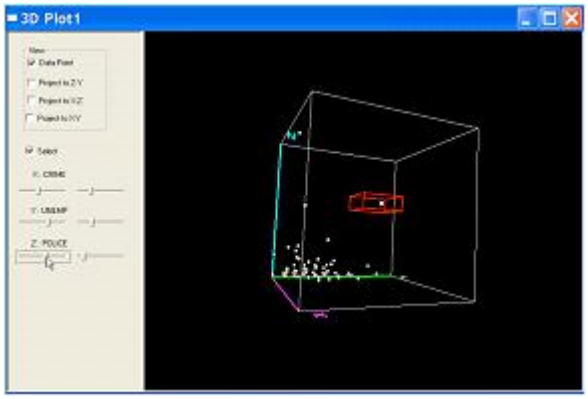


图 10.13 在 3D 图中移动选择形状

试着变化选择框的形状，移动选择框。你会发现常旋转立方体是很有帮助的，这样你可以看出选择框与点云图的关系。同样，你也可以按住 Control 键同时点击鼠标左键，来直接移动选择框。

3D 散点图与所有其它地图和图表是链接的。然而，进行选择的更新与 2 维情况下稍有不同。与标准情况相比，更新是连续的，3D 图中的选择在每次鼠标停止移动时更新。云图中的黄色点将在其它图表中匹配相应的观测值。如图 10.14，Mississipi 县图的云图中被选中的点被高亮显示。在其它方面，刷光功能与 2 维情况是相似的，但需要一些练习来了解选择框在 3 维空间的位置。

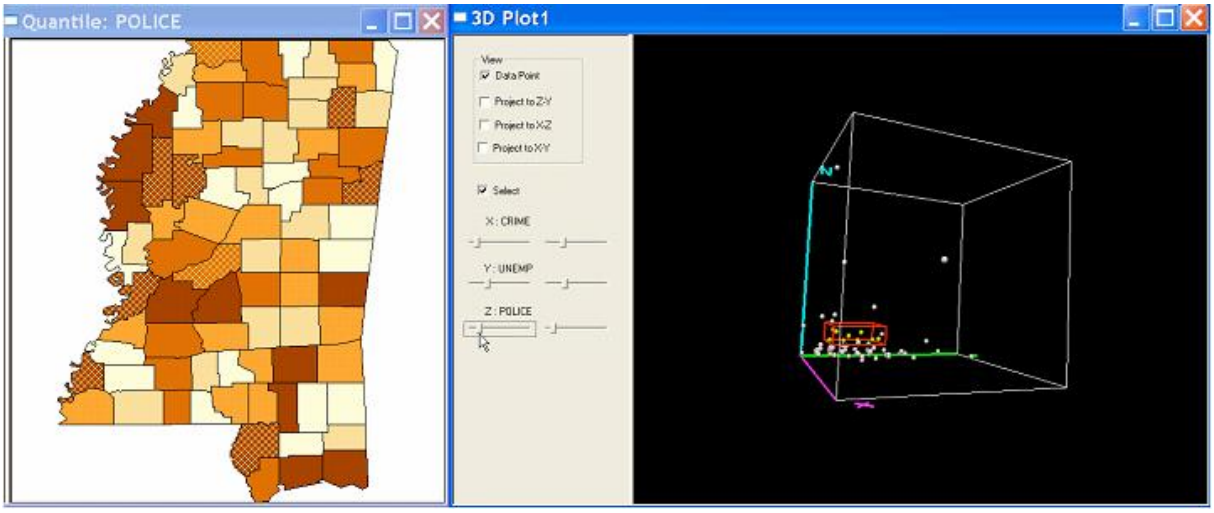


图 10.14 在 3D 图中刷光

在其它方向刷光也是有效的，但只有在 Select 选项关闭的情况下才能用。为了认识这一点，在县图中创建一个刷子，开始移动。每当刷子停止，3D 图中相匹配的选择将显示为黄色，如图 10.15。在练习中，你会发现经常缩放和旋转立方体是很有帮助的。

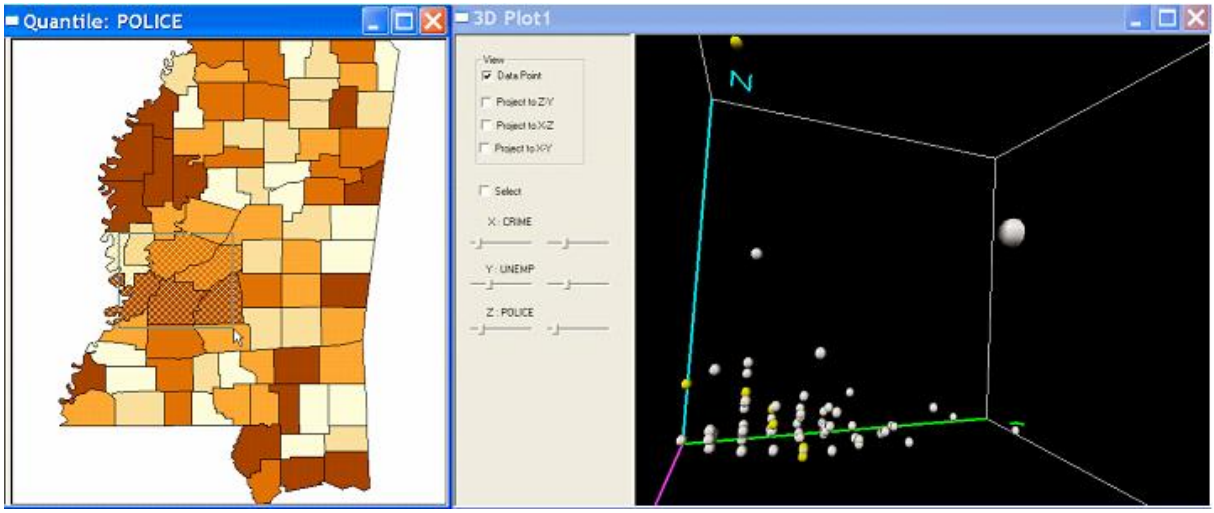


图 10.15 刷光与 3D 散点图链接的地图

10.4 练习

在 BOSTON 数据集(boston.shp，关键字为 ID)中，应用条件图和 3D 散点图探索中等房屋价值(CMEDV)和其它变量之间的关系。例如，考虑在条件散点图中由地理位置(X 和 Y)决定的房屋价格与空气质量(NOX)的关系。同样在 3D 散点图中探索房屋价格、空气质量和犯罪(CRIME)之间的联系。试着刷光由地区质心创建的 Thiessen 多边形地图。

现在你应该认识到(at a point：在一点之处的连续性)在什么情况下，将不同的传统 EDA 工具与探索即有空间又有非空间模式的地图结合，而大多数据即是有空间模式，也有非空间模式。

练习 11　ESDA 基础与地理可视化

11.1 目标

本练习开始解决具有明显空间特征的数据的探索的问题。此时(at this point)我们主要集中于制图和地理可视化。更高级的方法在练习 12 中论述。

在本练习末，你将知道：

- 创建一幅百分位图
- 创建一幅箱地图
- 在箱地图产中改变 Hinge 选项
- 创建一幅统计地图
- 改变统计地图的 Hinge 选项

有关这些操作的更详细的信息请参见 Users Guide 和 Release Notes。

11.2 百分位图

我们将会使用 BUENOSAIRES 数据集演示基本制图功能，包含了阿根廷议会在 Buenos Aires(Argentine)市 209 个地区 1999 年国家选举结果[1]。

文件名为 buenosaires.shp，关键字为 INDRANO。用此文件打开一个新的项目，底图如图 11.1。

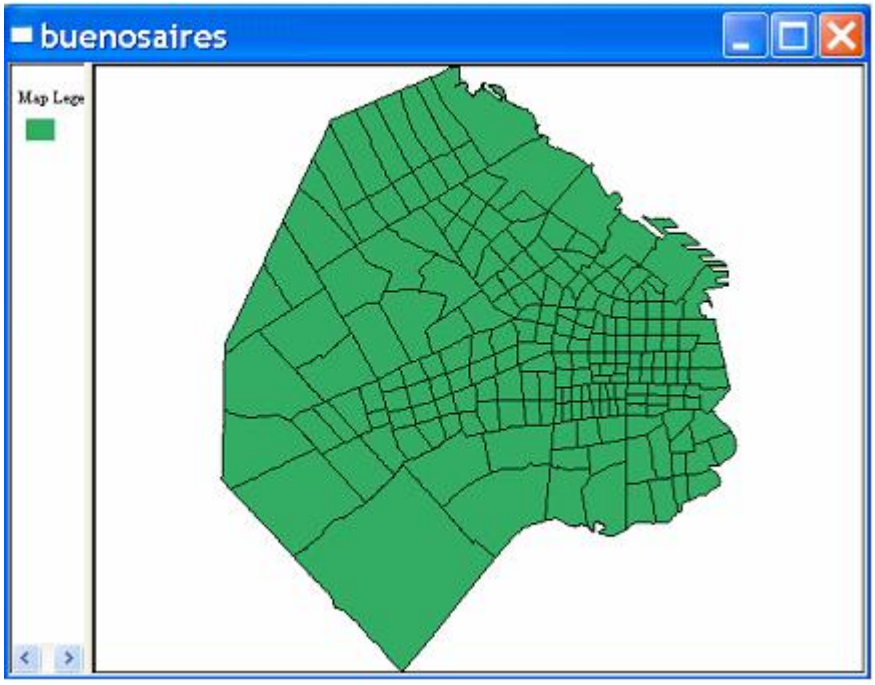


图 11.1 Buenos Aires 选举数据底图

从菜单中选择 Map>Percentile 或在底图中右击调用百分位图功能。后一方法将产生如图 11.2

[1] 有关数据的更详细的谈论及真实背景可参考 Calvo and Escobar(2003)。

的菜单。选择 Choropleth Map>Percentile 调用变量设置对话框。或点击工具栏按钮。

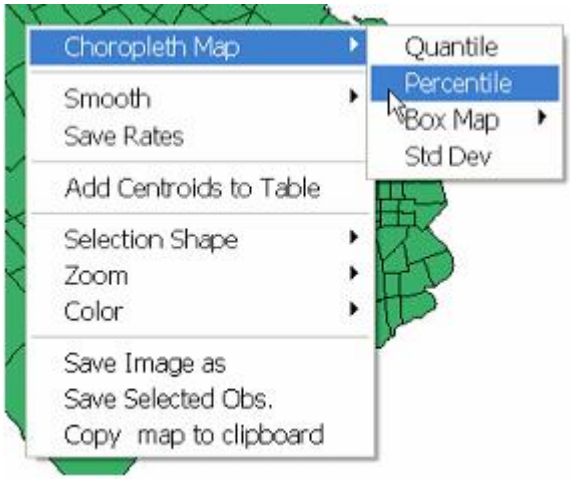


图 11.2 百分位图功能

选择变量 APR99PC(中右党 APR 选举结果，“Action por la Republica”)，如图 11.3，点击 OK 即产生一幅地图。图 11.4 所示的百分位图强调了非常小值(最低百分位)和非常高值(最高百分位)的重要性。注意该党的三个最高得票率(returns)是如何集中于三个小(面积)的选区(红色区域)。也注意简要分类大大简化了地图的空间模式。试着为其它党(AL99PC，中间党派“Alianza”)建一百分位图，及选举结果(TURN99PC)。关注一下这些地图所表示出的一般模式。

11.3 箱地图

箱地图是分位图的加强版本，其中第一和第四分位的离群值可以单独高亮显示。箱地图中的分类与箱图中所用的是完全相同的。从 Map 菜单或在地图中右击(图 11.5)调用箱地图功能。在当前地图中右击(或先创建一幅复制底图)，选择 Choropleth Map>Box Map>Hinge=1.5 产生变量选择对话框(见图 11.3)。仍然选择 APR99PC，点击 OK，产生一个箱地图，如图 11.6 所示。

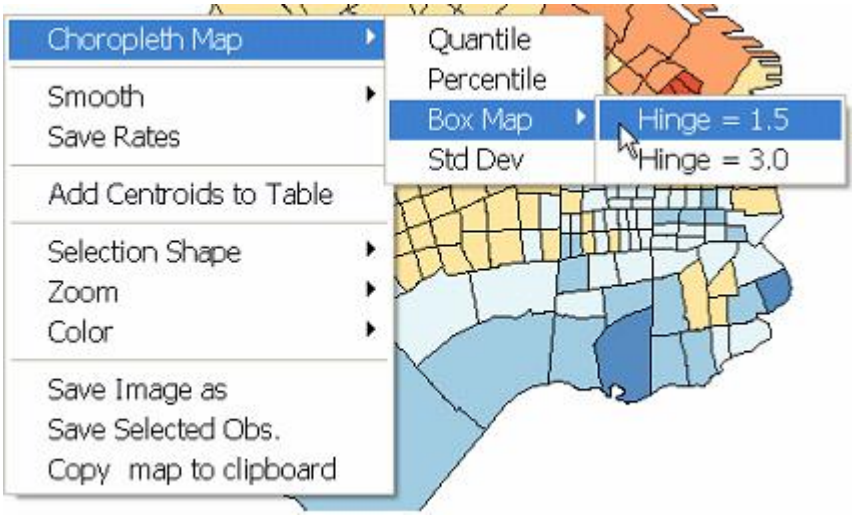


图 11.5 箱地图功能

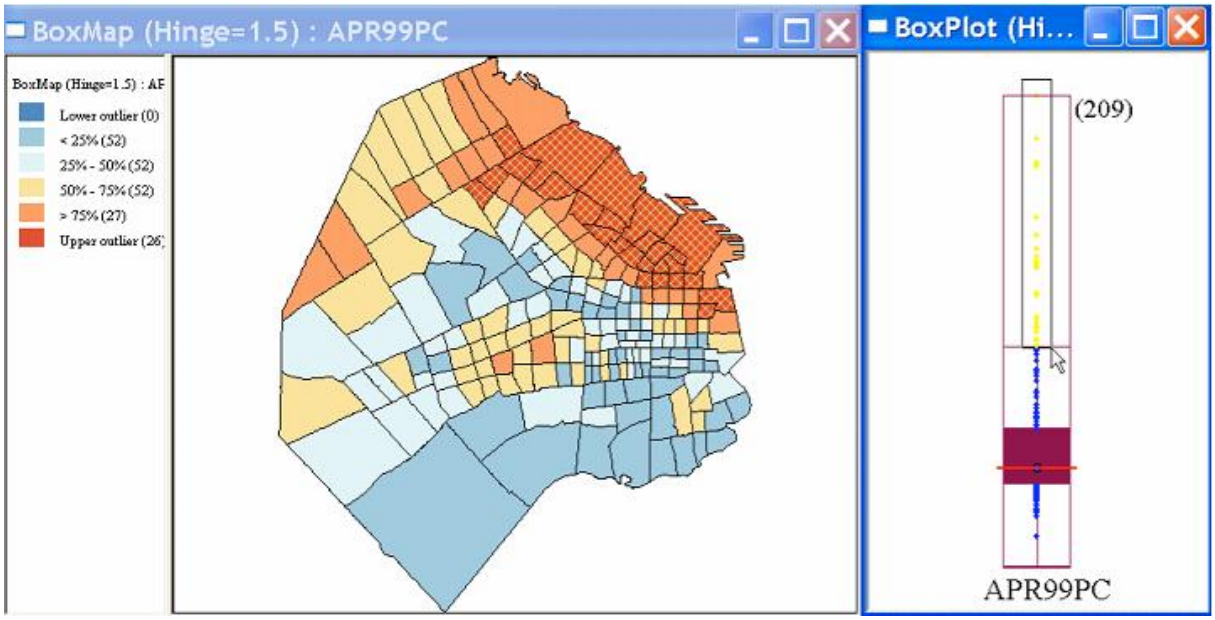
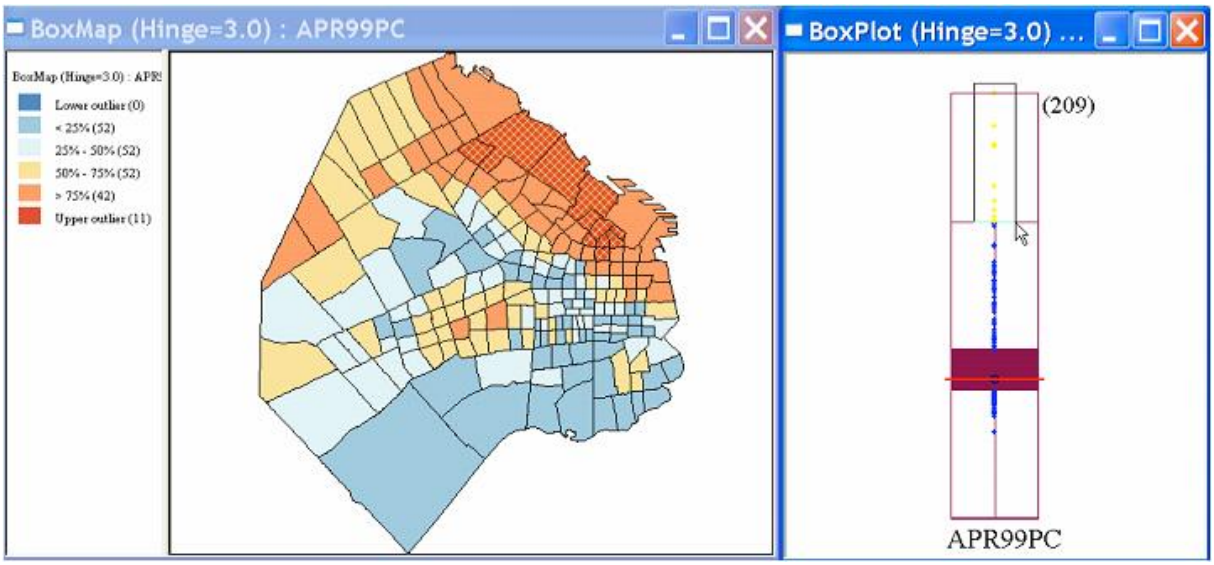


图 11.6 Hinge 为 1.5 的 APR 箱地图

为确认离群值的分类，为同一变量生成一个规则箱图(见 7.3 部分)，利用默认 hinge=1.5，选择上离群值，如图 11.6。注意在箱图中被选中的点如何准确地与箱地图中的暗红色区域相对应。另外要显示高值，地图暗示可以存在空间集聚，而标准箱图是作不到的。

与标准箱图相似，在箱地图定义离群值的标准可以设为 1.5 或 3.0。为 APR99PC 创建一幅新地图，Hinge=3.0，也要在箱图选项中改为此值。一次在箱图中选择离群值，查看它们的位置，如图 11.7 所示。注意在此图中的离群值的空间集聚也是非常明显的。如前面部分，试验 AL99PC 和 TURN99PC 的箱图。比较离群值的位置与它们暗示的空间集聚的程度。



11.7 Hinge=3.0 的 APR 箱图

11.4 统计地图

统计地图是高亮显示地图中极端值的第三种方法。GeoDa 生成一种圆圈统计地图，原始的空间单元被圆圈所代替。圆圈的面积与被选变量的值是成比例的。圆圈本身尽可能以非线性最优原则，紧密地排列在相匹配的空间单元的原始位置。

如同其它地图，统计地图可以从地图菜单(Cartogram)，或从右击任一打开的地图从其下拉列表中调用统计地图功能。

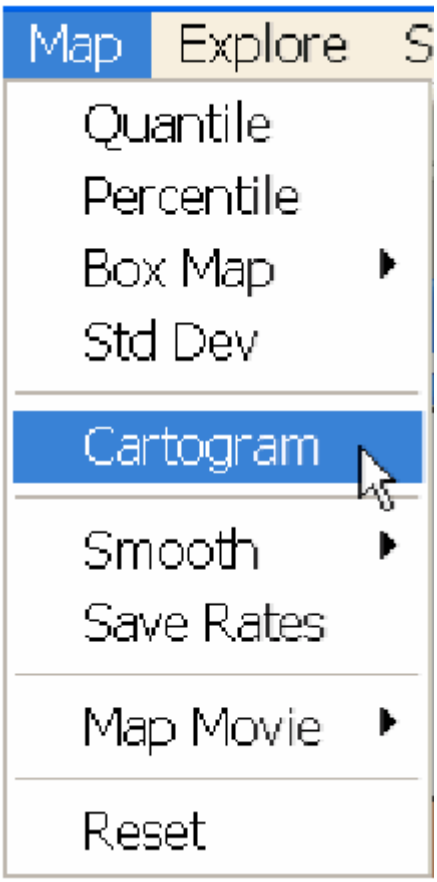
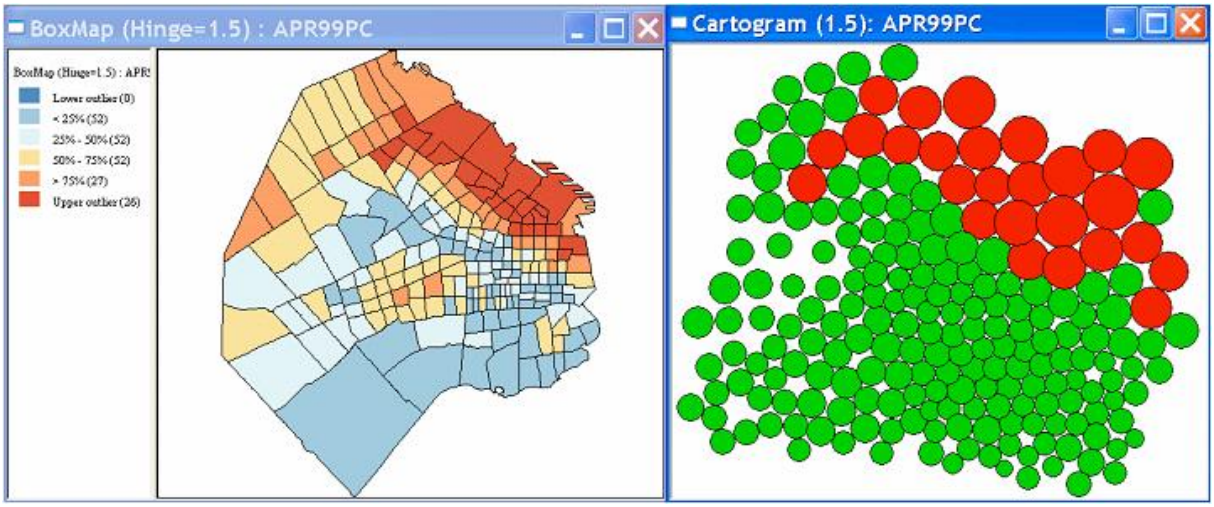


图 11.8 统计地图功能

将打开一个变量选择对话框(见图 11.3)。选择 APR99PC 创建统计地图，如图 11.9。统计地图以不同的颜色从其余的圆圈中高亮显示离群值(上离群值为红色，下离群值为蓝色)。注意图 11.9 中统计地图与箱图之间的普遍相同点。



11.9 Hinge=0.5 的工 APR 统计地图和箱图

圆圈的位置是一种叠代非线性方法的结果，在必要时可以重新定义。在统计地图中右击，选择 Improve cartogram with…>1000 iteratives，如图 11.10。短暂时间后，圆圈会出现跳动，出现

一个稍有不同的排列，如图 11.11。

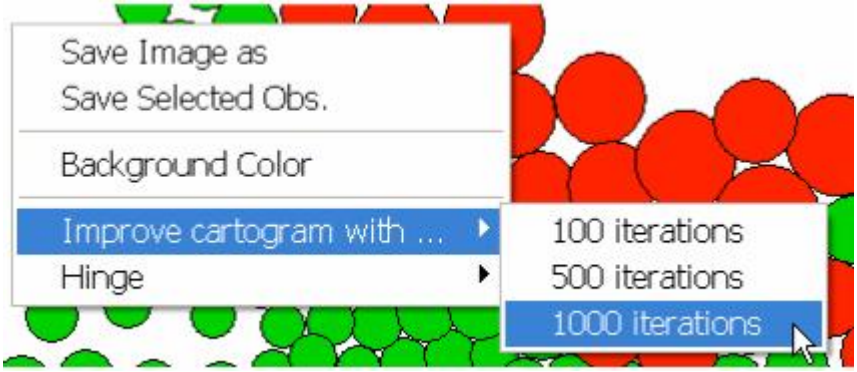
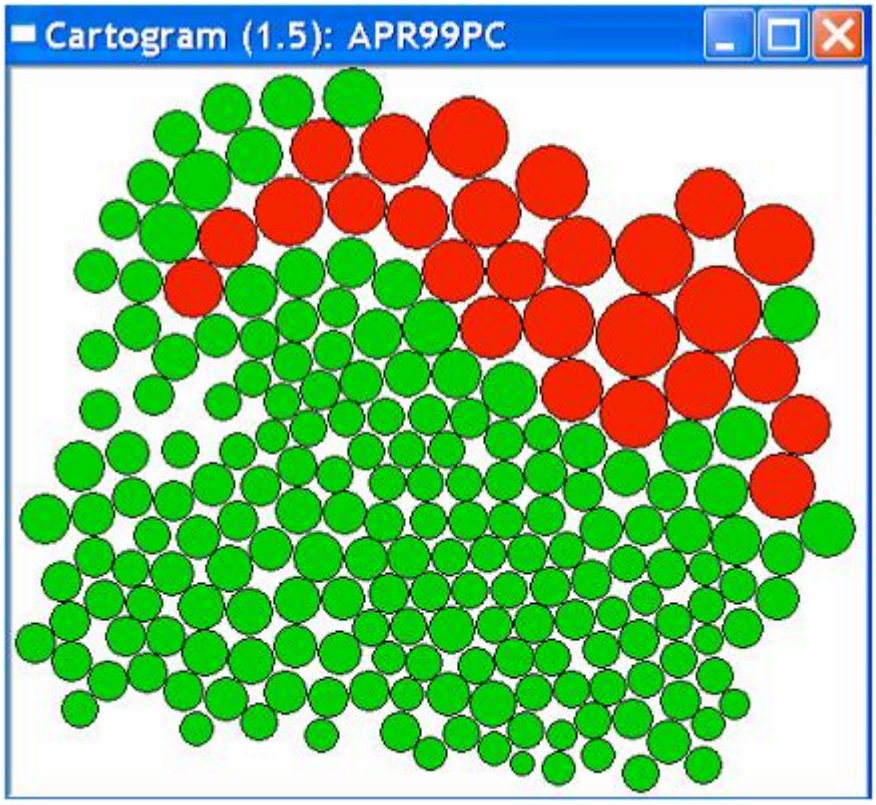


图 11.10 增强统计地图



11.11 增强后的统计地图

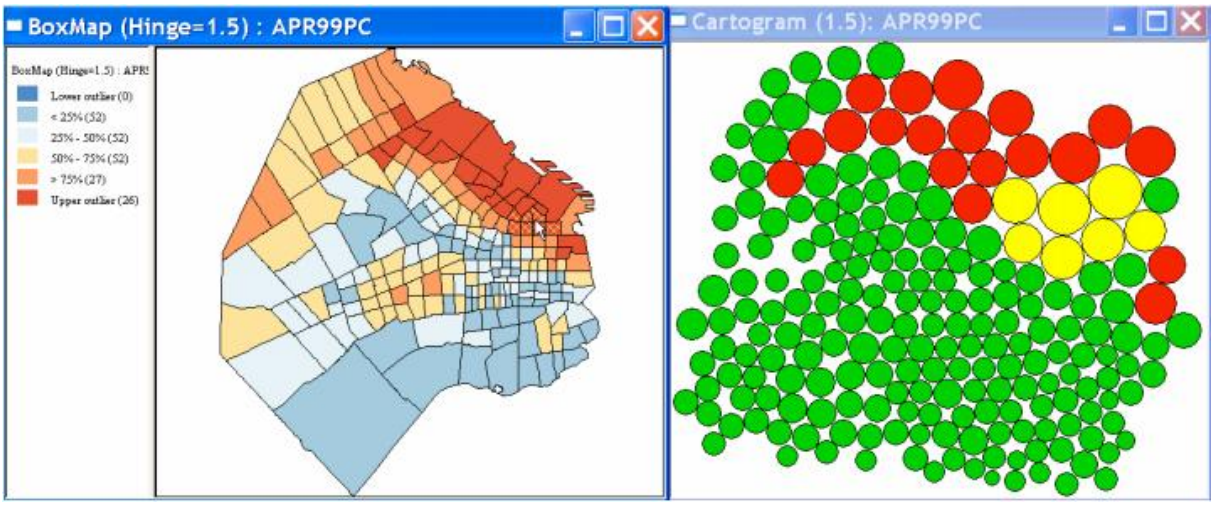


图 11.12 APR 链接的统计地图和箱图

注意统计地图中的其它选项关于 Hinge。因为统计地图高亮显示离群值的方法与箱地图(和箱图)相同，你可以用这一选项改变 hinge 标准。例如改变 Hinge 为 3，比较图 11.7 所示的箱地图的结果。

统计地图在正常的模式链接到所有的其它地图和图表。这对建立观测值的真实空间布局和统计地图中显示的理想布局之间的联系是非常有用的。在地图中，选择位于南部和东部边界的离群值集聚的小区域，注意它们在统计地图中的位置。如图 11.12 所示，它们明显地显示在统计地图中，而在标准 Choropleth 图中，它们几乎不被注意。再试验用 AL99PC 和 TURN99PC 建立统计地图。

11.5 练习

换一个主题，利用 rosas2001.shp 样本数据集(关键字为 ID)中的在位于 Argentina 的 Cordoba 精准农业试验中 1705 个有关谷物产量和相关投入变量的值。为产量变量和 BV(即低有机质显示器)建立百分位图。比较两者暗含的模式。你也为 N 制图，但你有何发现？

练习 12　高级 ESDA

12.1 目标

该练习显示一些 ESDA 更高级的可视化方法:地图动画(map animation)和条件地图(conditional maps)。

在本练习末，你将知道：

- 创建和控制地图动画
- 创建条件地图
- 在条件地图中选择条件类别

更详细的有在该操作的信息请参考 Users Guide 和 Release Notes。

12.2 地图动画

我们继续使用 BUENOSAIRES 样本数据集。如果它不是当前项目，关闭所有窗口，载入文件 buenosaires.shp，关键字为 INDRANO。GeoDa 中，地图动画的简单形式是针对一给定的变量，观测点自动移动，从最低值到最高值贯穿所有观测点。底图中相匹配的观测点要么单个显示，要么是累积显示。

如图 12.1，可以从菜单选择 Map>Map Movie>Cunulative，或点击工具条按钮来调用这一功能。出现变量设置对话框(见图 11.3)。选择 AL99PC(“Alianza”党选举结果)，点击 OK 产生原始的地图电影界面，如图 12.2。

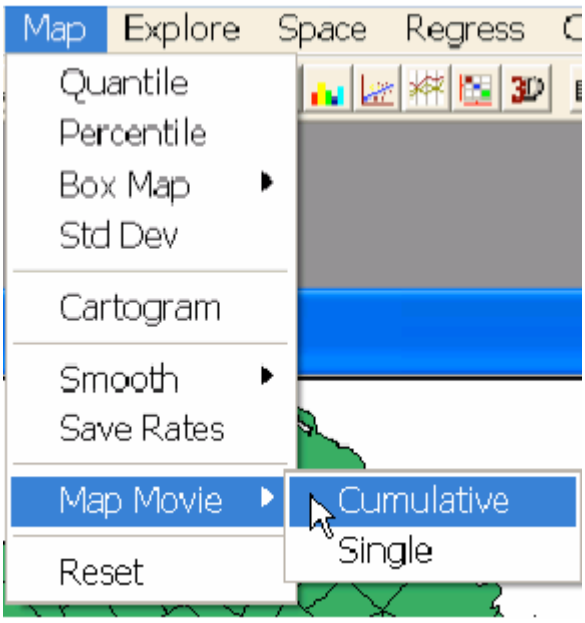


图 12.1 地图电影功能

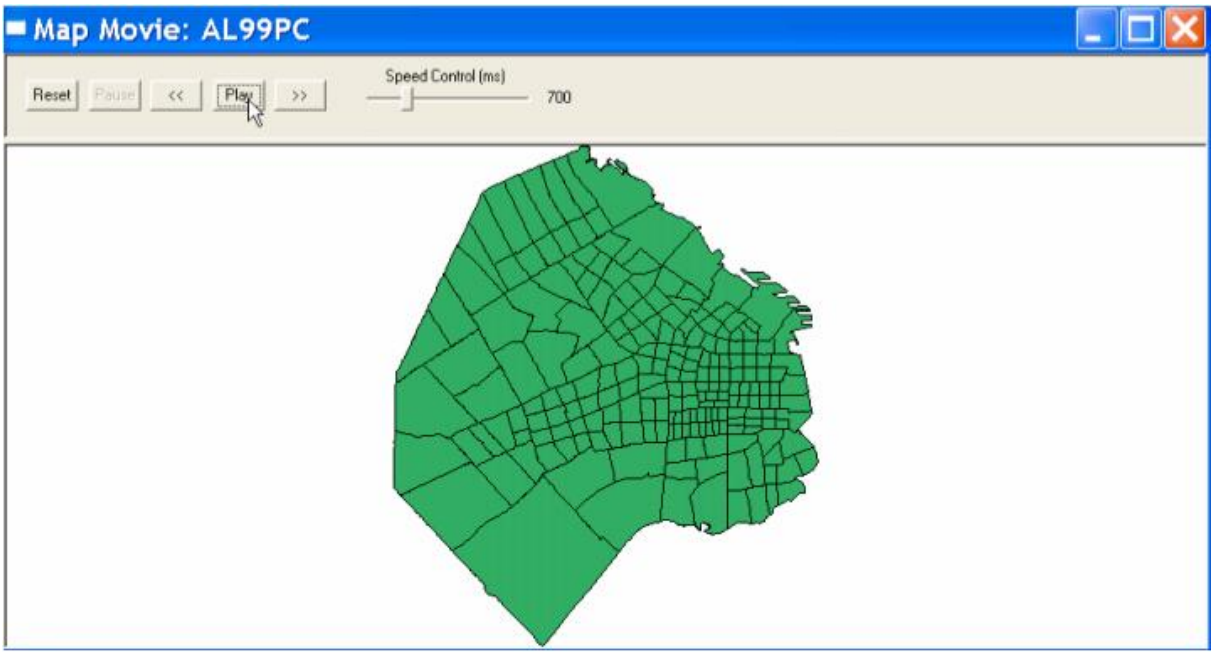


图 12.2 地图电影原始布局

点击 Play 按钮，开始播放电影。地图中的多边形会从最低值到最高值逐渐被品红色的阴影所填充。注意地图电影如何与项目中其它图表或地图相链接的，即电影中的选择是其它所有窗口的选择。你可以在任意时刻点击 Pause 停止电影播放，如图 12.3。点击 Reset 会清除所有被选多边形，开始一个空白的底图。你可以用滑块(Speed Control)改变电影的速度：向左移动滑块会增加速度。

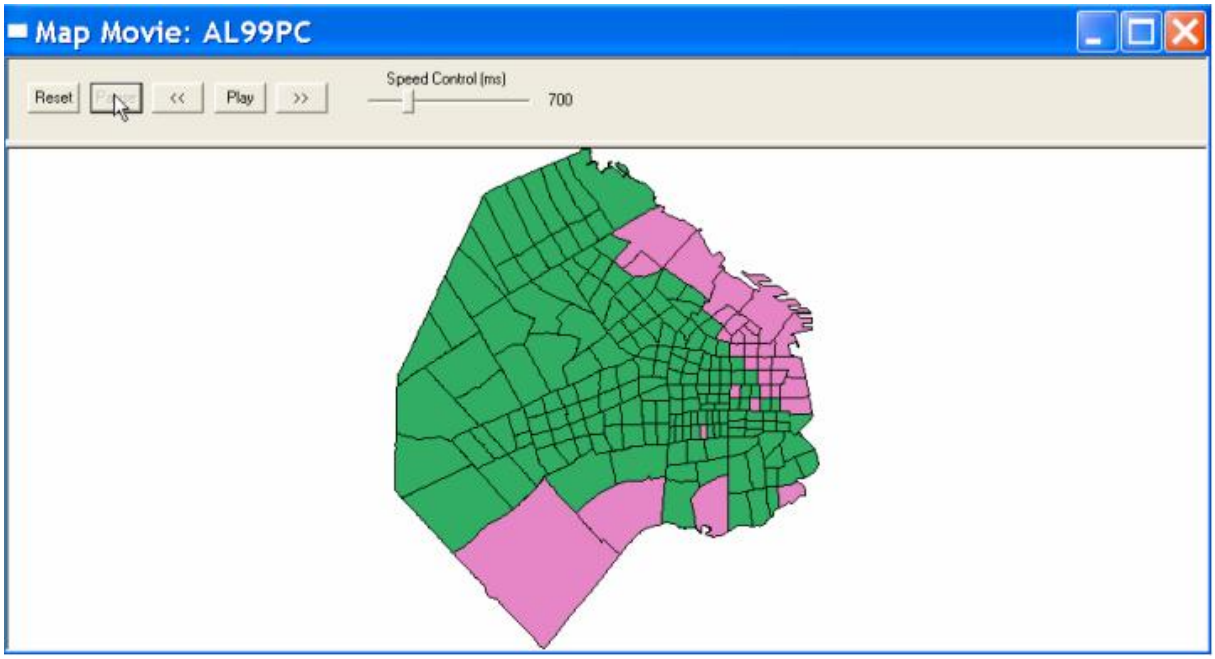


图 12.3 AL 选择结果的地图电影-暂停

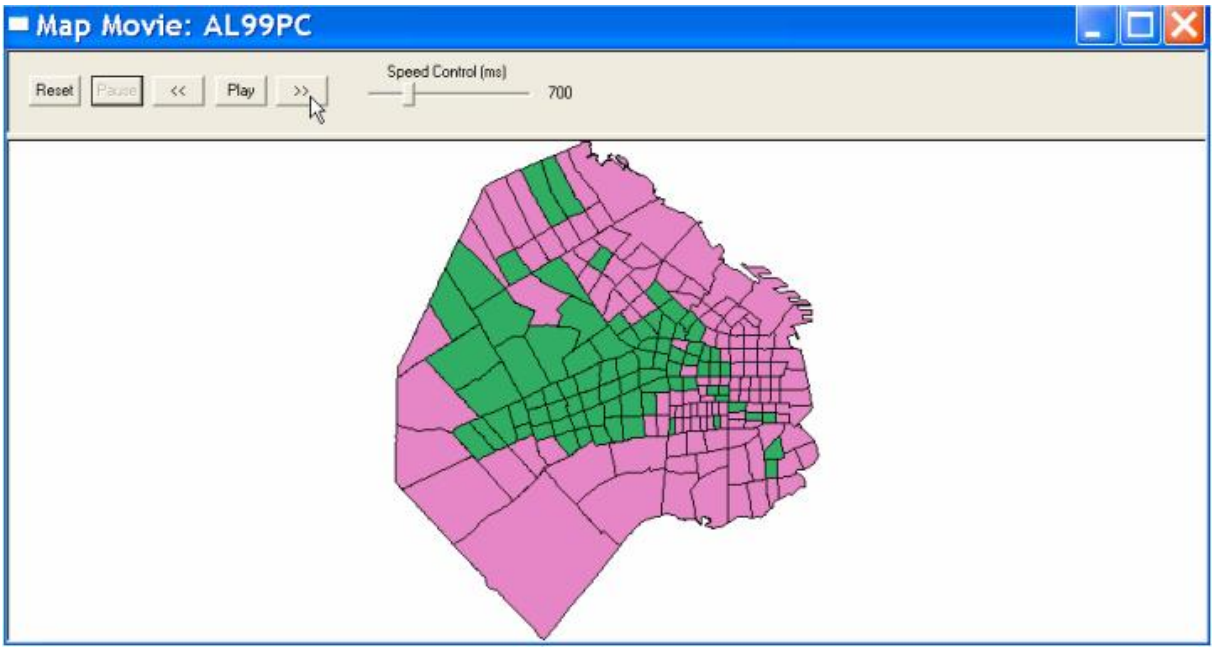


图 12.4 AL 选择结果的地图电影-逐步显示

一旦电影暂停，它可以用>>(或<<)一次前进(或后退)一个观测点。在后一阶段地图的动画处理中会涉及，如图 12.4。

地图动画的目的就是评估相似值位于相近位置的程度。例如在图 12.3 和 12.4 中，AL99PC 低值系统地开始于东北区域，沿外缘移动，高值在城市中心。这与随机模式是不同的，随机模式下，值会在地图中跳跃。在 grid100s.shp 样本数据集中可以看到这样的随机模式的例子。检查作一随机序列的变量(以 ranz 为开始的变量)。

12.3 条件地图

条件地图是 10.2 部分的条件图的一种特殊情况。如前(见图 10.1)开始条件图的功能，在视图类型对话框中选择 Map View 前面的单选按钮，如图 12.5。点击 OK，产生一个变量选择框，图 12.6。

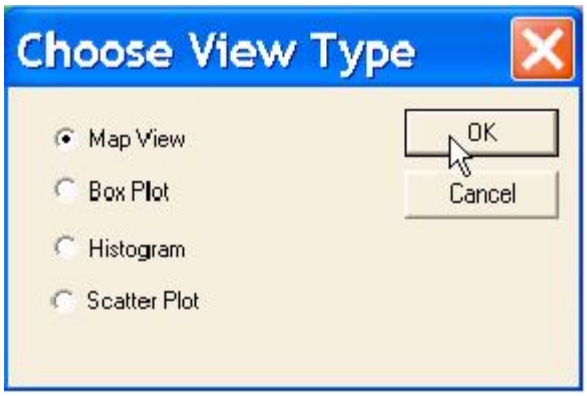


图 12.5 条件地图选项

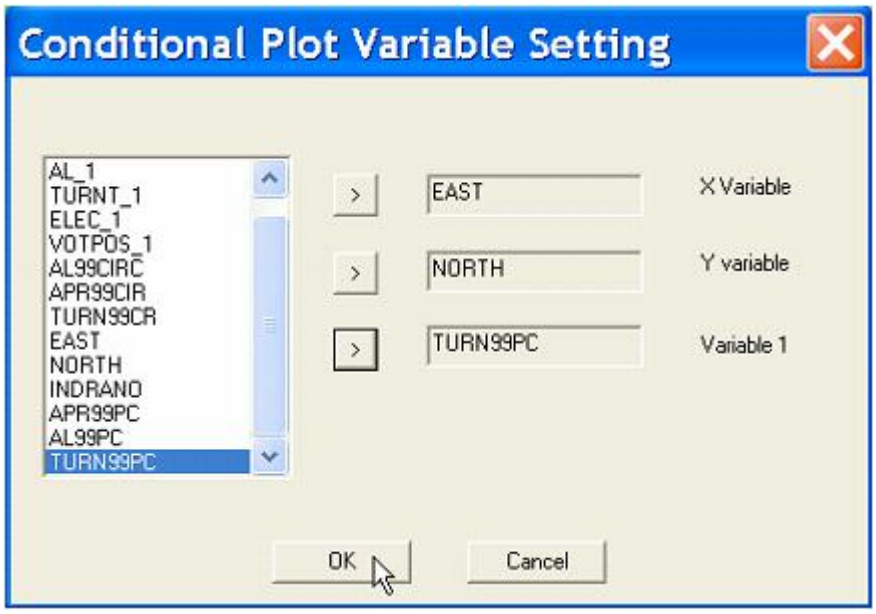


图 12.6 条件地图变量选择

选择 EAST 为 X 变量，NORTH 为 Y 变量。设置兴趣(Variable 1)变量为 TURN99PC。如前面的例子，根据区域位置的地理条件，分为 9 个亚区显示。选择任何其它的两个条件变量，如前面的条件图例子。通过从左向右移动相关手柄，可以改变条件变量的间距。

最后点击 OK，生成一幅条件地图，如图 12.7 所示。地图应用从低值端的蓝绿到高值端棕红色的渐变颜色。渐变颜色显示于图的顶部，指示变量 TURN990C 的变程。

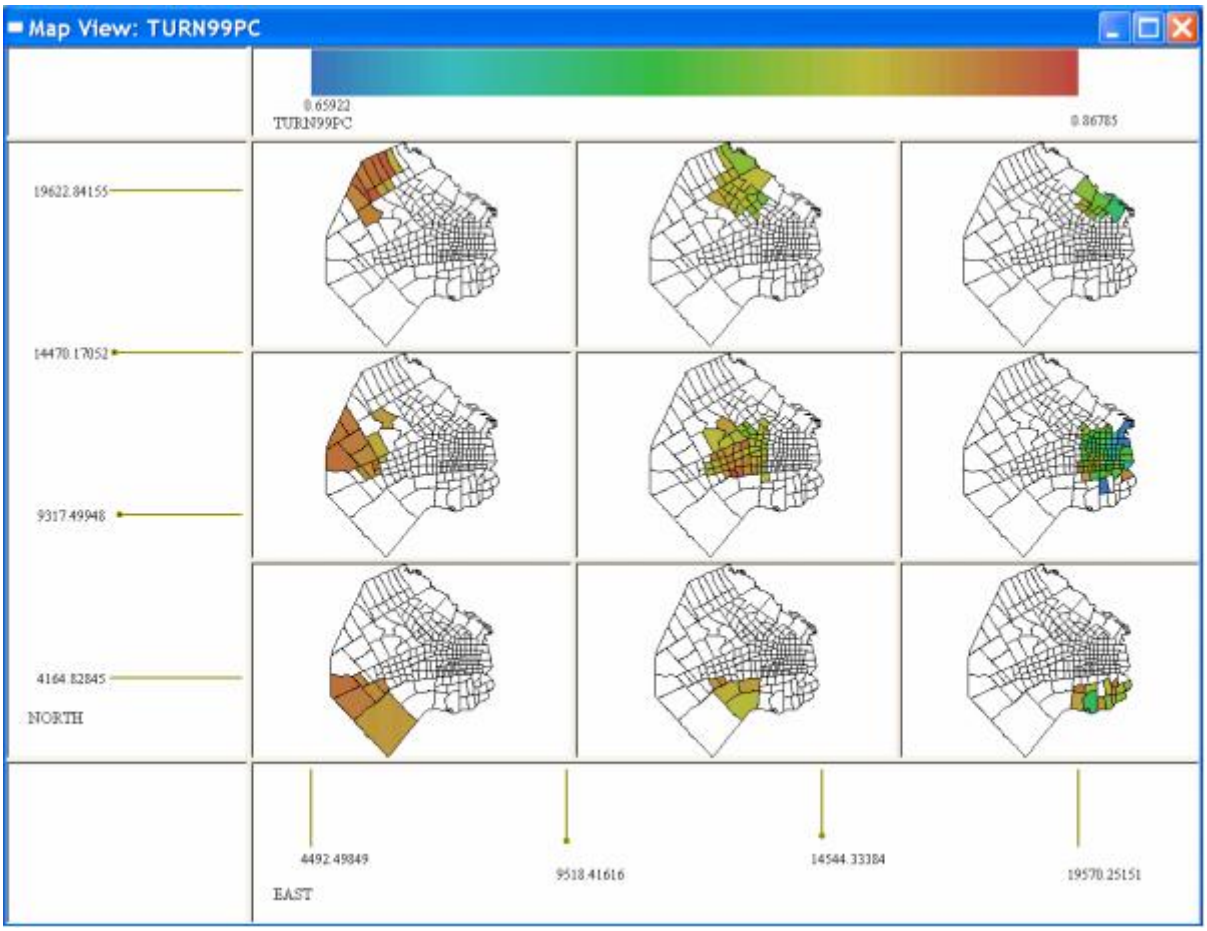


图 12.7 AL 选票结果条件地图

条件地图的目的是评估亚区间变量分布系统差异的程度。图 12.7 中的地图显示高值结果区域在西边，低值结果区域在东边。在更详细的探索中，为开始建立一系列的最终进行回归说明的假设，可以调查其它变量，它们的空间分布可能会显示出相似的模式。

12.4 练习

更近地考虑 Buenos Aires 的三个选区变量(APR99PC，AL99PC，TURN99PC)，评估它们显示出相似或对应地理模式的程度。或者，重新打开 rosas2001.shp 谷物产量样本，或练习 11 中你用过的其它的样本数据。

练习 13　基本的比率制图

13.1 目标

该练习显示一些在制作比率或比例地图时出现的基本概念。

在该练习末，你会知道：

- 创建一幅处于危险的事件或人口比率地图
- 保存计算的比率到数据表
- 创建过度(excess)危险地图

有关该操作的详细信息请参考 Users Guide。

13.2 原始比率地图

我们用 Ohio 州 88 个县肺癌数据制作比率地图，这些数据常用于近来有关疾病制图及空间分析。关闭当前项目，载入 ohlung.shp 样本数据集，关键字为 FIPSNO。出现 Ohio 县的底图，如图 13.1。

比率地图是一种特殊的 choropleth 地图，有独特的界面。不象通常的变量设置对话框，从数据集中选择比率变量，危险事件及人口被指定，比率会迅速计算。

图 13.1 Ohio 州肺癌数据底图

从菜单 Map>Smooth>Raw Rate 选择这一功能，如图 13.2。或者在窗口中右击底图，选择 Smooth>Raw Rate。现在还没有相应的工具按钮来制作比率地图。

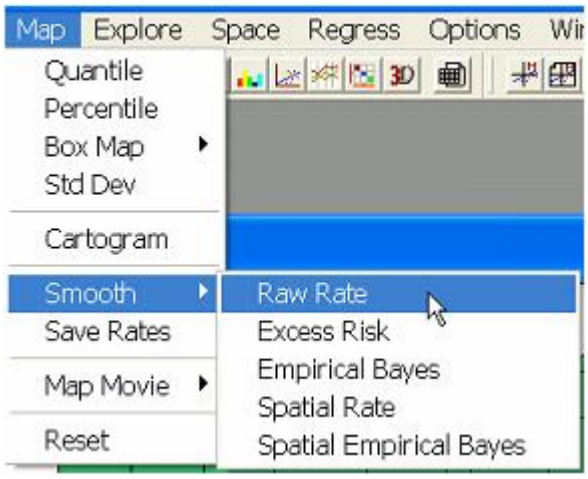


图 13.2 原始比率地图功能

出现 Rate Smoothing 对话框，带有候选事件变量和候选基础(Base)变量，如图 13.3。选择 LFW68 作为事件(1968 年白人女性死亡总人数肺癌)，POPFW68 作为危险的人口(1968 年白人女性人口总数)。然后确信从下拉列表中选择合适的地图类型，如图 13.4。默认为百分位地图，但在本例中并不适用(Ohio 州有 88 个县，少于有意义的百分位地图所要求的 100)。选择箱地图，Hinge 为 1.5，如图 13.4。最后，点击 Ok 产生如图 13.5 左边面板显示的箱地图。三个县由于高死亡率位于高的离群值。但是，由于比率的内在差异不稳定性，这种情况有可能是一种假象。我们回到练习 14。

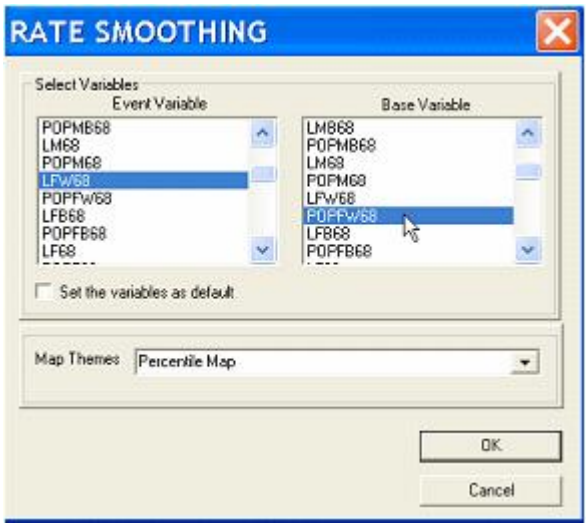


图 13.3 选择事件和基础变量

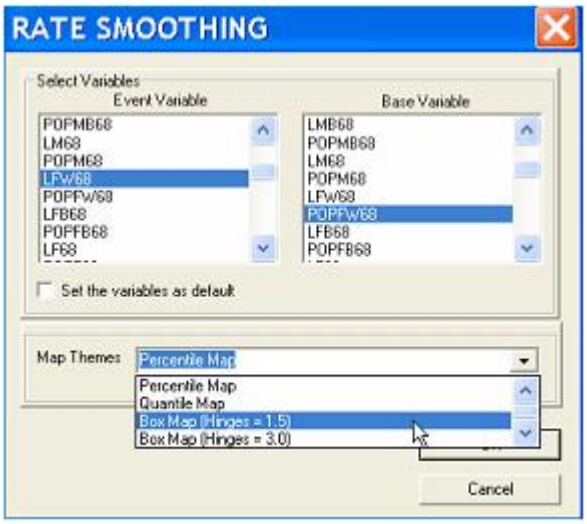


图 13.4 选择比率地图类型

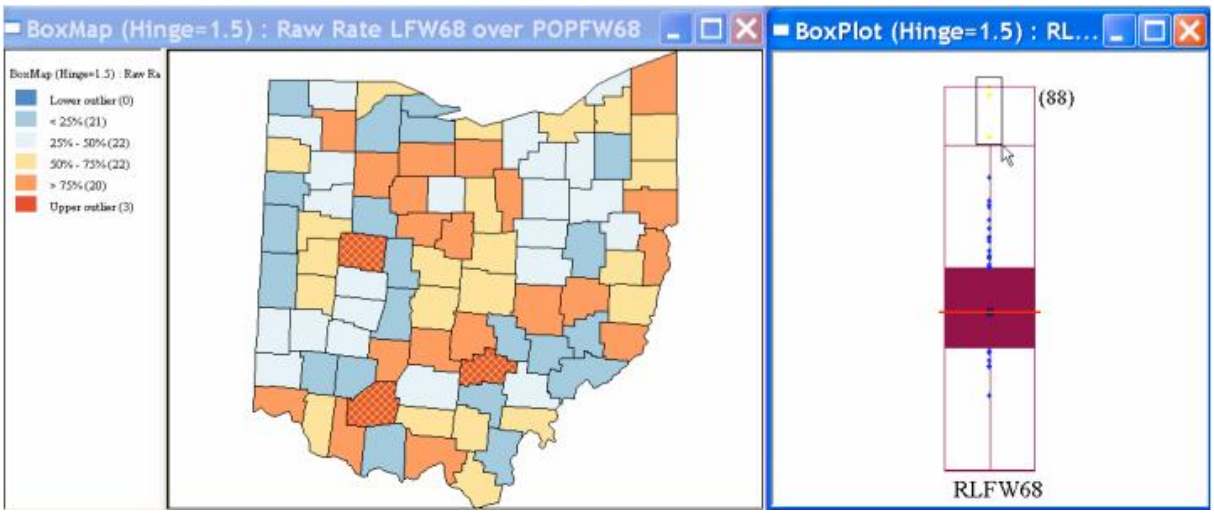


图 13.5 1968 年 Ohio 州白种女人肺癌死亡人数箱地图

虽然我们有肺癌死亡率的地图，但比率本身并不能用于其它分析。但是这将会很容易地得到。在地图中右击，产生菜单选项，如图 13.6 所示。选择 Save Rates 来建立比率变量。图 13.7 的对话框让你指定一个默认值之外的变量。以 RLFW68(或你容易识别的其它变量名)代替默认变量，点击 OK，将新变量添加到表格中。将数据表前方显示，看到已经添加了一个新栏，如图 13.8。

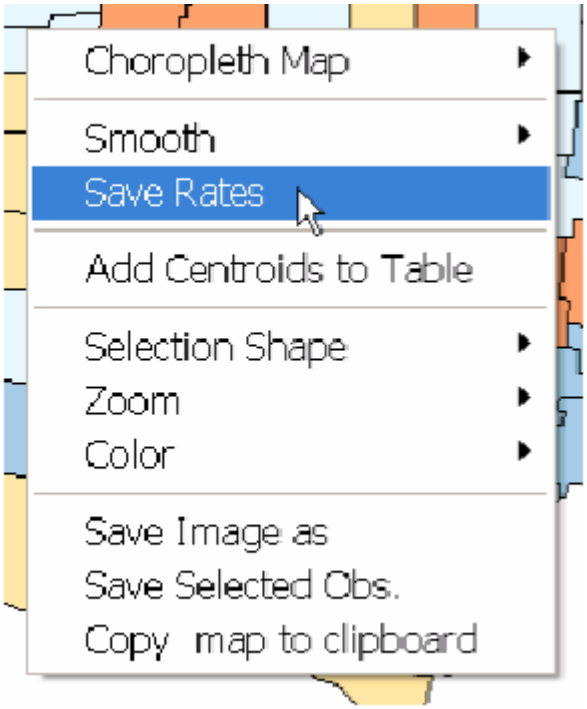
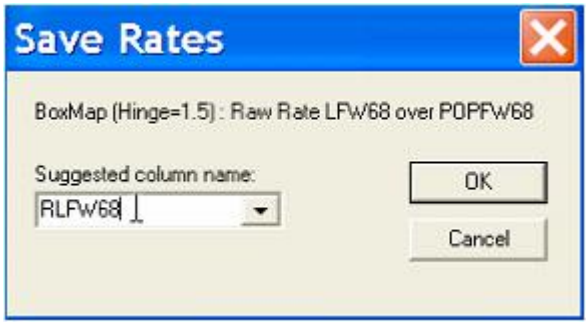


图 13.6 将比率保存到数据表中



13.7 保存比率的变量名

LF88	POPF88	RLFW68
116	241573	0.000123
2	19509	0.000061
12	39013	0.000096
7	18616	0.000000
352	760016	0.000120

图 13.8 添加到数据表中的原始数据

打开箱图功能(Explore>Box plot，或点击工具条按钮)选择 RLFW68 作为变量名。结果将如图 13.5 右侧所示图。选择 3 个离群值，查看它们在箱地图中的位置。你可以调用数据表来查找这三个县的名字(Logan，Highland，Hocking)，查看它们是否有不正常的小的人口基数(POPFW68)。

13.3 过度危险地图

在公共健康分析中常使用的概念是标准死亡率(SMR)的概念，或死亡率与一个国家(或地区)标准的比值。GeoDa 以过度危险地图(Excess Risk)作为 Map>Smooth 功能的一部分(见图 13.9)。

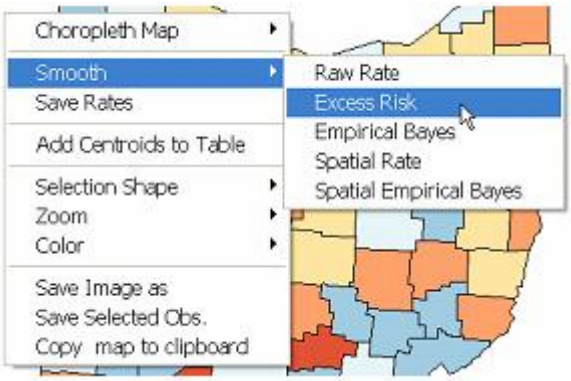


图 13.9 过度危险地图功能

过度地图是观测比率与所有数据计算的平均比率的比值。注意这一平均值并不是县比率的平均值。相反，它是以所有事件与所有危险人口总数的比(例如，在我们的例子中，州中所有白人女性死亡人数与据有白人女性人口总数的比)。

从 Map 菜单选择这一功能，或在任一底图右击选择 Smooth>Excess Risk。在变量选择对话框中，再一次用 LFW68 作为事件，以 POPFW68 作为基数(如图 13.3)。点击 OK，产生地图如图 13.10 所示。地图中的图例分类为硬编码，用蓝色调表示危险程度低于州平均值的县(过度危险比<1)，红色调表示危险程度高于州平均值的县(过度危险比>1)。图 13.5 中的三个离群值的过度比率为 2 到 4。

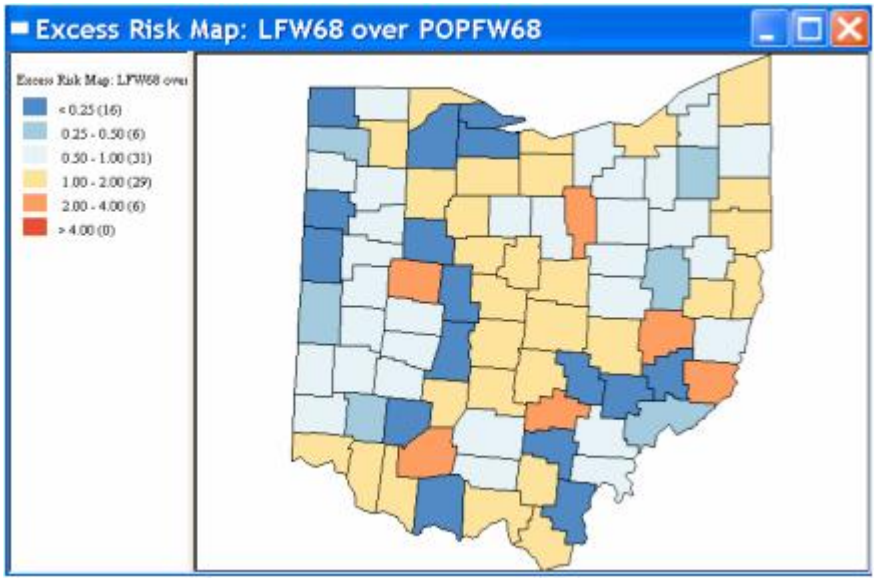


图 13.10 Ohio 州 1968 年白人女性肺癌死亡率过度危险地图

第一次可能会使你出错(throw off)的特点是这种地图的类型是硬编码，正常的地图选项(如图 13.4 中的箱地图，百分位地图)是可能忽略的。为创建一幅熟悉的过度危险比率(或标准死亡率)地图类型，你必须首先将计算的比率添加到数据表中。在地图中右击，选择 Save Rates，如图 13.6 所示。但是，在这一次，提示的新变量名为 R_Excess，如图 13.11。点击 OK 添加过度危险比率到，如图 11.12 所示。

标准比率可以用于任何类型的分析，图表或地图。例如图 13.13 显示一幅计算过的过度危险比率箱地图(R_Excess)。与图 13.5 的箱地图来比较。

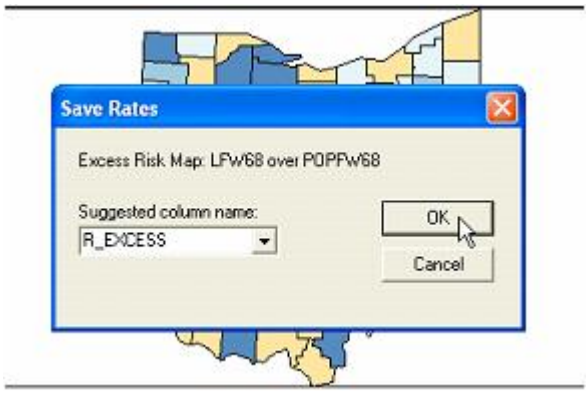


图 13.11 保存标准死亡率地图

LF88	POPF88	R_EXCESS
116	241573	1.123431
2	19509	0.559866
12	39013	0.882557
7	18616	0.000000
352	760016	1.097976
10	20284	0.000000
10	56848	0.000000
35	137489	0.835486
13	30898	0.000000
41	118184	0.910869
1	14458	1.362841
20	38983	1.039799
7	19907	0.494215
120	266561	0.842739
17	68919	0.304020

图 13.12 添加到数据表中的 SMR

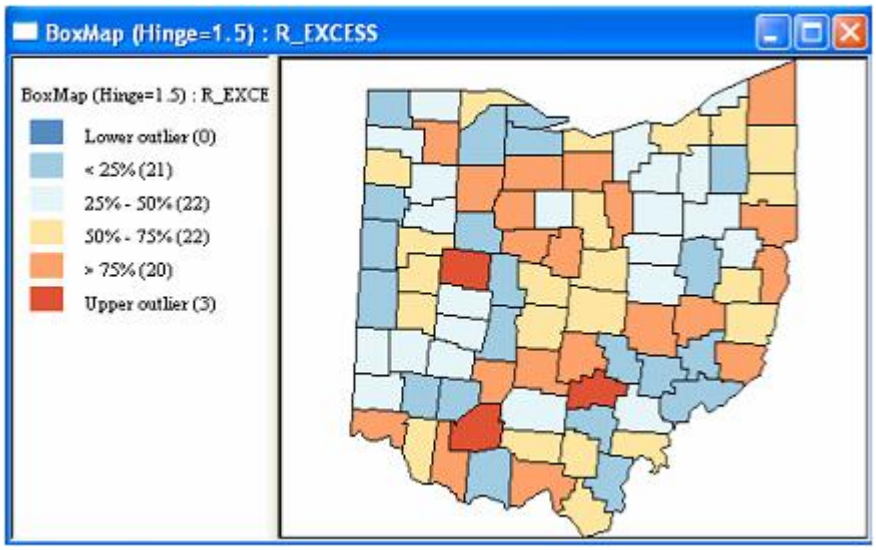


图 13.13 过度危险比率箱地图

13.4 练习

实验 Ohio 州肺癌数据集中其它人口类别和/或年代的比率计算和标准人口死亡率。或者，考虑

56 个 Scottish 地区著名的数据集中包含的唇癌死亡率(scotlip.shp，关键字为 CODENO)，或 100 个 North Carolina 县（sids.shp，关键字为 FIPSNO）的 SIDS 数据。杀人案数据集和 Buenosaires 选举结果也可以进行此类型的分析。

练习 14　比率平滑

14.1 目标

该练习显示了平滑比率地图来检查比率的内在差异不稳定性的一些方法。

在练习末，你将知道：

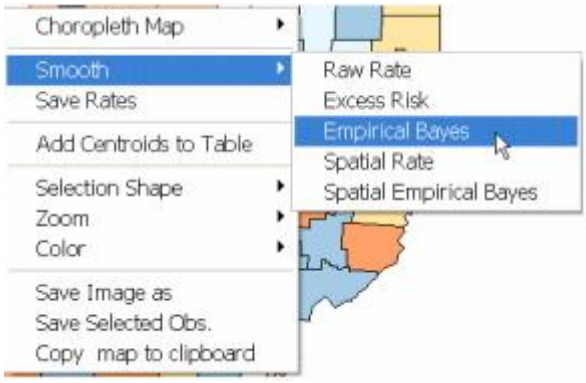
- 通过经验贝叶斯方法创建平滑比率的地图
- 创建 K-近邻空间权重文件
- 创建空间平滑比率地图
- 将计算的比率保存到数据表

有关本操作的详细信息请参考 Users Guide。

14.2 经验贝叶斯平滑

我们继续使用 Ohio 肺癌样本。如果你使用其它数据作前面的练习，首先关闭所有窗口，载入文件 ohlung.shp（关键字为 FIPSNO）。第一种平滑方法使用贝叶斯方法，但对于全州平均水平而言，原始数据被“缩水”了。在本质上，EB 方法包括计算每一个县的原始数据和州平均值之间的平均权重，与处于危险的潜在人口成比例的权重。简单说，小县（处于危险的人口少）的比率将会变化较大，但大的县比率将几乎不变。

从 Map 菜单调用这一功能，或在当前地图中右击，选择 Smooth>Empirical Bayes(图 14.1)。出现如前面的比率地图相同的变量选择对话框，如图 14.2。



14.1　经验贝叶斯比率平滑功能

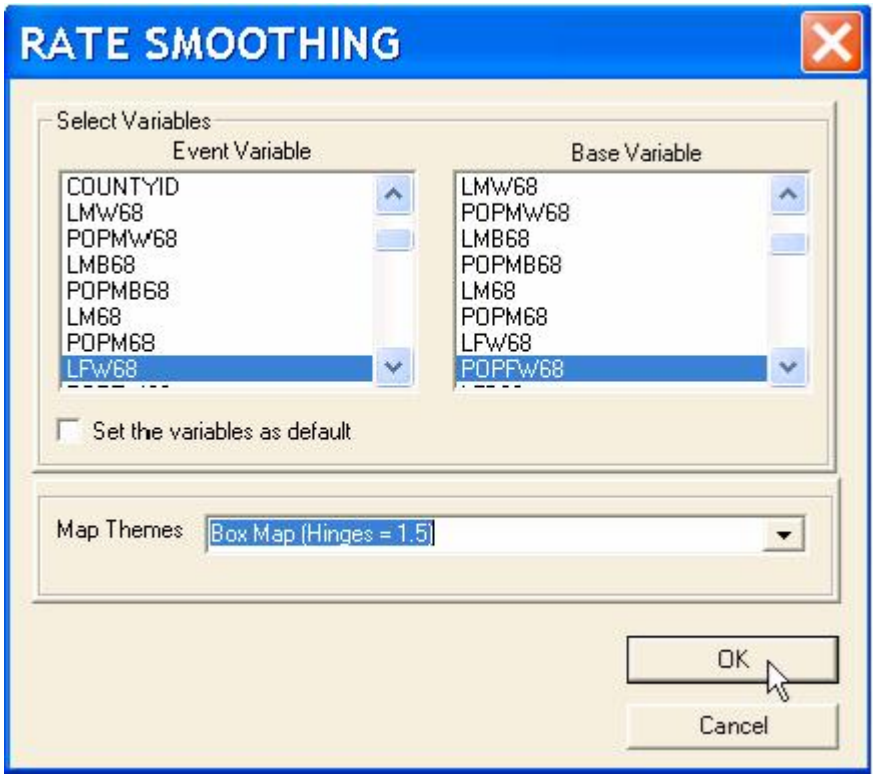


图 14.2　经验贝叶斯事件和基数变量选择

选择 LFW68 作为事件，POPFW68 作为基数，选择 Box Map，Hinge＝1.5，如图 14.2 所示。点击 OK，出现图 14.3 左边面板所示的平滑后的箱地图。将其与图 13.5 中原始箱地图相比较。注意平滑后原来的离群值已经不是离群值，但 Hamilton 成为新的离群值（在西南角）。

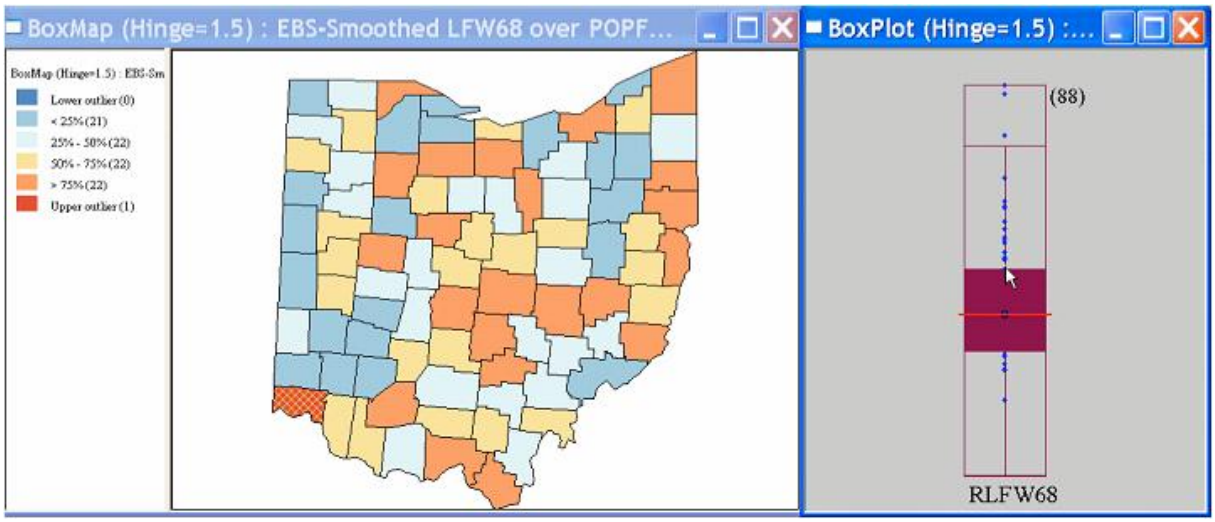


图 14.3　EB 平滑后的 Ohioe 各县肺癌发生率箱图

用练习 13 计算的原始数据 RLFW68（如果那时没有保存比率变量，创建一幅原数据箱图，随后保存比率）创建一幅箱图。在箱地图中点击离群值，看它在箱图中的位置。如图 14.3 右边面板中箭头所示，它在原始比率地图中大约 75%的位置。

由于许多原始离群县有处于危险的小的人口数（在数据表中查看），它们的 EB 平滑率与原始数据有较大不同。相反，Hamilton 县是有最多人口的县之一（它包括 Cincinnati 市），所以它的原始比率几乎不动。由于以上原因，它位于分布的顶端，成为一个离群值。你可以在箱图中为原始比率系统选择观测点，比较它们在累积分布的位置与平滑后比率的位置，查看哪个观测点是影响最大的。

利用数据表它们有小的人口。

14.3 空间人口比率平滑

空间比率平滑包括在轮流以每个县为中心的移动窗口中计算比率。移动窗口包括县及它的近邻。在 GeoDa 中，近邻通过空间权重文件来定义。这将在练习 15 和 16 中详细谈论。但是，为能够显示空间平滑，下面就进行创建空间权重的快速入门。

14.3.1 空间权重快速入门

我们将会创建一个简单的每个县有 8 个近邻的空间权重文件。在工具条中点击 Create weights，或从菜单 Tools>Weights>Create 调用此功能（如图 14.4）。

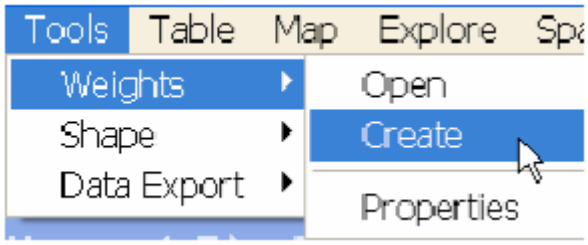


图 14.4　空间权重创建功能

这将产生权重创建对话框，如图 14.5 所示。键入输入文件 ohlung.shp 的路径，输出文件为 ohk8（一个文件扩展 GWT 将会加入到程序中），在下拉列表中为 ID 变量选择 FIPSNO。保留 Distance Weight 下所有选项为默认值（程序将会生成计算距离所需的 Ohio 各县质心）。最后，选择 K-nearest 近邻前的单选按钮，近邻数改为 8。

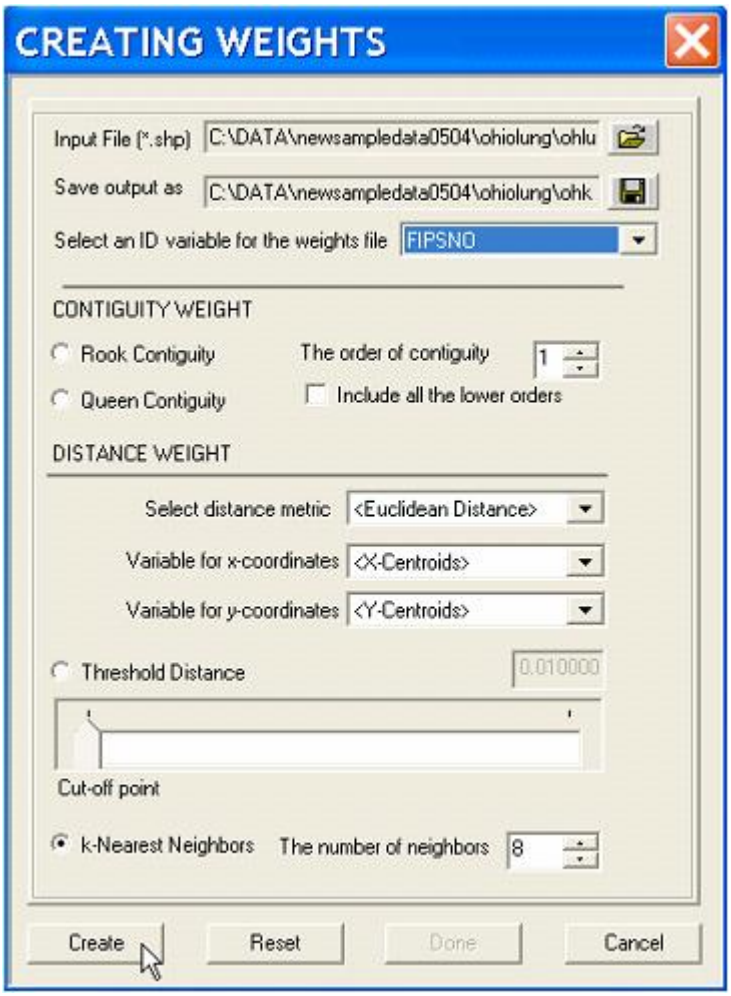


图 14.5　空间权重对话框

点击 Create 开始创建过程，当进度条（在 shp->gwt 窗口中的蓝色条）完成点击 Done。你现在有一个权重文件可以使用。

在开始空间平滑之前，你必须载入空间权重文件，使其可以被程序所调用。点击 Load weights 工具条按钮，或从菜单中选择 Tools>Weights>Open，如图 14.6。然后点击选择权重对话框中的单选按钮 Select from file（图 14.7），输入权重文件名。点击 OK 载入权重文件。现在已准备完毕。

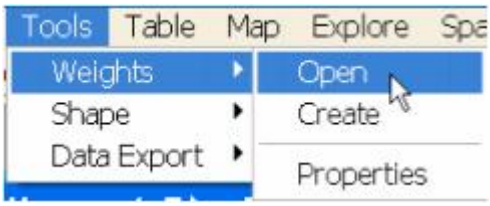


图 14.6　打开空间权重功能

图 14.7　选择空间权重对话框

14.3.2 空间平滑地图

载入空间权重文件后，从 Map 菜单调用空间平滑功能，或在地图中右击选择 Smoothing>Spatial Rate，如图 14.8。如果现在没有载入空间权重文件这时会出现一个错误消息。如果一切良好，在对话框中选择 LFW68 为事件，POPFW68 为基数（见图 14.2）。如前，从下拉菜单中选择 Box Map，Hinge=1.5，

点击 OK 创建地图。出现平滑后的地图，如图 14.9。

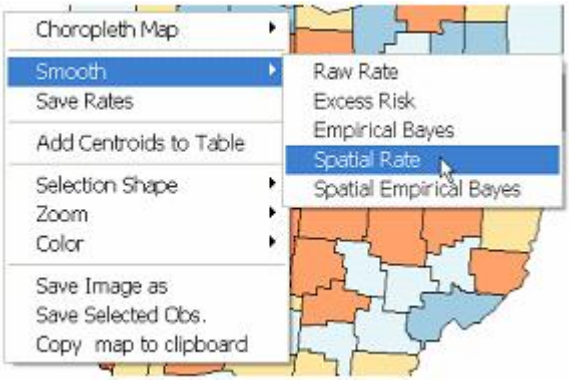


图 14.8　空间比率平滑功能

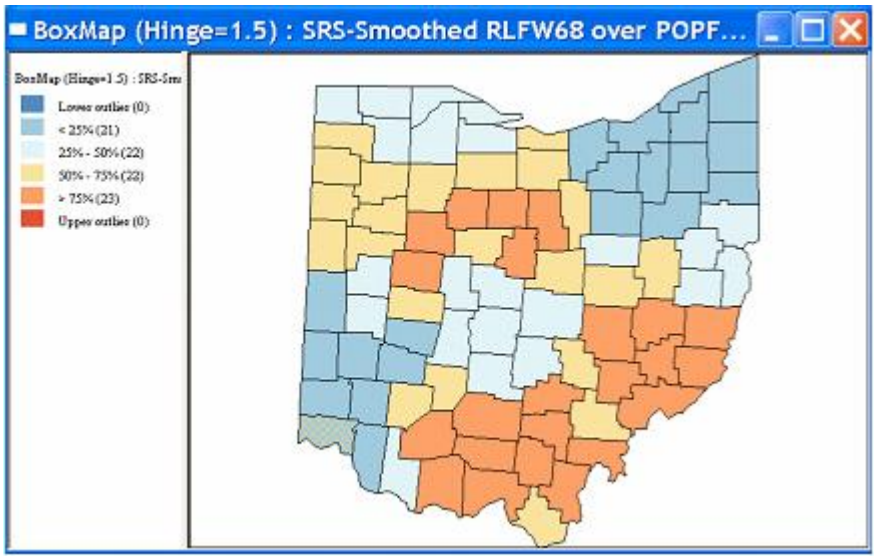


图 14.9　Ohio 州各县肺癌比率空间平滑箱地图

空间平滑地图强调地图中宽的区模式。注意没有离群值。还有，由于用 8 个近邻来平均，Hamilton 县成为第二分位县区域的一部分。

14.4 练习

如练习 13，进一步探索 Ohio 州肺癌数据中不同类别和/或年代原始比率与平滑后比率在空间模式方面的不同点和相同点。或者，使用前面用过的任何一个样本数据。关注原始比率地图和 EB 平滑地图的离群值存在的差异，以及关注空间平滑地图中出现的宽的区域模式。

练习 15　基于邻接（Contiguity）关系的空间权重

15.1 目标

本练习开始演示空间权重操作，建立基于空间邻接关系的权重。这里邻接的意思是具有公共边界。

在本练习末，你将知道：

- 用 rook 和 qween 标准建立多边形文件的一次邻接空间权重文件。
- 在直方图中分析连接结构。
- 将一次邻接权重文件转变为更高次邻接。

关于本操作的详细信息请参考 Users Guide，和 Release Notes。

15.2 基于 Rook 的邻接性

打开样本数据开始本次练习，它包含了 Sacramento，CA403 个人口调查区域的人口变量（sacramentot2.shp，关键字为 POLYID）。底图如图 15.1。

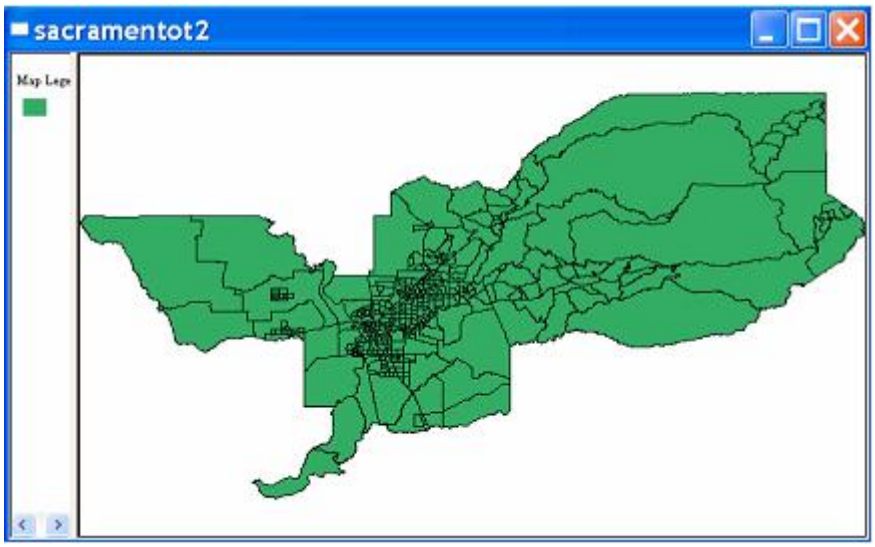


图 15.1　Sacramento 人口调查区数据底图

从菜单中来调用权重建立功能，选择 Tools>Weight>Create，如图 15.2。这可能在没有打开项目的情况下执行。换句话说，要创建权重文件，并不需要载入文件。或者，在一个项目是，可以单击相应的工具条按钮来打开这一功能。

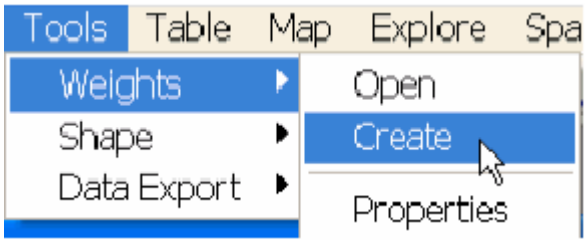


图 15.2　创建权重功能

权重创建功能产生一个对话框，需要指定相关的选项。首先是输入文件的文件名（对于邻接性权重，这个文件必须是多边形文件）和权重文件的文件名。前者输入 sacramentot2.shp，后者输入 sacrook，如图 15.3 所示。一个文件扩展 GAL 会被程序加入到权重文件中。指定关键字变量很重要的：输入 POLYID，如图 15.3。但这不是必须的，它确保表格中的数据与权重文件中相应的邻接实体能够完全匹配。

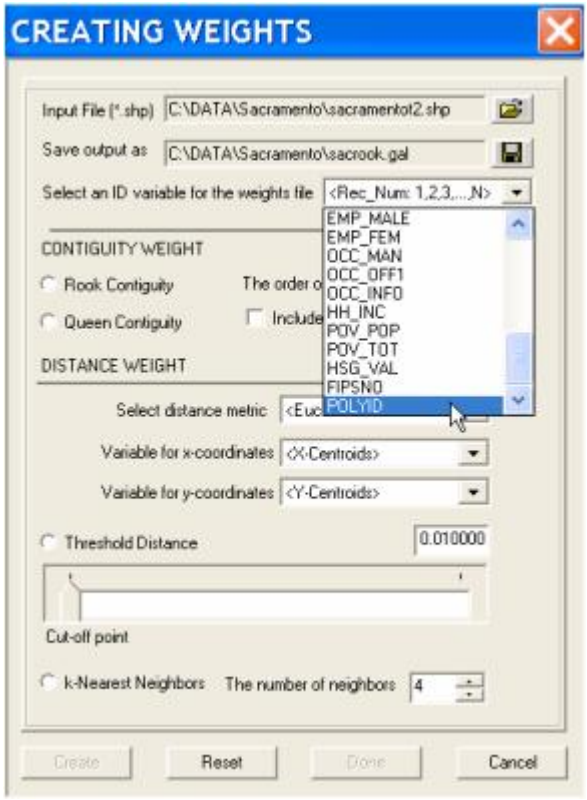


图 15.3　权重创建对话框

对 Rook 邻近权重文件所需的其它操作只是选择 Rook Contiguity 旁边的单选按钮，如图 15.4。然后点击 Create 开始创建权重。会出现一个进度条，如图 15.5，显示处理完成的时间（这会在很短的时间里完成）。点击 Done 结束这一过程（见图 15.5）回到标准界面。

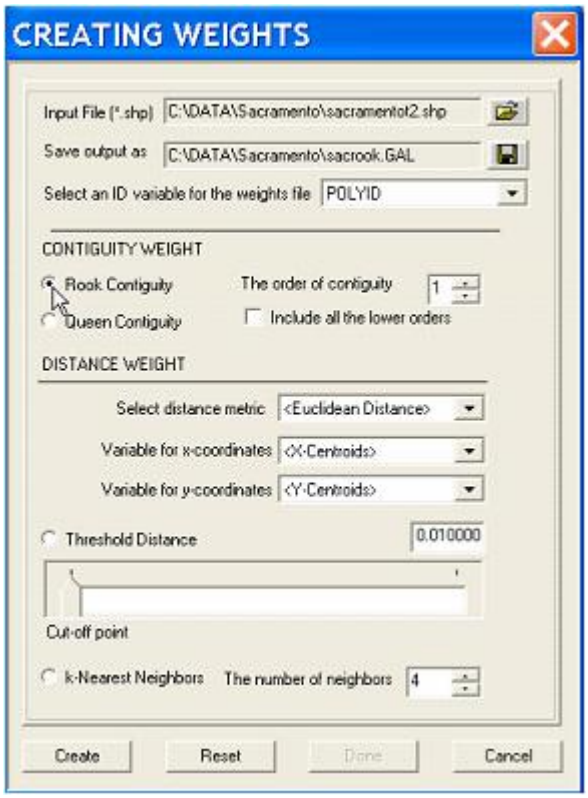


图 15.4　Rook 邻接性

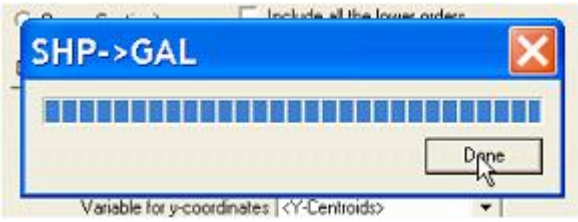


图 15.5　GAL 文件被创建

GAL 格式空间权重文件是一个简单的文本文件，可以用任一文本编辑器或文字处理器（确保能存储为文件文件）来修改。例如，对 Sacramento 人口区域，利用 POLYID 为关键字，文件 sacrook.gal 的部分内容如图 15.6 所示。第一行为本文件的头行，包括 0（将来要使用的标志），观测点数目，邻接结构来源的多边形文件名（Sacramento），关键字变量名（POLYID）。

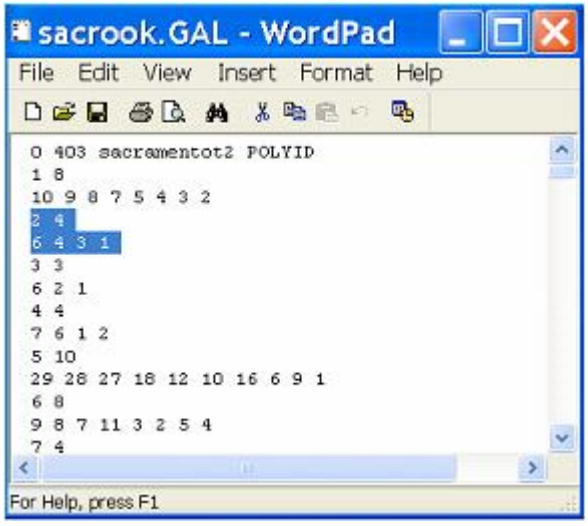
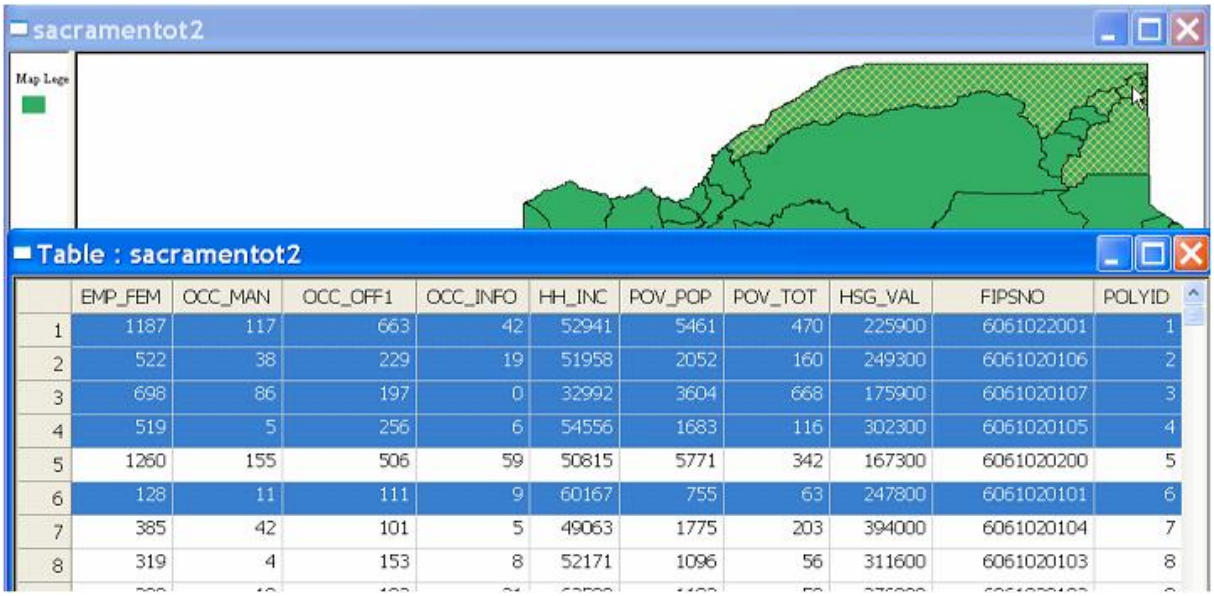


图 15.6　GAL 文件内容

注意，如果文件与权重文件不在同一目录下，GeoDa 会出现一个错误消息。如果在分析过程中，文件名或关键字变量发生变化，也经常会修改头行。在文件编辑器中打开 sacrook.gal，注意 POLYID2（FIPS 代码 6061020106）的邻居。打开表（点击表工具按钮），选择此区域。被选择的位置应当是在底图中所指的那一个，如图 15.7。第二个实体有 4 个邻居，图 15.6 所示的第一个高亮行中由所指示“4”。这些邻居的关键字变量的值（POLYID）在第二行列出。在表格中选择这些区域（用 shift-click），如图 15.7，地图中的相应位置会高亮显示。试下下其它位置。注意在一些情况下，Rook 标准会清除 Cornor 邻居，如在通常情况下没有完整边界的区域。这些位置将被包括在 Queen 标准中，在 15.4 部分会谈到。



	EMP_FEM	OCC_MAN	OCC_OFF1	OCC_INFO	HH_INC	POV_POP	POV_TOT	HSG_VAL	FIPSNO	POLYID
1	1187	117	663	42	52941	5461	470	225900	6061022001	1
2	522	38	229	19	51958	2052	160	249300	6061020106	2
3	698	86	197	0	32992	3604	668	175900	6061020107	3
4	519	5	256	6	54556	1683	116	302300	6061020105	4
5	1260	155	506	59	50815	5771	342	167300	6061020200	5
6	128	11	111	9	60167	755	63	247800	6061020101	6
7	385	42	101	5	49063	1775	203	394000	6061020104	7
8	319	4	153	8	52171	1096	56	311600	6061020103	8

图 15.7　Sacramento 人口区域 Rook 邻接结构

15.3 连接性（Connectivity）直方图

从菜单选择 Tools>Weight>Properties(如图 15.8)，创建一幅反映数据集人口区域连接性分布的直方图。或者，点击相应的工具条按钮。

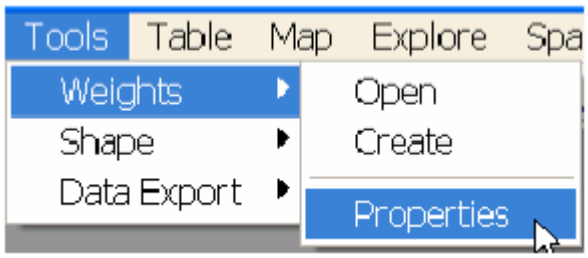


图 15.8　权重属性功能

直方图用于探测分布的奇怪特征是很重要的，它可以影响空间自相关统计和空间回归的解释。特别是需要注意两个方面。一是出现“岛”，或不连接的观测点，另一个是双峰分布，一些位置有非常少（如一个）的邻居，而其它有非常多的邻居。

选择权重属性功能产生一个对话框来指定权重文件，如图 15.9。输入 sacrook.gal，点击 OK。

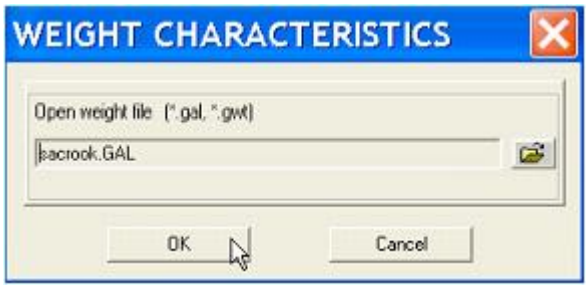


图 15.9　权重属性对话框

结果直方图如图 15.10，位于 Sacramento 地区底图的旁边。直方图用邻居数量（图例所示）描述了位置的分布（每一类的观测点数量显示于相应条块的顶端）。例如，最右边的条对应于有 14 个邻居的地区。点击直方图中的条块查找其在地图中的位置，如图 15.10 所示。或者，在地图中选择一个位置，从直方图中查看其所有的邻居数量。利用地图的缩放功能（在地图上右击，选择 Zoom，创建一个矩形选择区域）更清晰查看邻居结构。试验选择一些区域，将它们的邻接性直方图与全局分布相比较。

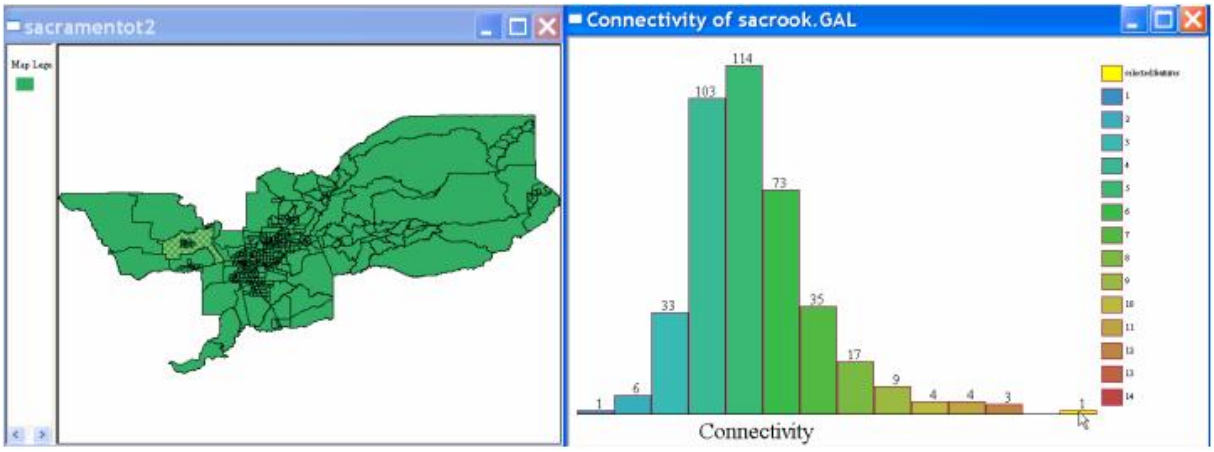


图 15.10　Sacramento 人口区域 Rook 邻接直方图

为显示岛状地区的连接性结构，载入数据集中 56 个 Scottish 地区（载入 scotlip.hp，关键字为 CODENO），创建一个 rook 邻接性文件（如 scotrook.gal）。建立连接直方图选择最左边的条块，对应于 0 个邻居（岛状）。图如 15.11 所示，在数据集中有三个岛状区域。

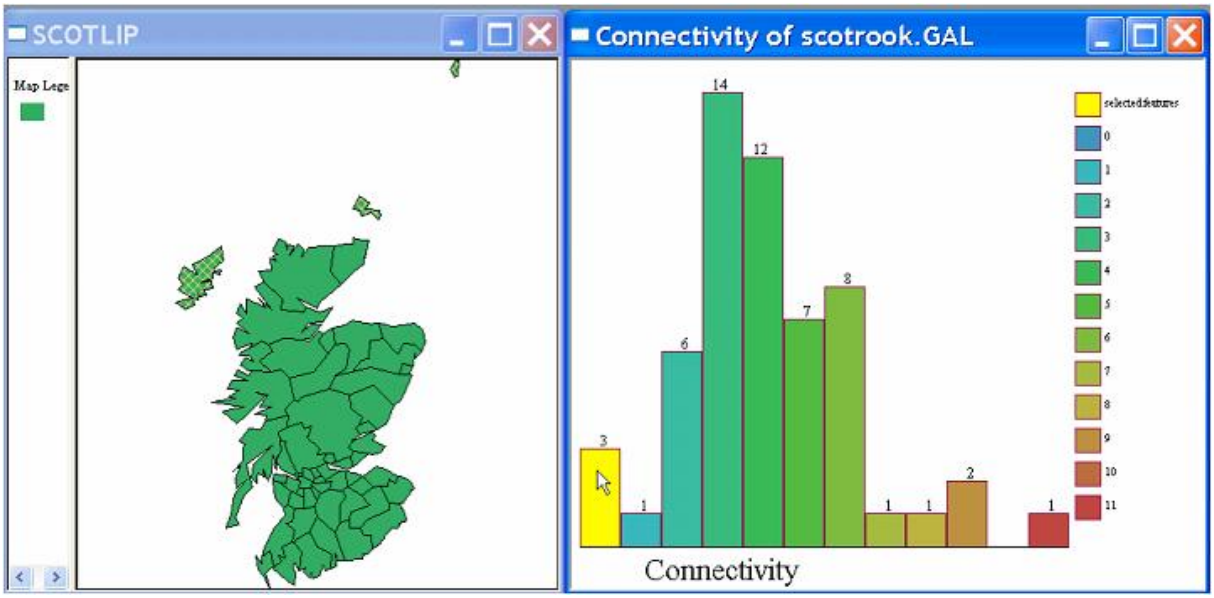


图 15.11　邻接性直方图中的岛状区域

15.4 基于 Queen 的邻接性

利用 queen 标准来定义邻居并创建邻接性权重文件与 Rook 标准具有相同的模式。如前，选择 Tools>Weight>Create(见图 15.2)产生一个权重创建对话框，图 15.12 所示。选择 sacramentot2.shpw 作为输入文件，指定 sacqueen 为输出文件名。确定设置 ID 变量为 POLYID。选择 Queen Contiguity 旁的单选按钮，如图 15.12 所示，点击 Create。与前面相同的进度条会出现（图 15.5）。点击 Done 完成处理过程。

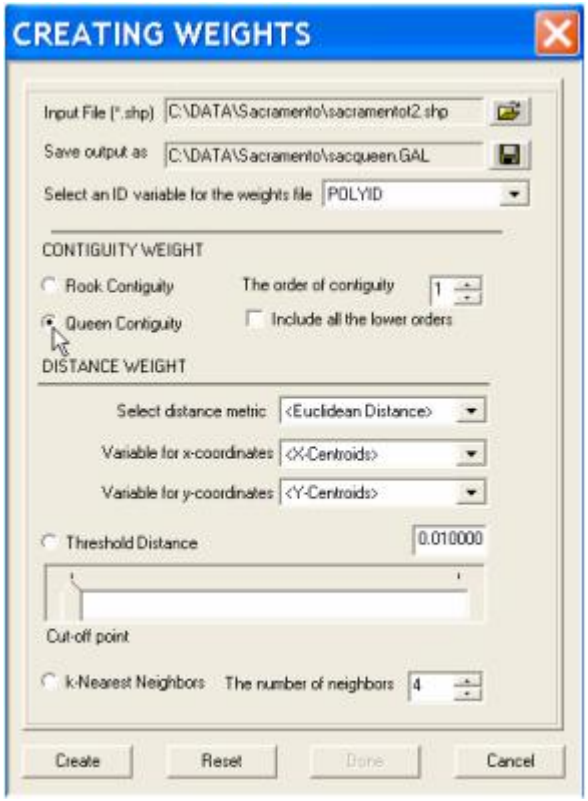


图 15.12　Queen 邻接性

比较 Sacramento 数据的 Rook 和 Queen 标准的邻接性结构。这两个直方图显示于 15.13。单击 rook 标准中 5 个邻居相对应的条块，注意 queen 标准的分布有 5 个或更多的邻居。在地图中查看选择来查找有不同的地区。

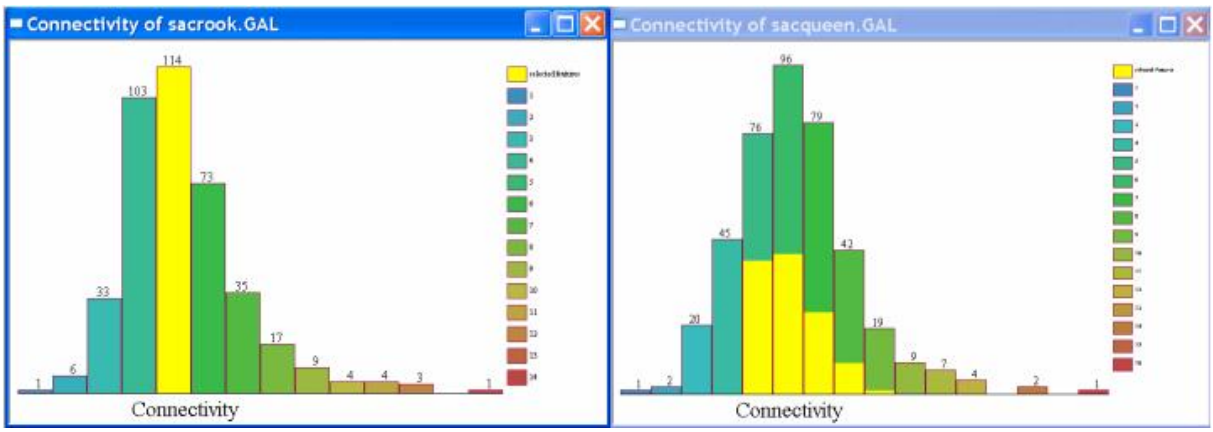


图 15.13　rook 和 queen 邻接性的连接结构比较

15.5 高次邻接性

以相同的方式来创建高次邻接性空间权重文件。权重创建对话框是相同的。选择 rook 或 queen 旁的单选按钮和邻接性的次数。例如，在图 15.14 中，选择 rook 标准 2 次邻接性。

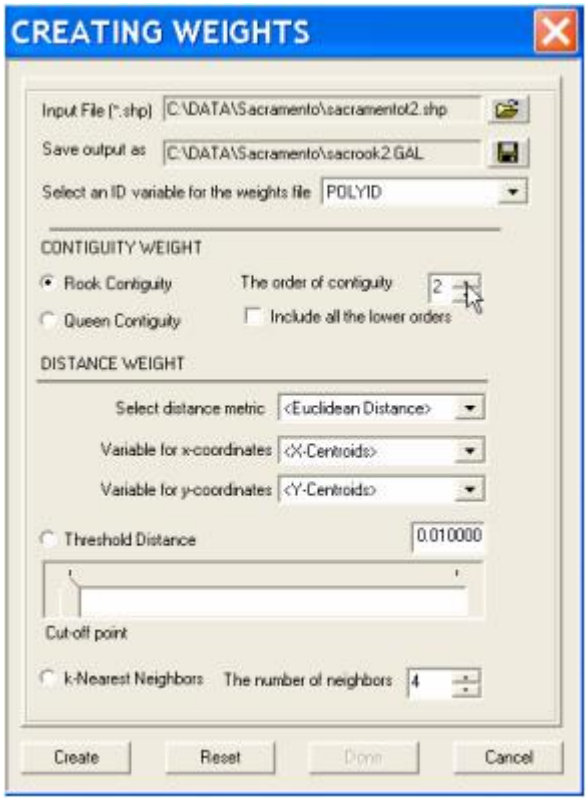


图 15.14　二次 rook 邻接性

注意邻接性次数下面的选择框。GeoDa 允许两种高次邻接性定义。一是“pure”，不包括低次时的邻接性位置（这是高次邻接性的书面定义）。另一个是累积（cumulative），包括所有低次邻居。试着用这一功能为 Sacramento 创建一幅 pure 和 cumulative 二次权重文件。比较这两个的连接结构。图 15.15 和 15.16 显示了这两种情况下二次 rook 邻接性的连接性直方图，选中如前的观测点（5 个一次 rook 邻居）。

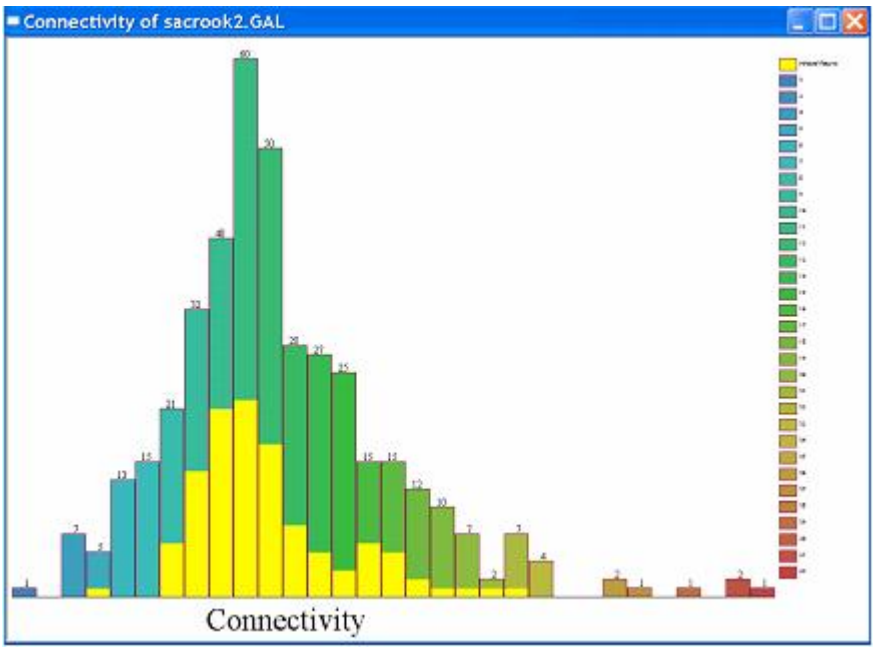


图 15.15　Pure 二次 rook 连接性直方图

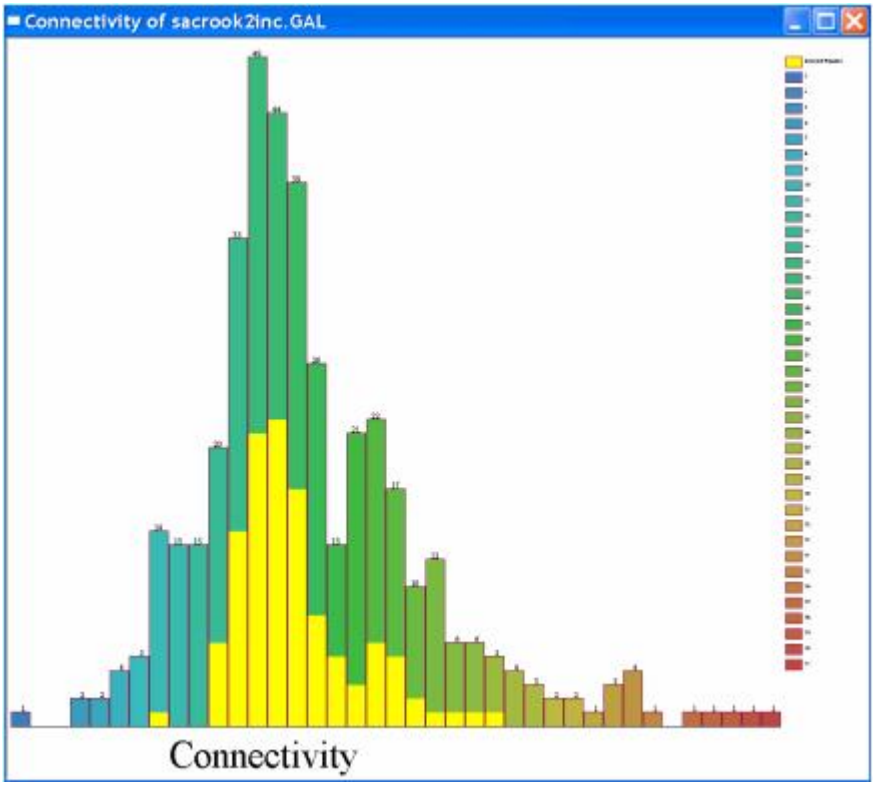


图 15.16　Cumulative 二次 rook 连接性直方图

15.6 练习

在样本数据集中利用任何的多边形文件练习创建 rook 和 queen 邻接性权重和高次权重。这些操作将会在空间自相关和空间回归模型估计中反复用到。也创建一些更高次的邻接性权重。查看连接结构，使用链接功能查找地图中被选中位置的邻居数量。

练习 16　基于距离的空间权重

16.1 目标

本练习显示了基于距离的权重的建立方法，邻居的定义是基于点或多边形质心的距离。

在本练习末，你将知道：

- 指定距离，由点文件创建基于距离束（distance band）的空间权重文件。
- 调整距离
- 创建基于 K 最近邻居标准的空间权重文件

有关本操作的详细信息请参考 Users Guide 第 83—85 页。和 Release Notes 第 18—19 页。

16.2 距离束权重

打开 Boston 房屋样本数据（输入 boston.shp，关键字为 ID），载入 506 个人口区的质心文件开始这一练习。底图如图 16.1 所示。对于邻接权重，建立基于距离权重矩阵可以在不打开项目的前提下进行，可以直接利用 Tools 菜单。在一个项目中，也可能使用两样（图 15.2）的方法，或点击工具条中相应的按钮。

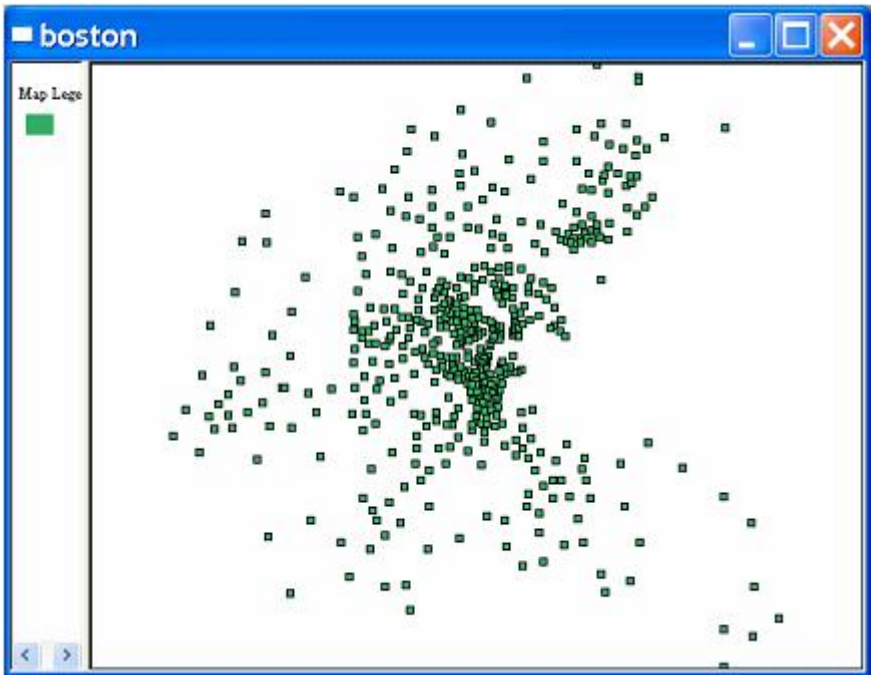


图 16.1　Boston 人口区质心数据底图

选择 Tools>Weight>Create 来打开权重创建对话框，如图 16.2。输入文件为 boston.shp，空间权重文件名为 bostondist(程序会打开一个 GWT 文件扩展)，指定 ID 为 ID 变量。然后，到关于 Distance Weight 对话框部分。保留默认<Euclidean Distance>，因为 Boston 数据是 UTM 投影的坐标。如果点以经度和纬度表示，你需要选择<Arc Distance>选项。

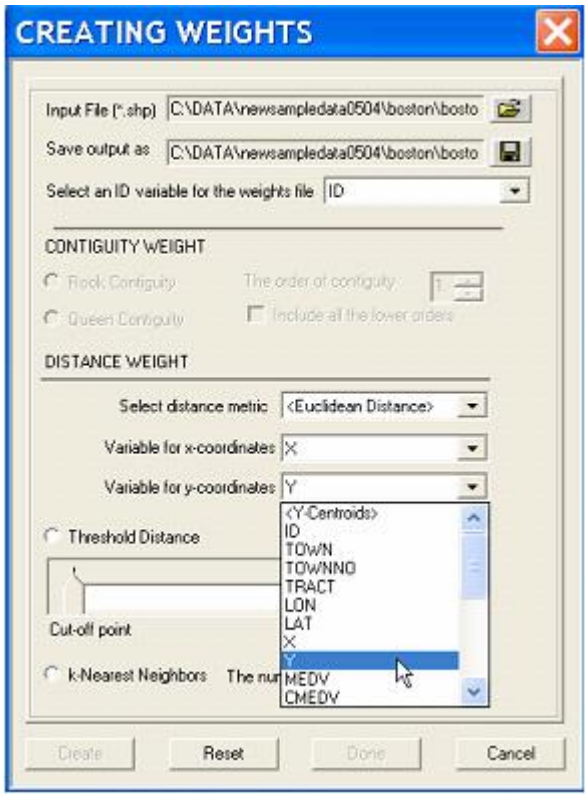


图 16.2　距离权重对话框

第二，指定 X 和 Y 坐标的变量名分别为 X 和 Y，如图 16.2 所示。注意，相对于纯（pure）邻接权重，基于距离的空间权重可以计算点文件和多边形文件的距离。对于后者，如果不指定坐标变量，会计算多边形的质心，用于距离的计算。

选中 Threshold distance 旁的单选按钮，如图 16.3。注意文本框中的值变为 3.972568。这是保证每一位置都至少有一个邻居的门槛距离。如果设定门槛距离小于此值，会出现“岛”。特别地，一些实验（和连接结构检查）需要设定比最小门槛大些的值。

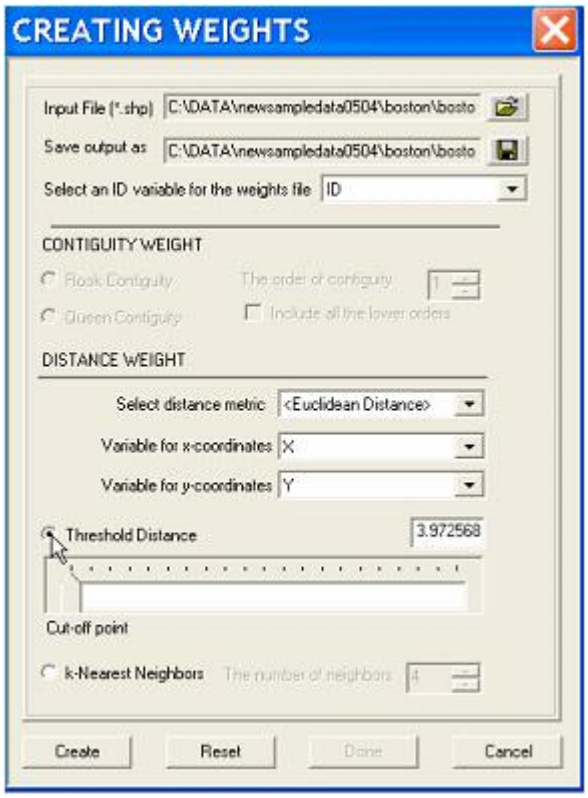


图 16.3　指定门槛（Threshold）距离

点击 Create 开始创建。会出现一个进度条，如图 16.4。点击 Done 返回主界面（standard interface）。

图 16.4　创建 GWT 文件

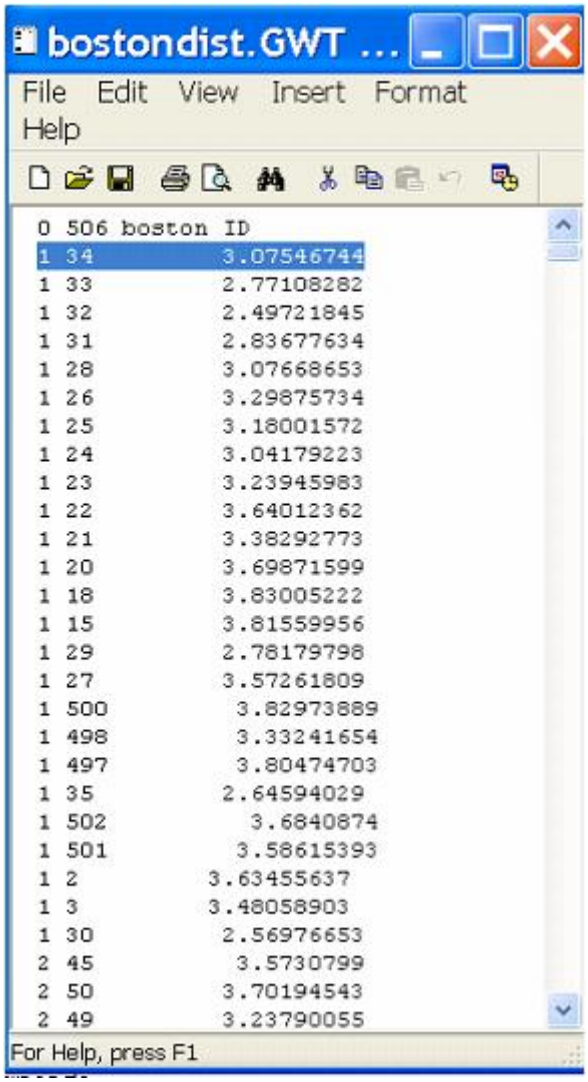


图 16.5　GWT 文件内容

包含空间权重信息的 GWT 文件是一个标准的文本文件，如前面提到的 GAL 文件。格式稍有差别，如图 16.5。利用文件编辑器，打开你刚刚建立的 bostondist.gwt 文件，查看其内容。如图 15.6，每一行为头行，包括了如 GAL 相同的信息（地方（place holder），观测点数目，文件名及 ID 变量）。文件剩余的部分包括每一个邻居对，“原始”ID，“目标”ID 和两个点之间的“距离”。注意第三个值并不是空间权重值。在当前版本中，GeoDa 并不产生权重值，只考虑邻居是否存在邻居关系。

用 15.3 部分提到的方法查看 bostondist.gwt 的连接结构。结果如图 16.6。注意，相对于基于邻接的权重，距离有一个很宽的范围。实际上，当点的分布是不规则分布（如一些点集聚在一起，而其它的距离很远）时，这种基于距离的权重是很典型的。在这种情况下，数据集中的为避免“岛”的最小门槛距离对多数位置而言可能是太大了。这些门槛距离是由最远的点对所决定的，并不能代表其它点的分布。在这种情况下，在解释距离门槛时要小心谨慎，使用 k-nearest 邻近权重也许会更合适。

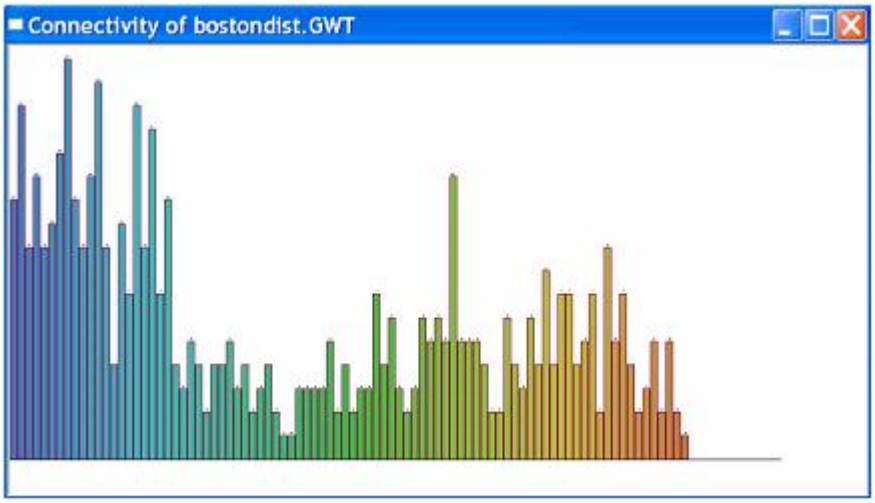


图 16.6　基于距离权重连接性

16.3 k-nearest 邻近权重

用与其它权重相同的方法（Tools>Weight>Create）开始创建一个 k-nearest 邻近权重文件。产生一个权重对话框，如图 16.7 所示。在对话框中，输入文件为 boston.shp，输出文件为 bosotonk6，ID 变量为 ID。向下看对话框，选中 k-nearest neighbors，如图 16.7。将邻居数量改为 6，点击 Create 来创建权重。当进度条显示结束时，点击 Done 按钮返回主界面。

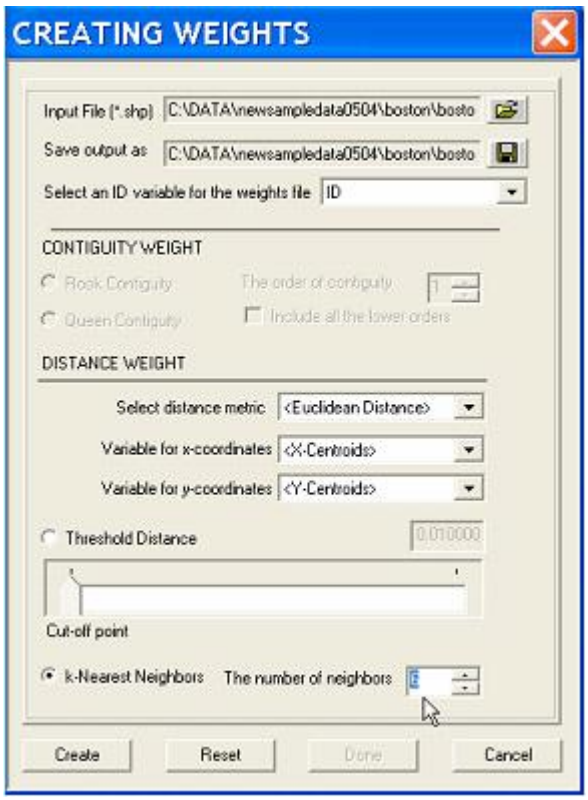


图 16.7　最近邻权重对话框

K 最近邻标准保证每一个观测点有相同的邻居数量。检查你刚刚创建的 GWT 文件，查看这一情况。或者相查看权重属性。连接性直方图并没有太大的意义，如图 16.8，但它显示每个位置有 6 个邻居。这不是很有用，但保证邻居数量的正确性是有用的，因为 GeoDa 中没有空间权重文件的元数据。

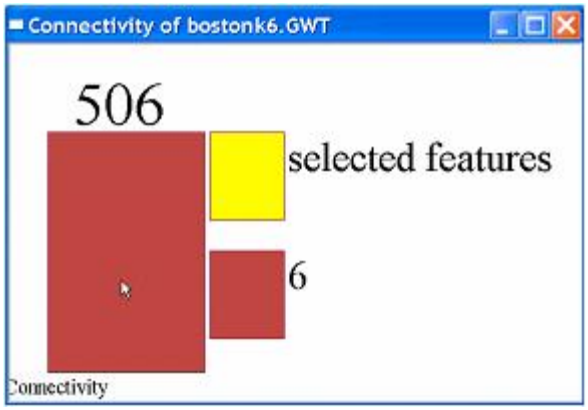


图 16.8　最近邻连接属性

16.4 练习

利用其它数据集中的点文件练习创建基于距离的空间权重，如 oz9799.shp（关键字为 STATION）的 30 个空气质量监测点，或 baltimore.shp 点文件（关键字为 STATION）中 211 个房屋销售。或者，查看多边形文件的默认特征。例如，利用 ohlung.shp 多边形文件中 Ohio 的 88 个县（关键字为 FIPSNO）来创建距离束和 K 最近邻权重。

在后面的练习中，你将会需要空间权重作为进行空间自相关分析的重要输入文件，所以你现在创建的任一文件，在需要时不需要再重新创建，只需要“打开”。

练习 17　空间滞后变量

17.1 目标

本练习显示建立空间滞后变量及它在 Moran 散点图中的用处。

在本练习末，你将知道：

- 为指定的权重文件创建一个空间滞后变量
- 用空间滞后来“手工”建立一个 Moran 散点图

17.2 建立空间滞后

空间滞后变量是空间自相关计算和空间回归模型解释的重要部分。迅速计算空间滞后变量是 GeoDa 的特色，但在一些情况下，明确地计算空间滞后变量是很有用的。例如，如果想要在其它统计软件包中使用这些变量，将是很容易的。空间滞后计算是 GeoDa 中 Table 功能的一部分（见练习 3，特别是 3.4 部分）。载入 Sacramento，CA 的 403 个人口区的样本数据（sacramentot2.shp，关键字为 POLYID）。底图如图 15.1 所示。

在开始 Table 操作之前，确保有一个空间权重文件已经被打开。如果当前没有这样的文件，空间滞后计算会产生一个错误消息。从菜单打开权重，用 Tools>Weights>Open，如图 17.1，或在相应的工具条按钮上点击。在权重对话框中指定 sacrook.GAL 作为文件名，如图 17.2 所示。

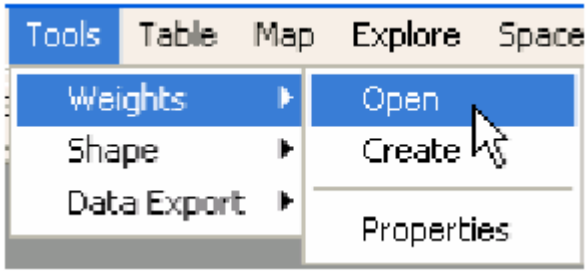


图 17.1　打开权重文件

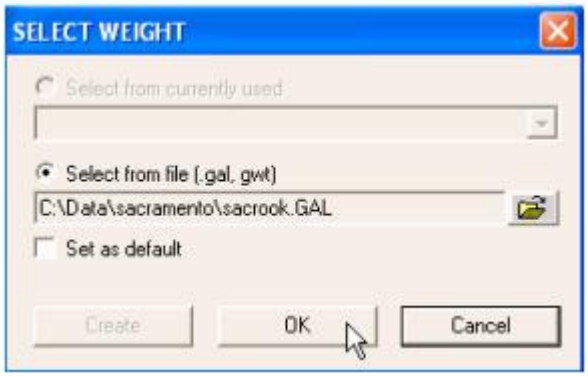


图 17.2　选择空间权重文件

开始滞后计算的一切已经准备就绪。打开数据表（点击 Table 工具按钮），右击从菜单中选择 Field Calculations（图 17.3）。然后在 Field Calculation 对话框中选择 Lag Operations 标签，

如图 17.4。在最左边的文件框中输入 W_INC，如图 17.5，保留默认的权重文件，选择 HH__INC（人口区中等家庭收入）为滞后变量。点击 OK 计算新变量。新的一栏会加入到数据表中，如图 17.6 所示。

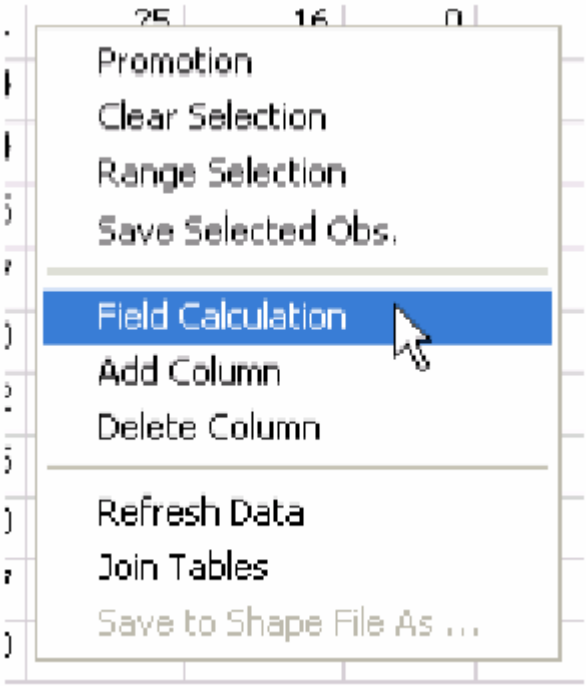


图 17.3　数据表计算选项

图 17.4　表格中的空间滞后计算标签

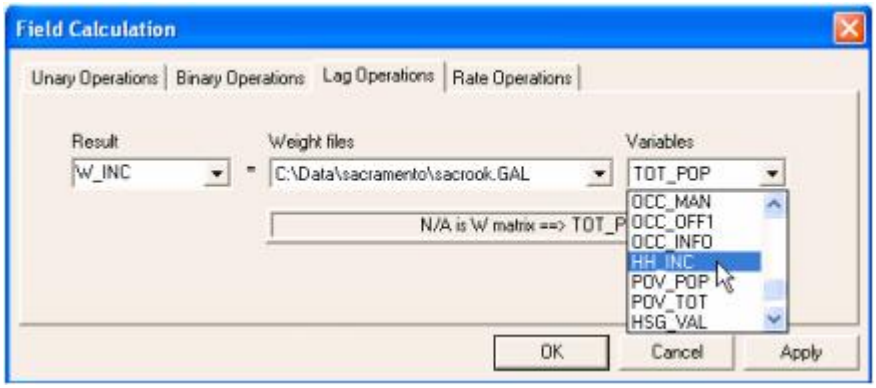
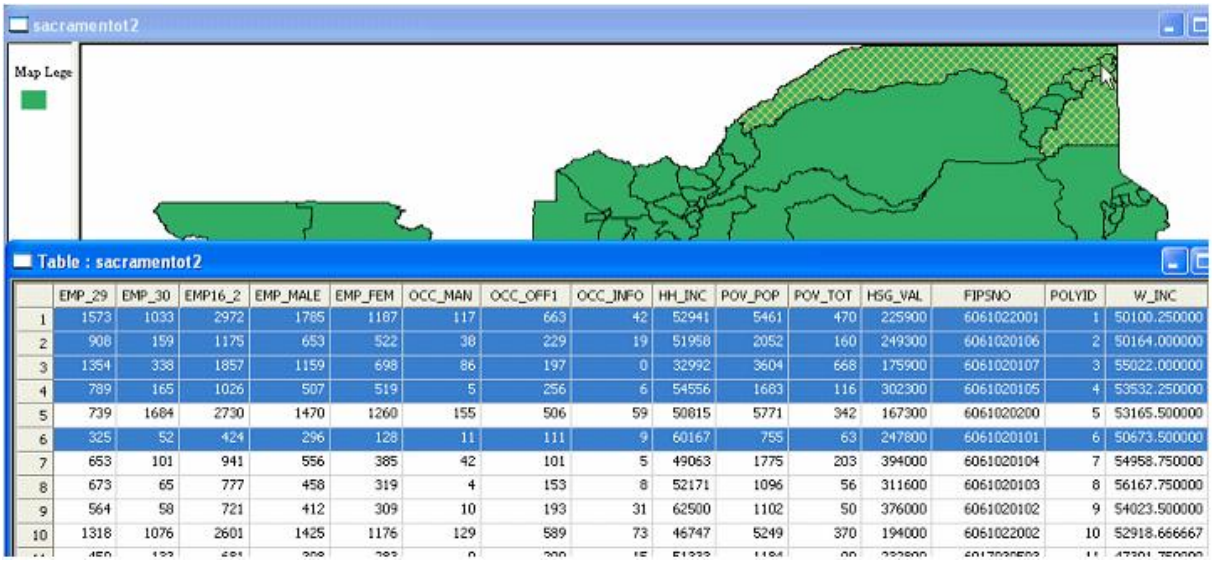


图 17.5　Sacramento 地区家庭收入空间滞后对话框



	EMP_29	EMP_30	EMP16_2	EMP_MALE	EMP_FEM	OCC_MAN	OCC_OFF1	OCC_INFO	HH_INC	POV_POP	POV_TOT	HSG_VAL	FIPSNO	POLYID	W_INC
1	1573	1033	2972	1785	1187	117	663	42	52941	5461	470	225900	6061022001	1	50100.250000
2	908	159	1175	653	522	38	229	19	51958	2052	160	249300	6061020106	2	50164.000000
3	1354	338	1857	1159	698	86	197	0	32992	3604	668	175900	6061020107	3	55022.000000
4	789	165	1026	507	519	5	256	6	54556	1683	116	302300	6061020105	4	53532.250000
5	739	1684	2730	1470	1260	155	506	59	50815	5771	342	167300	6061020200	5	53165.500000
6	325	52	424	296	128	11	111	9	60167	755	63	247800	6061020101	6	50673.500000
7	653	101	941	556	385	42	101	5	49063	1775	203	394000	6061020104	7	54958.750000
8	673	65	777	458	319	4	153	8	52171	1096	56	311600	6061020103	8	56167.750000
9	564	58	721	412	309	10	193	31	62500	1102	50	376000	6061020102	9	54023.500000
10	1318	1076	2601	1425	1176	129	589	73	46747	5249	370	194000	6061022002	10	52918.666667

图 17.6　加入到数据表中的空间滞后变量

回想图 15.7 中 POLYID2 有 4 个邻居。地图和表格中的相关区域被高亮显示，如图 17.6。

对于邻接权重文件，如 sacrook.GAL，空间滞后变量等于相邻单元的简单平均值。在图 17.6 中查看第 2 行的 W__INC 值（50164）是否是第 1，3，4，6 行 HH__INC 的平均值。

17.3 空间自相关

Moran 散点图是一幅兴趣变量为 X 轴，空间滞后为 Y 轴（详见 18.2.2 部分）的图。因为刚刚计算的空间滞后可以立即用于任何分析，所以你可以在规则散点图中，用空间滞后 W__INC 和 X 轴变量 HH__INC 手工建立一幅 Moran 散点图（8.2 部分）。

从菜单中选择 Explore>Scatter plot 开始散点图功能，或点击相应的工具条按钮。在变量选择对话框，在左边栏中指定 W__INC，在右边栏中指定 HH__INC，如图 17.7。然后点击 OK 产生散点图，如图 17.8。

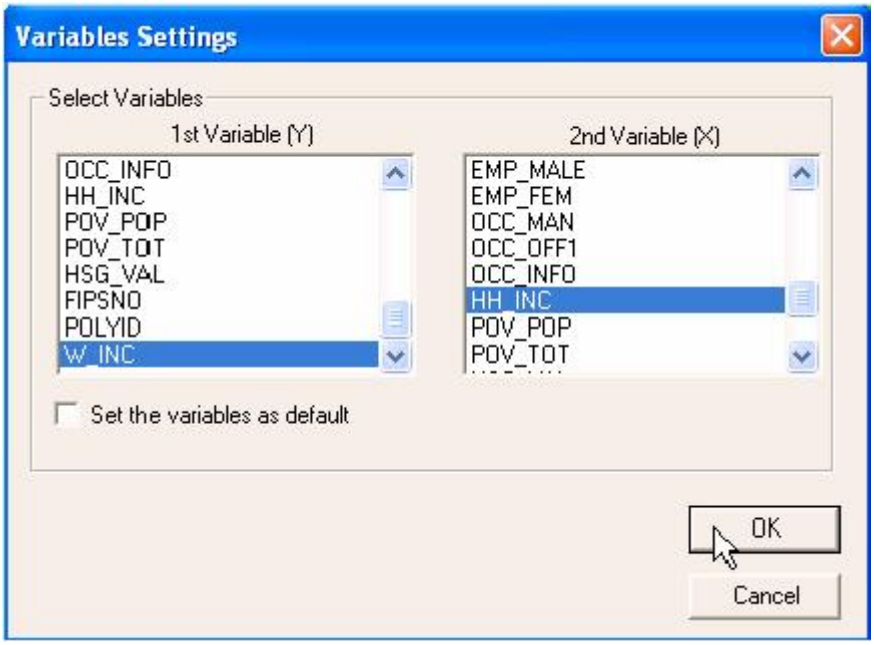


图 17.7　收入的空间滞后与收入的变量选择

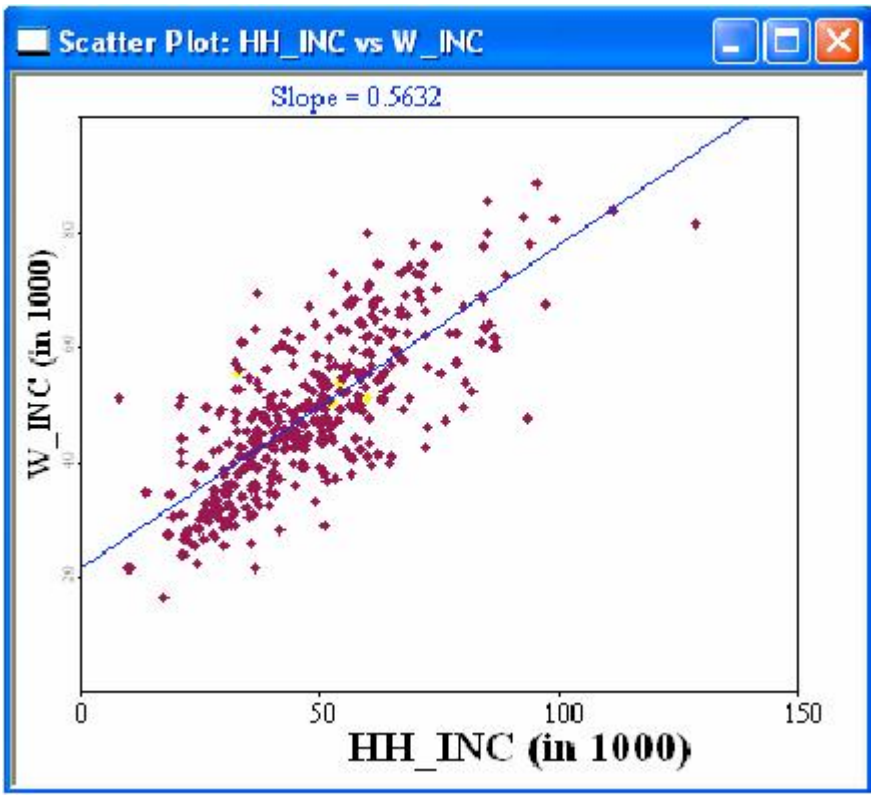


图 17.8　作为规则散点图创建的 Moran 散点图

回归直线的斜率是 HH__INC 以 rook 邻接权重定义计算的 Moran's I 的统计量（statistic）。自由提前运行，按照 18.2.2 的指导（但替换为相关的 Sacramento 数据和权重）查看这一情况属实。当然，因为图 17.8 并不是一个非空间散点图，它不包括评估 Moran's 意义的方法。

17.4 练习

用练习 15 和 16 中建立的空间权重来创建空间滞后变量及手工建立 Moran 散点图。

练习 18　全局空间自相关

18.1 目标

本练习开始演示单变量和 Moran 散点图空间自相关分析。关于这些方法的方法背景，见 Anselin(1995,1996)。

在本练习末，你将知道：

- 创建一个单变量空间自相关散点图
- 通过序列检验进行显著性评估
- 创建显著性信封（envelope）
- 刷光 Moran 散点图
- 保存空间滞后和标准变量

有关本操作的更详细信息请参考 User's Guide。

18.2 Moran 散点图

18.2.1 准备工作

我们将要对 56 个 Scottish 地区的唇癌数据进行操作。载入文件 scotlip.shp，关键字为 CODENO。结果如图 18.1 所示。

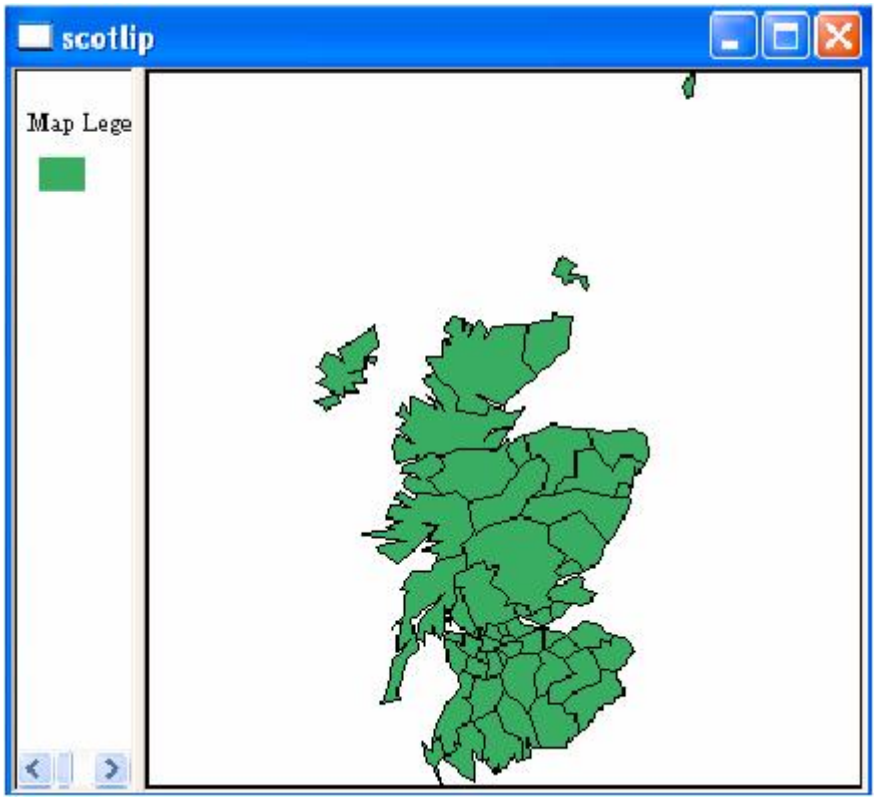


图 18.1 Scottish 唇癌数据的底图

为能够比较“原始”数据的散点图和练习 20 中 EB 标准比率散点图，要确保在数据集中有原始数据的变量。如果你以前没有计算过，一个简单的方法就是创建一幅箱地图，在菜单中用 Map>Smooth>Raw Rate。选择 Cancer 为事件，Pop 为基数变量，如图 18.2 所示。确保设定地图类型为箱地图。点击 OK，产生地图如图 18.3 所示。

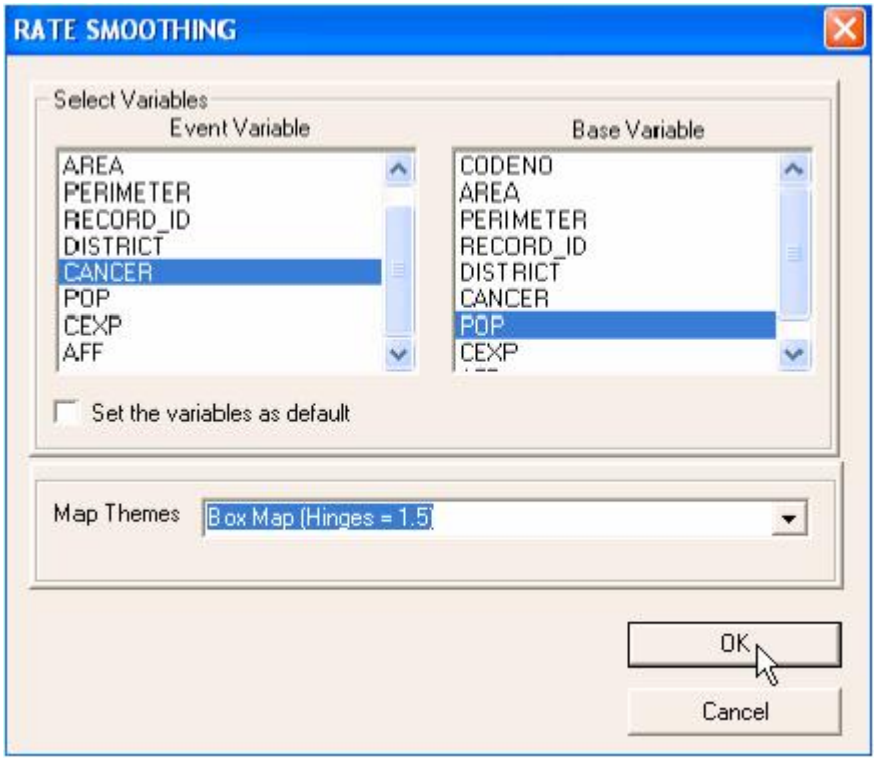


图 18.2 分地区 Scottish 唇癌原始比率计算

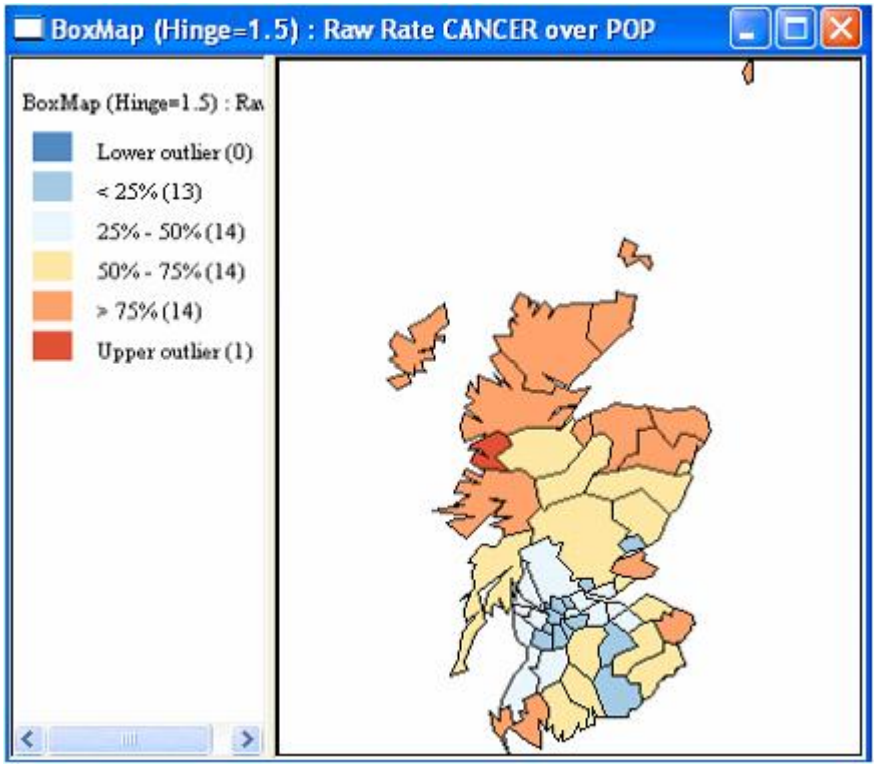


图 18.3 分地区 Scottish 唇癌原始比率箱地图

在箱地图中右击，选择 Save Rates(见图 13.6)，将原始比率添加到数据表中。为简便起见，保留变量名为默认的 R_RAWRATE。

最后，确保有一个 Scottish 地区文件的空间权重文件。在本例中，我们将使用 5 最近邻。如果你在前面的练习中没有创建一文件，就创建这样一个文件，文件名为 scot5k.GWT。

18.2.2 Moran 散点图功能

当所有准备工作完成后，点击工具条按钮，或从菜单选择 Space>Univariate Moran(图 18.4)打开 Moran 散点图功能。将产生一个变量选择对话框，如图 18.5。

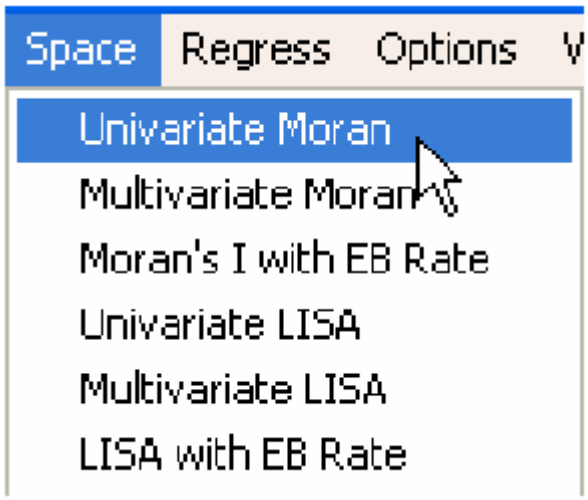


图 18.4 单变量 Moran 散点图功能

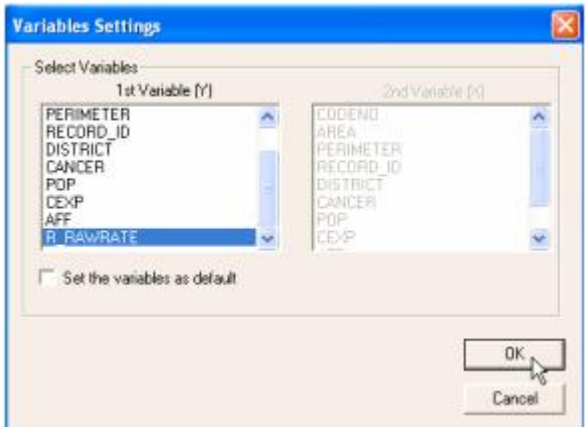


图 18.5 单变量 Moran 散点图变量选择对话框

选择 R_RAWRATE 作为变量，点击 OK。然后选择 scot5k.GWT 作为权重文件。在图 18.6 中，显示了空间权重文件中需要为被指定的情况(如从文件选择)。如果这个文件已经被打开，对话框会稍有不同(例子见图 20.3)。点击 OK 创建 Moran 散点图，如图 18.7 所示。

图 18.6 单变量 Moran 散点图空间权重选择对话框

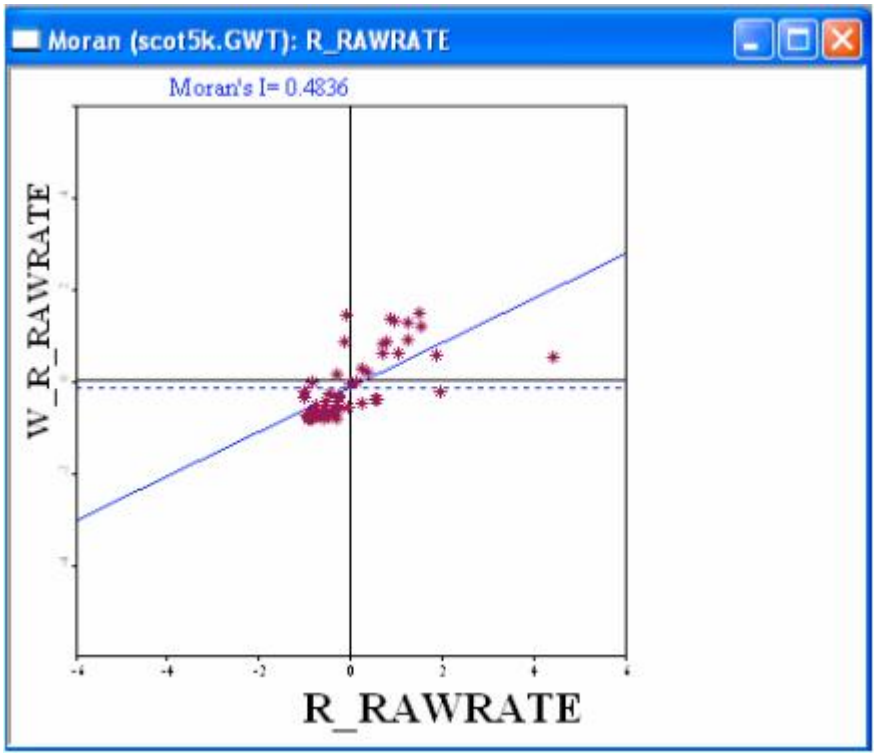


图 18.7 Scottish 唇癌比率 Moran 散点图

注意：Y 轴被指定为 W_R_RAWRATE，而没有必要进行明确的空间滞后的计算。R_RAWRATE 在 X 轴，它已经相对于标准差(超过 2 倍标准差即被认定为离群值)进行了标准化。散点图以平均值为轴的中心，将图分为 4 个象限。每个象限对应于不同的空间自相关类型：高高和低低为正相关；低高和高低为负相关。利用选择工具每一种空间自相关类型对应的空间位置(通过与地图的链接)。

图的顶部所列的值 0.4836 是 Moran's I 统计量。因为此图是一种特殊的散点图，可以应用 Exclude Selected。试验这一选项(在图中右击或从 Options 菜单中调用此选项)，评估当排除指定观测点时，空间自变量相关系数(回归直线的斜率)的变化。相似地，你可以象在其它散点图中以同样的方式刷光 Moran 散点图。

用于创建散点图的中间计算可以保存于当前数据表中。在图中右击，产生 Optoins 菜单，如图 18.8，选择 Save Results 会产生变量选择对话框，如图 18.9。这一次，你不能保留默认的变量名，因为 GeoDa 只支持长度不超过 12 个字符的变量名。编辑变量名(如使用 W_RAWRATE)，将 R_RAWRATE 的标准值及相应的空间滞后加入到表格中。

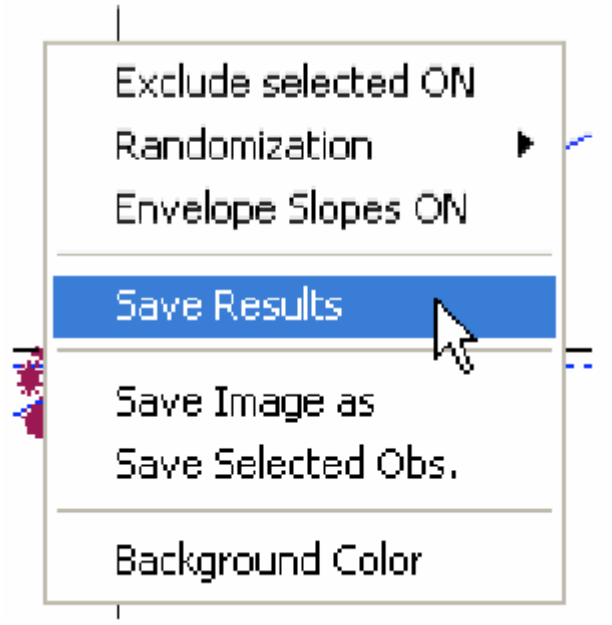


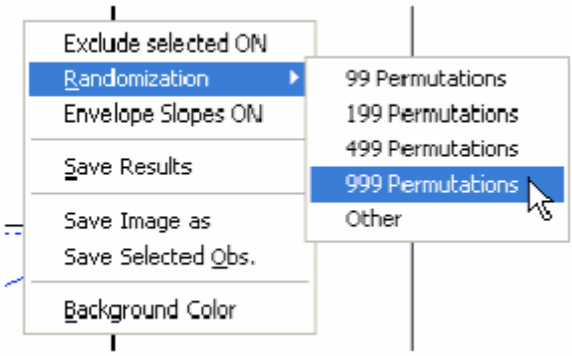
图 18.8 Moran 散点图保存结果选项

图 18.9 在 Moran 散点图中保存结果的变量对话框

18.3 推断

对 Moran's I 的推断是基于随机序列，它多次重新计算统计量产生一个参考分布。得到的统计量与参考分析相比较，计算一个假设显著性。

在散点图中右击调用选项菜单来开始推断计算，如图 18.10。选择 Randomization>999 permutations，产生直方图，如图 18.11 所示。



18.10 Moran 散点图中随机对话框

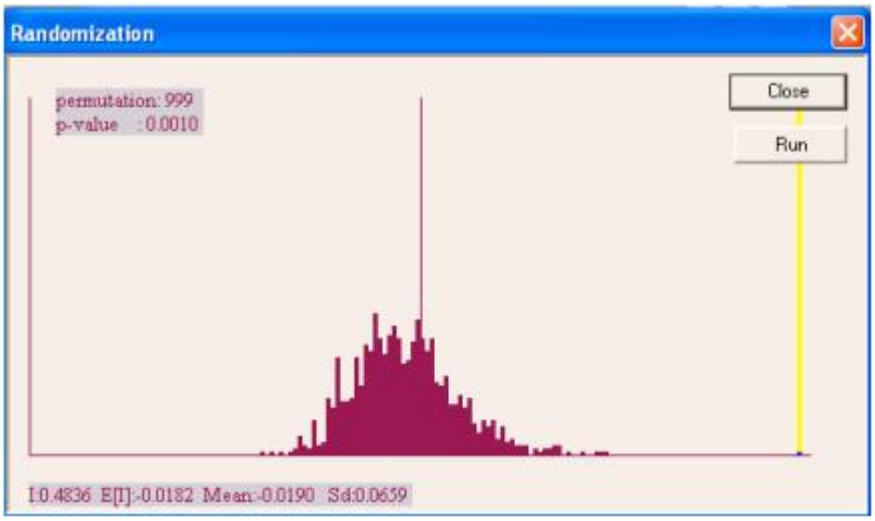


图 18.11 Moran's I 序列经验分布

除了参考分布(棕色)和统计量(黄色条)，图中左上角还列出了序列数量和假设显著性水平，统计值为 0.4836，理论平均值为 E[I]=-0.0182，经验分布的平均值和标准差。在图 18.11 中，这些值分别是-0.0191 和 0.0659(这些值取决于特定随机序列，序列之间稍有不同)。点击 Run 按钮，评估结果对特定随机序列的敏感性。特别地，对 999 序列，这些结果不会有太大的变化，但对数字较小的序列，如 99，将会有非常大的差别。注意：最显著的 P-level 直接取决于序列的数目。例如，对 99 序列，p=0.01，对 999 序列，p=0.001。

将 Moran's I 统计量的显著性可视化的不同方法是在图中绘出随机信封(envelope)。这些斜率对应于 2.5%到 97.5%的参考分布，包括了在空间随机数据中计算的 Moran's I 统计量分布的 95%。

在图中右击，选择打开选项(Exclude Selected，这是一个形状选项)，如图 18.12 所示。随后，图中会出现两个破折线，如图 18.13 所示。注意：真正的 Moran 散点图斜率正好位于随机转置(permute)序列范围之外。

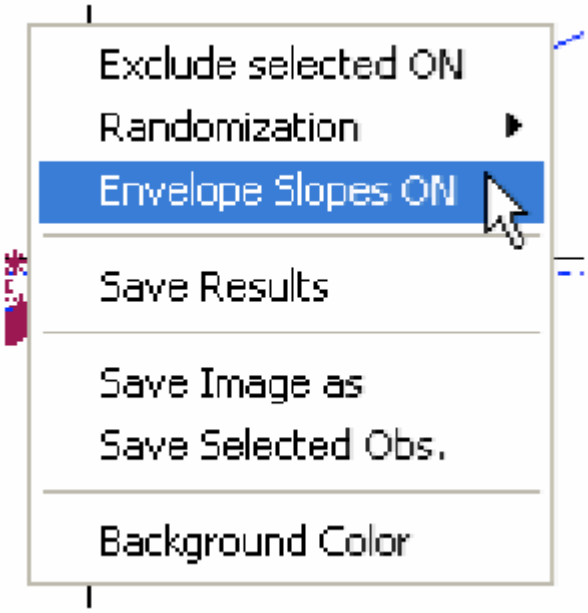


图 18.12 Moran 散点图的信封(envelope)斜率

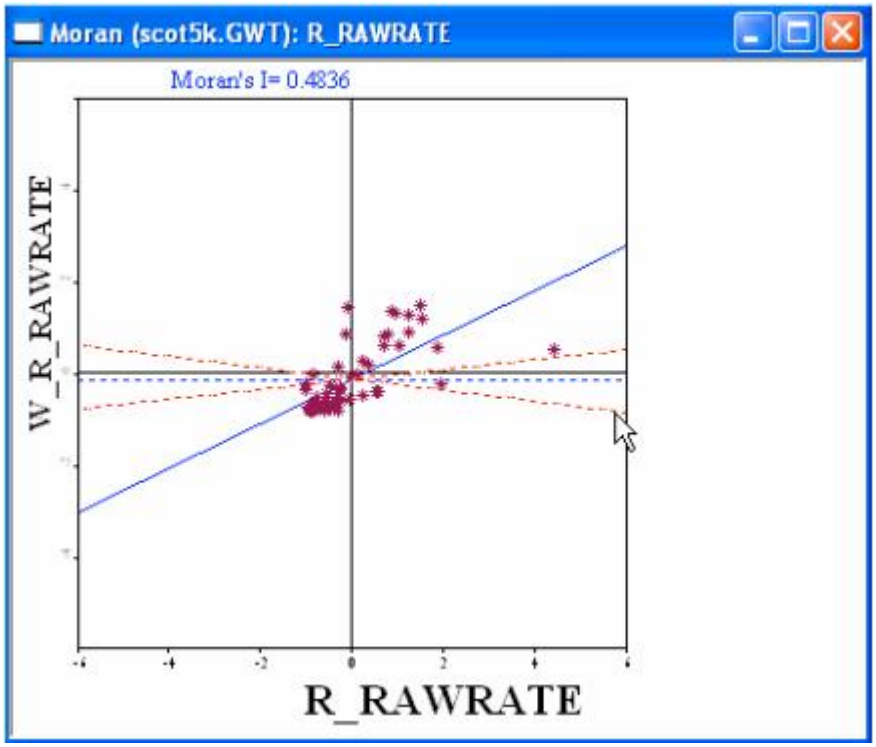


图 18.13 加入到 Moran 散点图的信封(envelope)斜率

18.4 练习

一些样本数据适合于进行空间自相关的比率分析，允许(allowing)在练习 20 中原始比率与 EB 标准化比率进行比较。包括 ohlung.shp 中 Ohio 州 88 个县的肺癌数据(关键字为 FIPSNO)，sids.shp 中 North Carolina 州 100 个县的 SIDS 死亡率(关键字为 FIPSNO)，都是经典的公共健康方面的应用。Buenosaires.shp 中 Buenos Aires 城的 209 个选区(关键字为 INDRANO)是一个政治科学的下例子，sto_hom 中 78 个县的杀人案数据(关键字为 FIPSNO)显示了在犯罪学方面的应用。在每一个这样的例子中，可以分析不同的变量，评估空间自相关的敏感性选择空间权重。

练习 19　局部空间自相关

19.1 目标

本练习关注局部空间自相关的定义,特别是局部 Moran 统计。其方法背景可参考 Anselin(1995)。在本练习末，你将知道：

- 计算局部 Moran 统计和相关的显著性地图及丛聚地图
- 评估丛聚地图对序列数字及显著性水平的敏感性
- 解释空间聚集和空间离群值的定义

有关本操作的详细内容请参考 User's Guide 第 99-105 页。

19.2 LISA 地图

19.2.1 基础

为显示局部空间自相关功能，载入包含 78 个县杀人案数据的文件 St Louis，OM(stl_hom.shp，关键字为 FIPSNO)。底图如图 19.1。你将会需要一个空间权重文件。如果你还没有制作权重文件，在开始分析前创建一个 rook 权重文件(取名为 stlrook.GAL)。从菜单选择 Space>Univariate LISA(如图 19.2)，或点击相应的工具按钮，开始 Local Moran 功能。将会出现一个熟悉的变量选择对话框。为变量选择 HR8893，如图 19.3。然后选择 OK，将出现一个权重选择对话框。指定权重文件 stlrook.GAL，如图 19.4，继续点击 OK。最后一步，出现结果选项对话框，如图 19.5 所示。

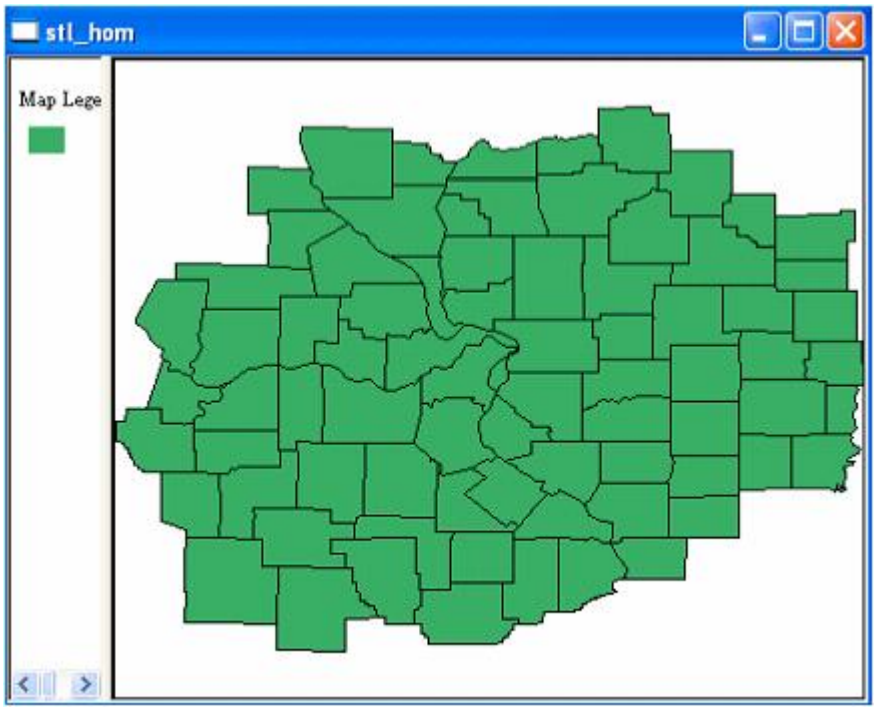


图 19.1 St Louis 区域各县杀人案底图

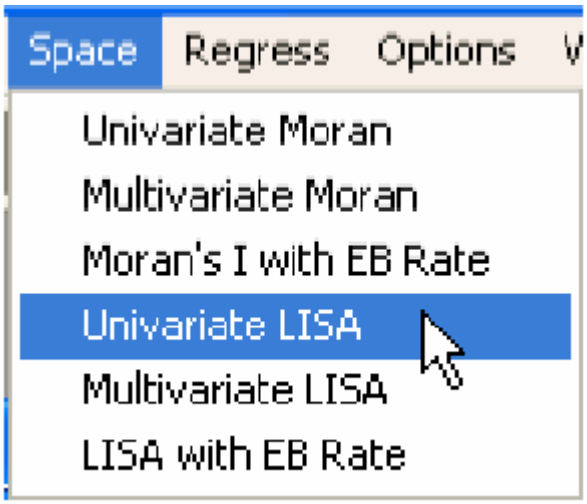


图 19.2 局部空间自相关功能

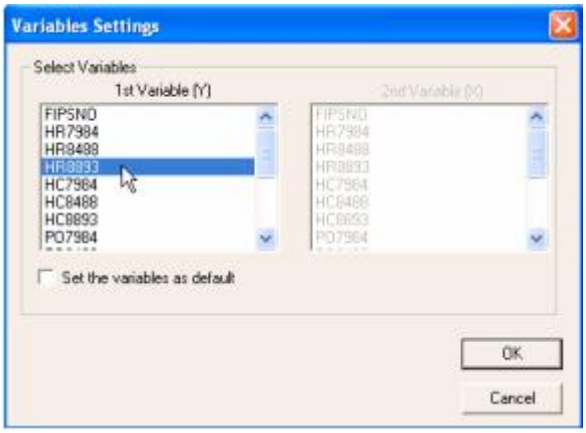


图 19.3 局部空间自相关变量选择对话框

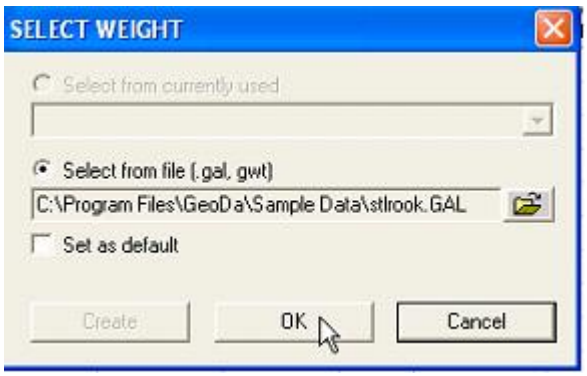


图 19.4 局部空间自相关空间权重选择

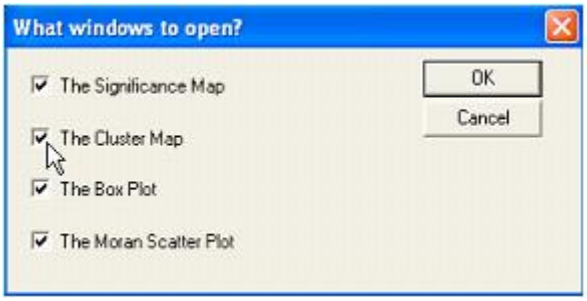


图 19.5 LISA 结果选项窗口

得到 4 个不同类型的图表和地图：一幅显著性地图，一幅聚集地图，一幅箱图，一幅 Moran 散点图。现在，选中所有的 4 个选项，如图 19.5，但这并没有必要。点击 OK 产生 4 个图表。

19.2.2 LISA 显著性地图

如图 19.6 所示的显著性地图用不同的绿色阴影(在图例中给出了相应的 p 值)显示了显著的局部 Moran 统计的位置。你的结果可以会稍有不同，因为第一幅图是根据 99 序列的结果(这里显示的地图是 9999 序列的运行结果，避免对特定随机性的太多的敏感性)。我们转到 19.3 部分的打断。注意结果 p=0.05 好象有些不太可信。因为它们似乎忽略了与多元比较相关的问题(作为结果，真实的 p 值可能高于 0.05)。

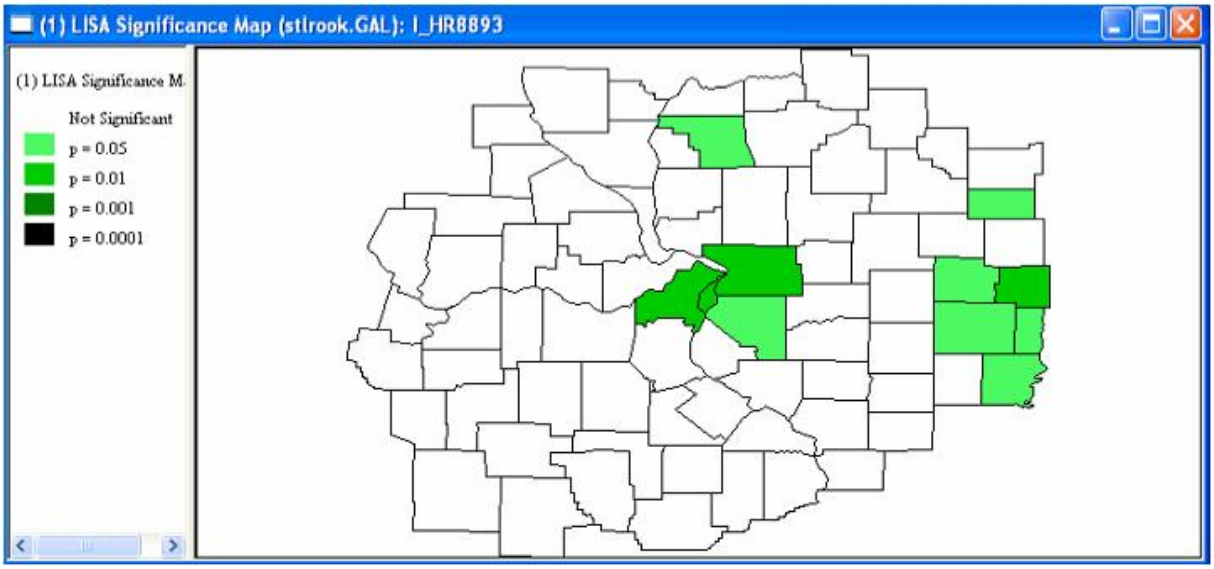


图 19.6 St Louis 区域杀人案发生率 LISA 显著性地图

19.2.3 LISA 聚集地图

按理说，最有用的图表是所谓的 LISA 聚集地图，如图 19.7 所示。它在本质上提供与显著性地图相同的信息(为什么是相同的？)(实际上，这两幅地图是同步的)，但用不同的颜色表示不同的空间自相关类型。这 4 种颜色如图中图例所示：深红色表示高高，深绿色表示低低，品红色表示高低，浅绿色表示低高(样图中没有低高的区域)。这四种类型对应于 Moran 散点图中的 4 个象限。选择相同颜色的县，查看它们在 Moran 散点图中的位置。

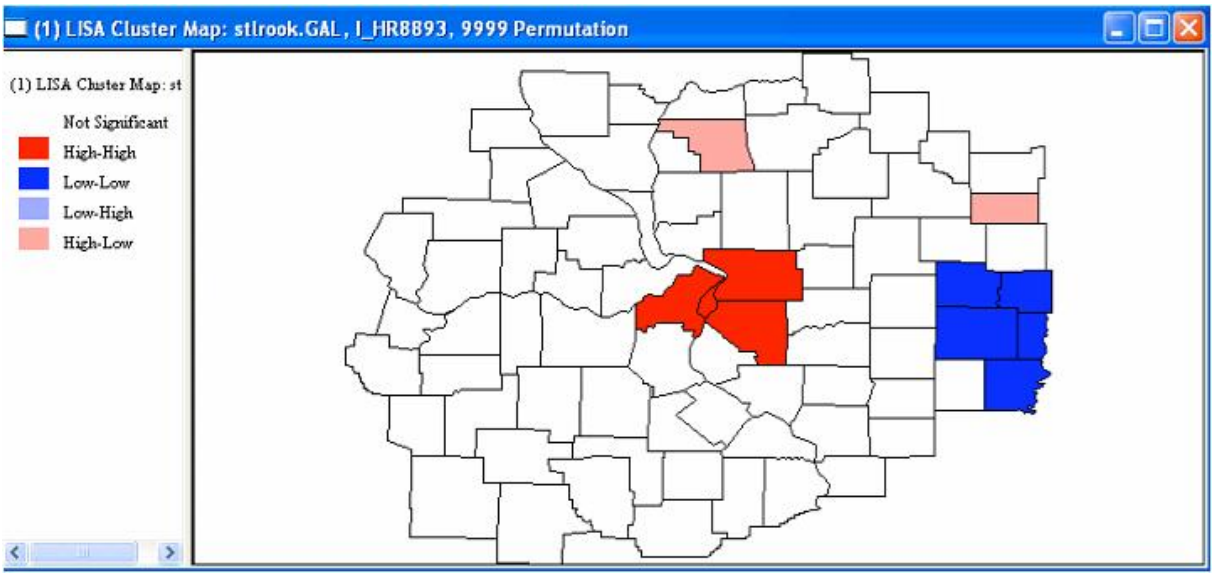


图 19.7 St Louis 区域杀人案发生率 LISA 聚集地图

19.2.4 其它 LISA 结果图表

另一个结果是所有观测点的局部 Moran 统计分布箱地图。这主要是专业兴趣(technical interest)，暗示具有不同局部自相关模式的潜在位置。

全局 Moran's I 统计是局部 Moran 统计的平均值。所示，如果这些局部统计的分布高度不均匀，或被一些少数很大的值所控制(如图 19.8)，总体显示会具有欺骗性，或对一些观测点非常敏感。从高值到低值刷光箱地图，在地图中定位这些相应的县。注意局部 Moran 正值可能与高高及低低模式相对应。

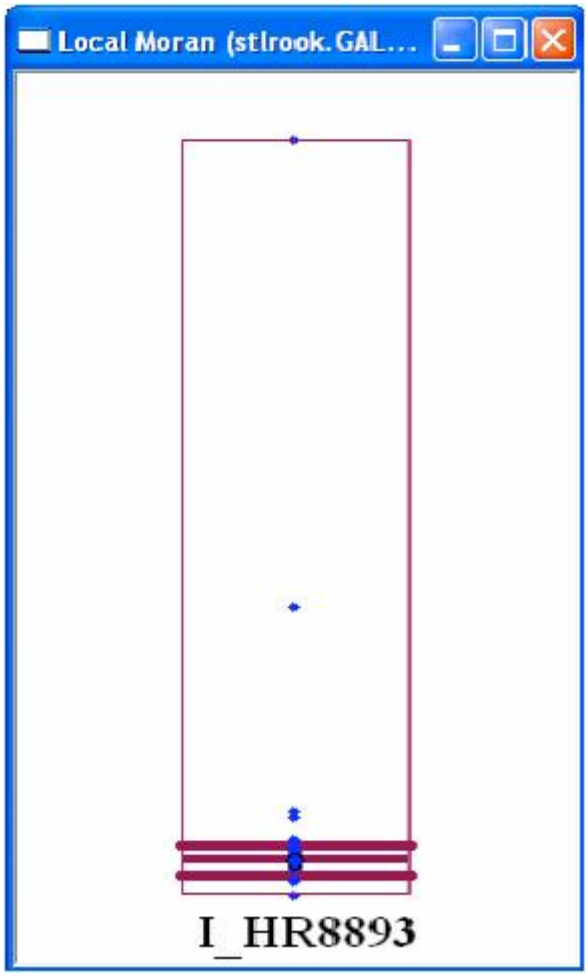


图 19.8 LISA 箱地图

最后的一幅结果图是 Moran 散点图，图 19.9 所示(详见练习 18)。

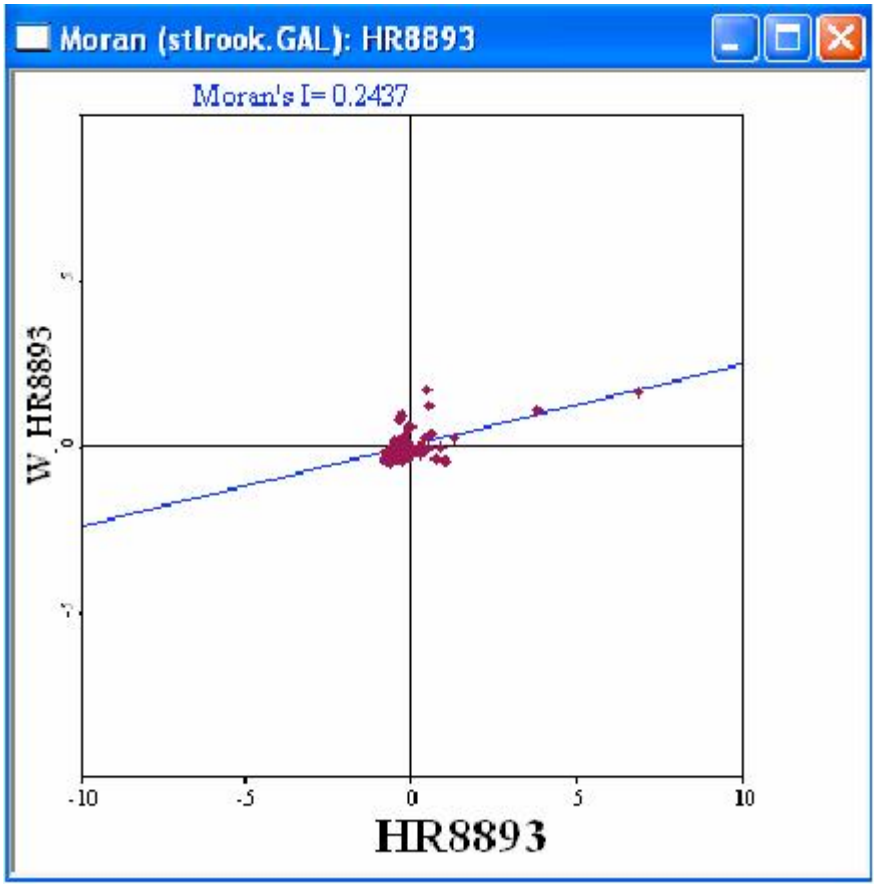
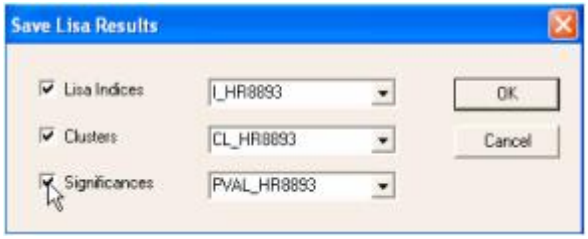


图 19.9 LISA 散点图

19.2.5 保存 LISA 统计值

一些中间过程的结果可以加入到数据表中。在任一地图中右击或以常用的方式从选项菜单中选择 Option>Save Results。出现一个对话框，建议为局部 Moran 统计值的变量名或 Lisa Indices(CL_HR8893)，显著位置的聚集类型(CL_HR8893)，序列程序(routine)p 值(P_HR8893)。选中默认值前的多选框，如图 19.10，点击 OK 将这些变量添加到数据表中。核实它们是否被加入，如图 19.11 所示。



19.10 LISA 保存结果选项

I_HR8893	CL_HR8893	PVAL_HR8893
-0.077288	0.000000	0.194000
0.172823	0.000000	0.068000
0.101758	0.000000	0.392000
0.089104	0.000000	0.250000
0.162797	0.000000	0.344000
-0.292601	0.000000	0.064000
-0.050662	0.000000	0.290000
0.145500	0.000000	0.242000
-0.000924	4.000000	0.042000

图 19.11 加入到数据表中的 LISA 统计值

19.3 推断

首先产生的显著性地图和聚集地图对是根据很快的计算，只利用 99 个序列和默认的显著性水平 p=0.05。在多数应用中，这已经是很好的估计了，但它好象对特定随机过程还有某些敏感性。为得到更多的稳定(robust)结果，增加序列数量到 999 或 9999 是很好的练习(对合理大小的数据集)，执行多次运行，直到结果稳定。

在显著性地图或聚集地图中右击，产生 Option 菜单，如图 19.12。选择 Randomization>Other，输入习惯的序列数值 9999，如图 19.13 所示。点击 OK，注意一些显著的县发生的些许变化。特别的，这个过程只会影响边缘显著(p=0.05)的县，但它会导致结果的总体分布出现相当大的改变。在当前的例子中，它似乎主要影响了 Moran 县中空间离群值的出现，IL(在地图北部中心部分点击，在表格中找到它的名字)，及其低低聚集的程度。

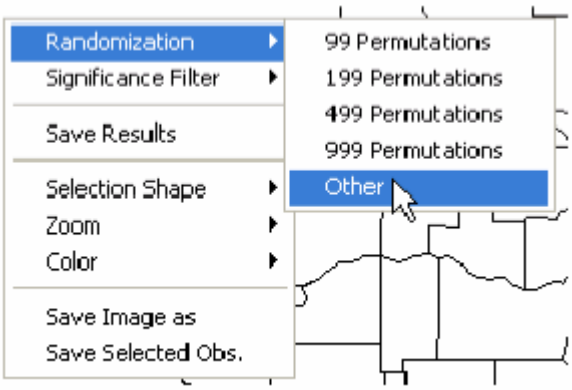


图 9.12 LISA 随机选项

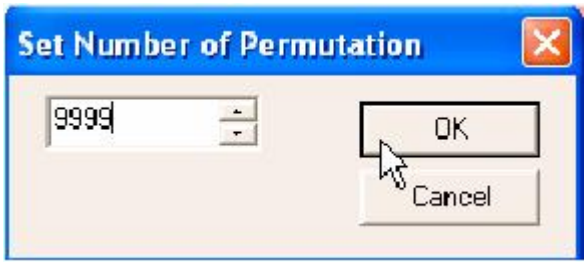


图 19.13 设置序列数值

可以改变地图显著性取舍点(cut-off)值来进一步评估结果的敏感性。在任一地图中右击，产生一

个 Option 菜单，但现在选择 Significance Filter>0.01，如图 19.14。注意显著性小于 0.05 而不是小于 0.01 的从地图中消失。

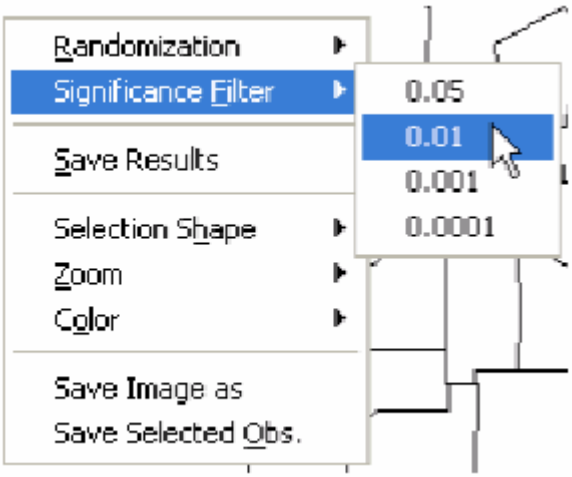


图 19.14 LISA 显著性过滤选项

例如在图 19.15 中，结果 LISA 聚集地图缺少显著的高低位置，使低-低聚集减少为一个县。将显著性水平设为 p=0.001，看一下其效果。即使在这样的需求水平下，一贯的显著县是相当稳定的空间聚集或离群值的指示。

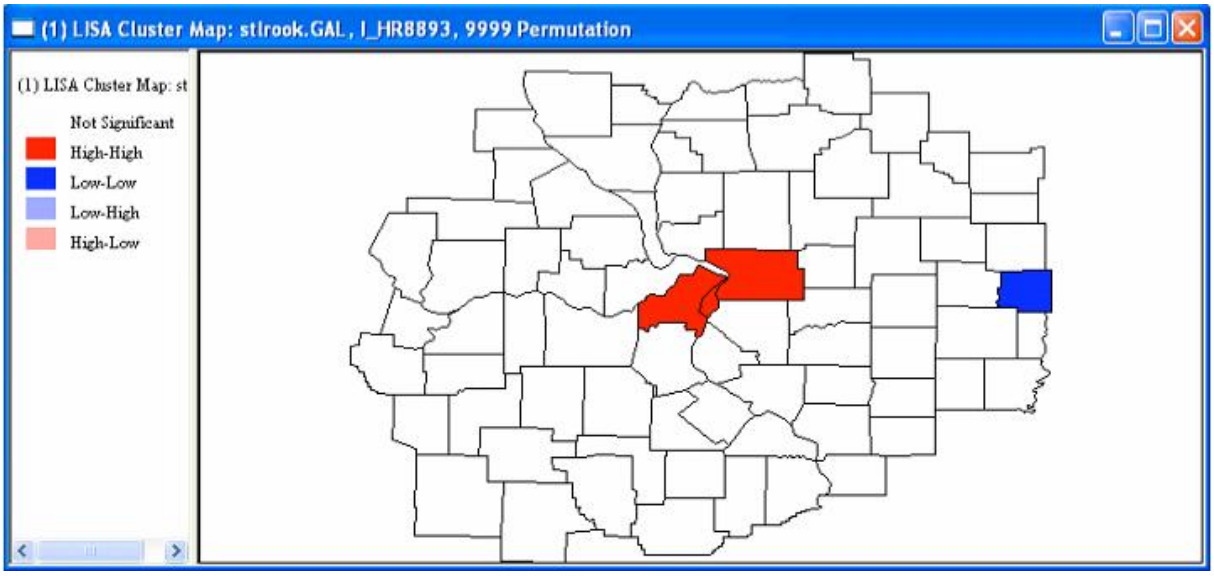


图 19.15 LISA 聚集地图 p<0.01

19.4 空间聚集和空间离群值

高-高和低-低位置(正局部空间自相关)是典型的被称为空间聚集，但高-低和低-高位置(负局部空间自相关)被称为空间离群。当根据定义，离群位置是单个，就不是聚集的情况。

要记住，显示在 LISA 聚集图中的所谓的空间聚集只是聚集中心。当一个位置的值(高或低)是与其邻居值(被相邻值的权重平均值所概括)相似超过空间随机的情况，聚集就这样划分。这种情况下的任何位置都会被标上地图。但聚集本身也可能会成为这个位置的邻居。

例如，在图 19.16 中，聚集县(p=0.01)的邻居为更好地显示空间聚集程度，用阴影交叉线表示。一些邻居被重叠，但不足为奇。总体而言，聚集的空间程度的印记比核心本身所暗示的大的多。但是，由更自由的 p=0.05 的显著性所暗示的核心与它并没有不同。

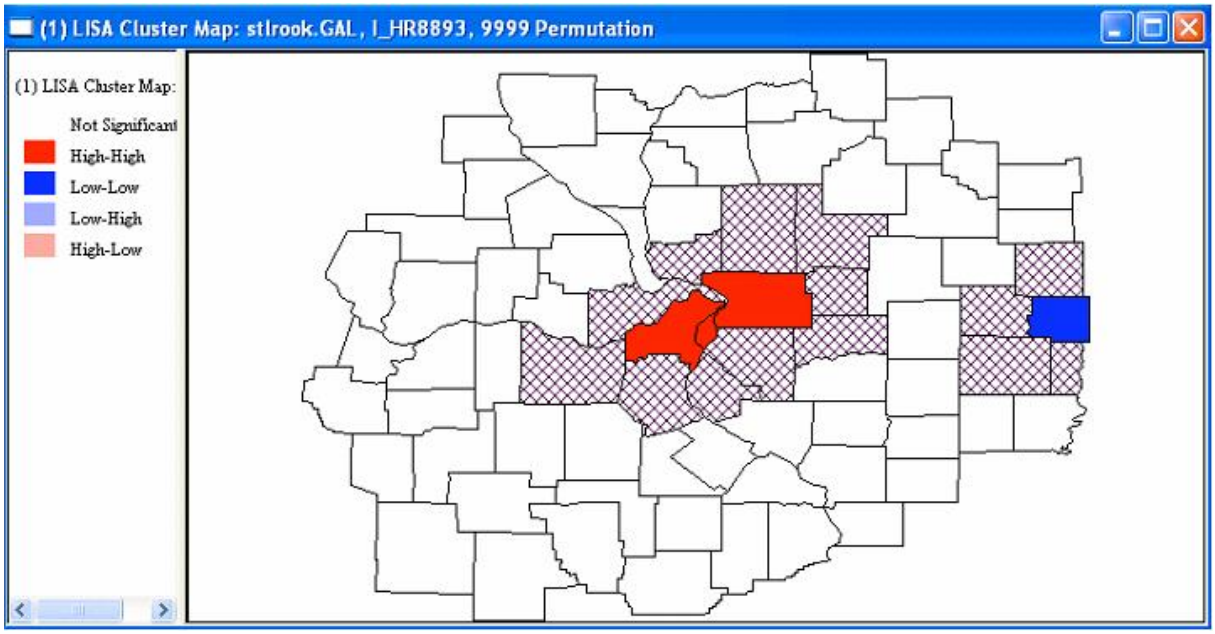


图 19.16 空间聚集

19.5 练习

从局部的角度，重新访问你在 18.4 部分执行的全局空间自相关分析。试着估计哪一个结果对显著性水平、序列数量或空间权重选择特别敏感。**根据这一仔细的分析，鉴别对这些因素稳定的空间聚集与离群**。

练习 20　比率的空间自相关分析

20.1 目标

本练习显示了空间自相关分析，比率的方差不稳定性调整。有关这些方法的方法背景见 Assuncao and Reis(1999)。

在本练习末，你将知道：

- 为比率创建一幅 Moran 散点图
- 用经验 Bayes(EB)调整来考虑 Moran 散点图中比率的方差不稳定性
- 用经验 Bayes(EB)调整来考虑局部空间自相关分析中比率的方差不稳定性

有关本操作更详细的信息请参考 User's Guide 第 97-98 页。

20.2 准备工作

当为比率或比例计算 Moran's I 统计值时，暗含的稳定假设可能会被比率的内在方差不稳定性所推翻。当处于危险的人口(基数)在所有观测点变化相当大时，后者就是具有方差不稳定性。方差不稳定性会导致 Moran's I 的错误推断。为了纠正，GeoDa 执行 Assuncao and Reis 推荐的 Empirical Bayes 标准化方法。这与计算 EB-smoothed 比率的 Moran's I 不同，但是一种利用相似(但不相同)的理由进行直接变量标准化方法。在 GeoDa 中，可以在全局(Moran 散点图)和局部空间自相关统计中执行这一功能。

为了能够与练习 18 和练习 19 的非标准化比率进行比较，我们将使用相似的数据和权重文件。对于全局尺度，我们将使用 Scottish 唇癌数据和相关的空间权重文件。
对于局部分析，我们将使用 St Louis 杀人案数据和一次 rook 权重文件。

20.3 EB 调整后的 Moran 散点图

载入 Scottish lip 数据，从菜单 Space>Moran's I with EB Rate 调用 EB-adjusted Moran's I，或点击相应的工具条按钮。会出现一个变量选择对话框，与前面用过的 smoothing 操作具有相同的格式。选择 Cancer 作为事件，Pop 作为基数变量，如图 20.2。

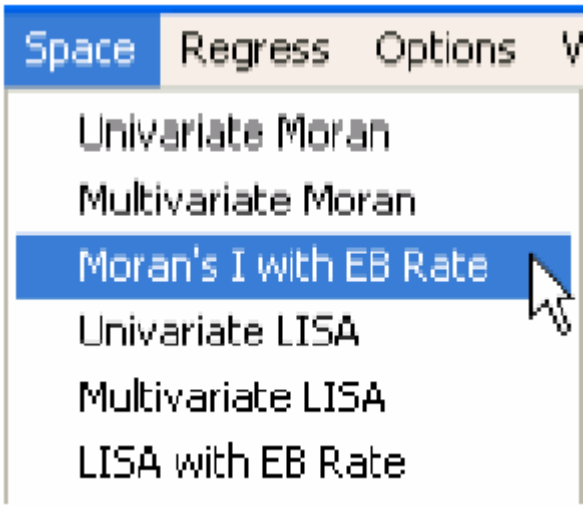


图 20.1 Empirical Bayes 调整后的 Moran 散点图

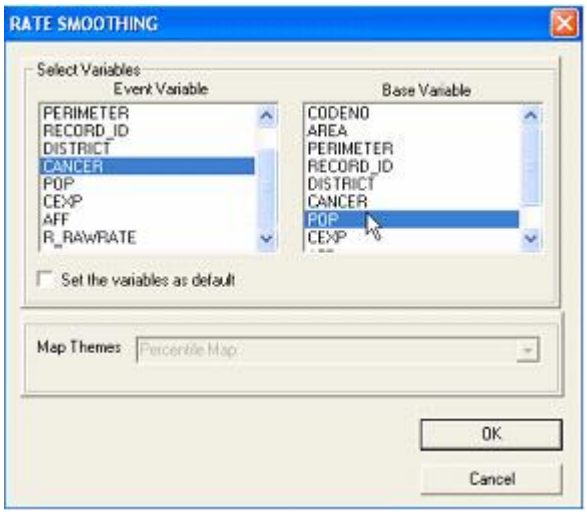


图 20.2 EB Moran 散点图变量选择

点击 OK 会出现一个权重选择对话框。如果你已经打开文件 scot5k,GWT，对话框会如图 20.3。

图 20.3 选择当前空间权重

点击 OK 会产生 Moran 散点图，如图 20.4 所示。注意 Moran's I 的值为 0.5311，与未经标准化比率统计值的 Moran's I 的值不同(在图 18.7 为 0.4836)。

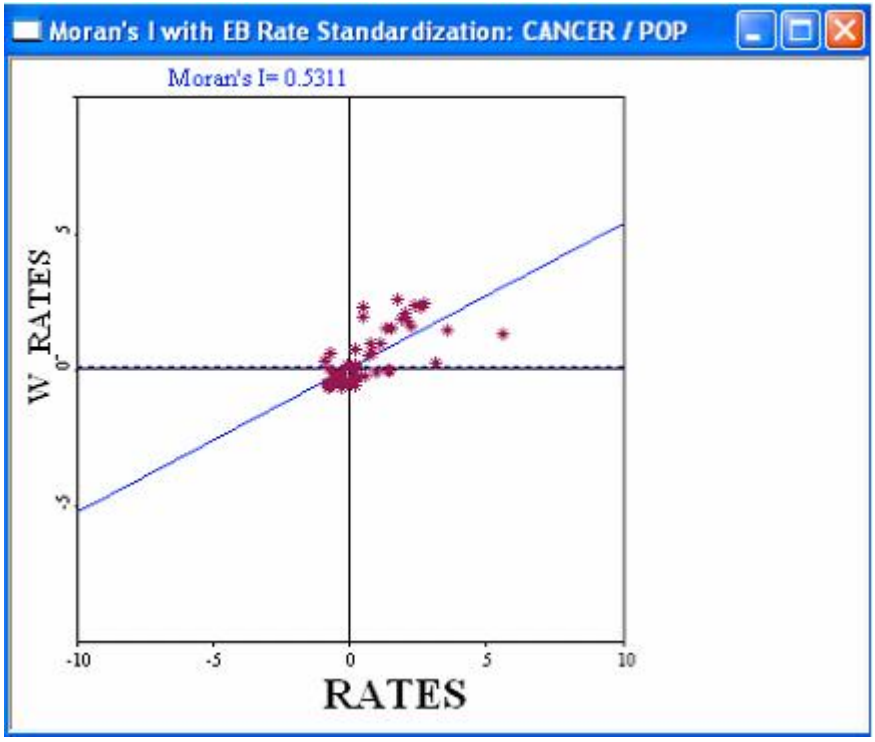


图 20.4 Scottish lip Cancer 比率的 Empirical Bayes 调整后的 Moran 散点图

更重要的是评估推断结果是否受到影响。如前，在图中右击出现 Option 菜单，选择 Randomization>999 Permutations。图 20.5 中结果序列经验分布仍暗示一种高的显著性统计，尽管伪显著性水平较低，p=0.04(你的结果可以由于随机序列而稍有不同)。点击 Run 几次，估计结果的稳定性。

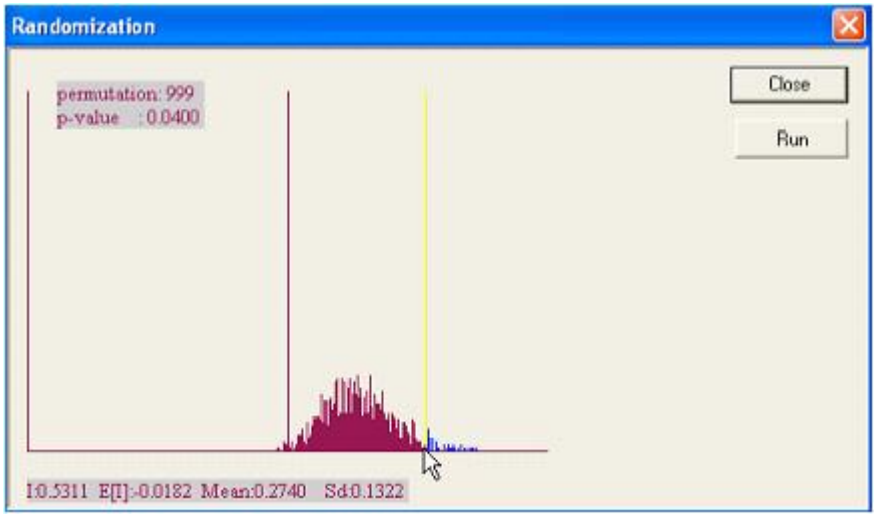


图 20.5 EB 调整后序列经验分布

20.4 EB 调整后 LISA 地图

在 GeoDa 中，EB 标准化也在局部 Moran 统计中执行。我们将使用练习 19 中由杀人案数目与县人口数计算的杀人案发生率来重复进行分析。

载入 St Louis 数据，点击工具条按钮或从菜单 Space>LISA with EB Rate 来调用 EB adjusted 局部 Moran(图 20.6)。

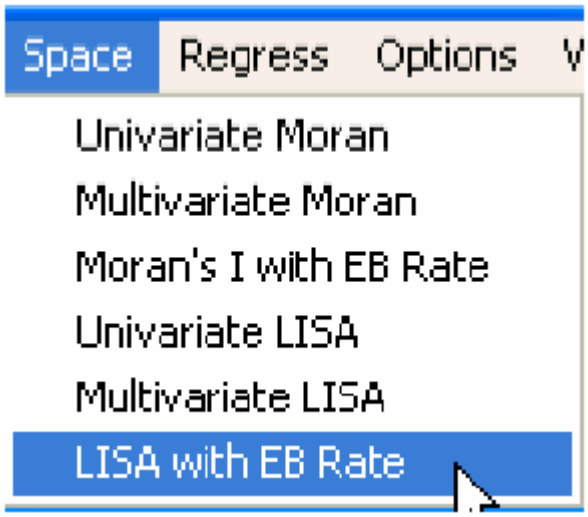


图 20.6 EB 调整后 LISA 功能

如前，在变量选择对话框中指定 Event 变量为 HC8893，Base 变量为 PO8893，如图 20.7。然后选择 stlrook.GAL 作为权重文件(图 20.8)。

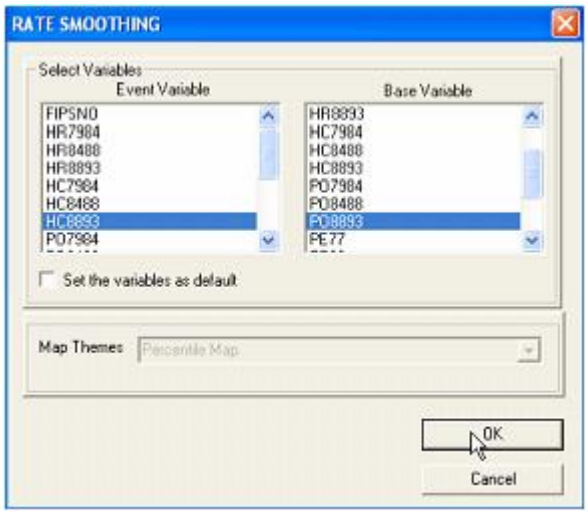


图 20.7 EB LISA 变量选择对话框

最后在结果窗口对话框中选中 Cluster Map 选项，如图 20.9。点击 OK 生成地图。

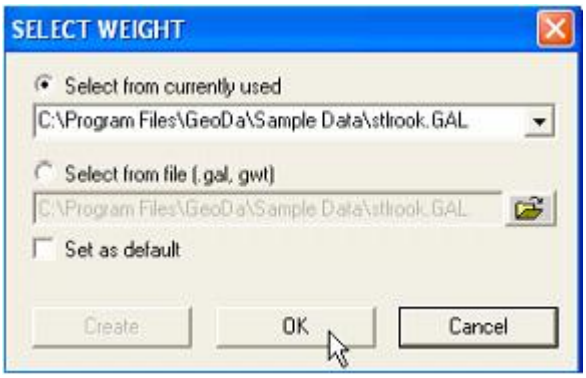


图 20.8 EB LISA 空间权重选择