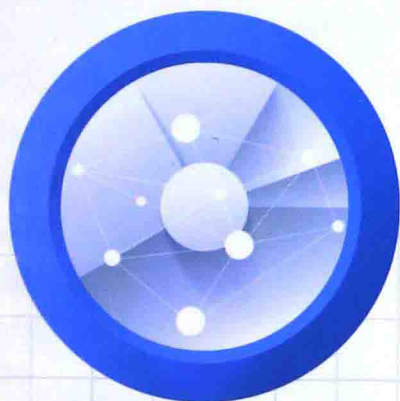




Spatial
Econometrics

空间计量学入门 与GeoDa软件应用

本书主要介绍空间计量学的入门知识，以及相对应的“入门级”软件GeoDa的应用和常见数据文件的导入。本书特别强调空间计量学的实践性，并以实际案例为导向讲述空间计量学在人文学科和自然学科中的应用。该书适用于区域经济学、法学、政治学等社会学科，以及医学、地理学、植物学、土壤学、地质学、水文学和气候学等自然学科本科生和硕士研究生以及对该学科感兴趣的一般读者。

ISBN 978-7-308-13837-6



9 787308 138376 >

定价：29.80元

空间计量学入门与 GeoDa 软件应用

陈安宁 著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目(CIP)数据

空间计量学入门与GeoDa软件应用/陈安宁著.

—杭州: 浙江大学出版社, 2014.10

ISBN 978-7-308-13837-6

I. ①空… II. ①陈… III. ①计量学 IV. ①TB9

中国版本图书馆CIP数据核字(2014)第210862号

空间计量学入门与GeoDa软件应用*

陈安宁 著



责任编辑 伍秀芳 (wxfw@zju.edu.cn)

封面设计 林智广告

出版发行 浙江大学出版社

(杭州市天目山路148号 邮政编码310007)

(网址: <http://www.zjupress.com>)

排版 杭州立飞图文制作有限公司

印刷 杭州杭新印务有限公司

开本 710mm×1000mm 1/16

印张 7.5

字数 122千

版印次 2014年10月第1版 2014年10月第1次印刷

书号 ISBN 978-7-308-13837-6

定价 29.80元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部联系方式: (0571)88925591; <http://zjdxcb.tmall.com>

前言

近年来,一些出国攻读硕、博士学位的学生陆续给我寄来他们学校(人文学科)硕、博士学位的课程表或教学大纲,甚至教材。我发现西方,特别是美国学校除哲学外的人文学科普遍比较重视与计量有关的知识教授,且不少课程在国内还属空白或尚处起步阶段。其中,有一门叫“空间计量经济学”的课程相当吸引我的注意。我曾在网上做过一番调查,注意到几乎没有经济学“出身”的学院专门开设此门课程,只有极少数地理学“出身”的学院在开此门课程。为此,2012年深秋,我们特请北京大学公共管理学院博士生导师沈体雁和他的博士后张晓欢来我校做空间计量经济学讲座,受益非浅,在此深表谢意。另外,在收集资料过程中,我曾收获到郑光辉翻译安塞林的《用GeoDa探索空间数据:工作手册》中文版,着实大大减轻了我学习和写作过程的劳动,在此本人也深表万分感谢。

在我国应用经济学中,区域经济学是硕士点最多的二级学科。不过,规模大并不意味着实力强。据我所知,在我国高校中,诸如“新经济地理学”和“空间计量经济学”此类“现代化”区域经济学的基础课程普及率极低。如果长期如此,我们培养的大多数区域经济学学生将会缺少一种与国外同行交流的语言。因而,我们需要普及开设相关课程。然而,要开设“空间计量经济学”首先遇到的困难就是缺乏教材,国内至今没有一本让初学者学完后就能实践的教材。故2012年我就打算写一本入门级的空间计量经济学教材。在上述空间计量经济学讲座期间,我曾与一些外校学生交谈过,发现他们学习中遇到的最大障碍却是计算机软件的应用。诚然,今日学习“计量”与“统计”之类的课程必须要掌握一门以上有关的软件,否则,学生就会因缺乏实践,或实践中支付的劳动成本过高(假若用手工计算的话),而无法获得阶段性成就感,进而便会失去继续学习的兴趣。至于学习有关“空间计量经济学”的软件也非登天难事。像我们常见的Matlab、Stata、R语言之类的软件都有相关空间计量的软件包和工具箱(有些需要另外下载),只是国内一般大学的人文类学院很少会开设相

关课程。而这些软件又因要涉及编写程序和命令等,往往造成许多文科学生“谈软色变”。不过,两年前,我发现一种可以“很傻瓜”(只需点点菜单、工具栏、左键和右键就行)处理空间计量问题的自由软件——GeoDa,很适合中国师生“快学快用”的需要。因而,我就打算写一本用 GeoDa 为工具的空间计量经济学。只是在使用 GeoDa 过程中,我发现这款软件不仅可以在空间计量经济学中应用,而且还可以在法学、政治学等人文学科中应用,甚至还可以在医学、地理学、植物学、土壤学、地质学、水文学和气候学等自然学科中应用。所以,我决定将教材的名称由“空间计量经济学及 GeoDa 应用”改为“空间计量学及 GeoDa 应用”,以扩大其适用性。

尽管在写作过程中自己付出了很大努力,不过因时间仓促和水平有限,预期书中仍然会存在一些尚未被发现的疏漏之处。所以,特别期待各位同仁和读者对本教材提出指正与建议,便于本教材的完善。

陈安宁,2014年夏于杭州文汇苑

目 录

前 言	iii
第1章 空间计量学概述	1
1.1 什么是空间计量学	1
1.2 空间计量学的发展	2
1.3 空间计量学的应用软件介绍	3
1.4 GeoDa使用简介	4
第2章 数字地图操作	9
2.1 分位数地图	9
2.2 箱形地图	12
2.3 百分位地图	13
2.4 标准差地图	14
2.5 平滑地图	15
2.6 地图的输出	17
2.7 多幅地图制作	18
2.8 选择、链接和修饰观察对象	18
2.9 动画地图	21
第3章 数据表格处理	24
3.1 浏览与选择	24
3.2 排序	26
3.3 查询	27
3.4 计算	28
第4章 点与多边形应用与转换	32
4.1 点状“地图”文档的输入	32
4.2 多边形“地图”文档的输入	35
4.3 底图与数据表格的链接	37
4.4 创建规则格网多边形 shape 文档	40

4.5	创建包含质心坐标的点文档	42
4.6	在数据表中添加质心坐标	44
4.7	创建泰森多边形文档	46
	附 录	48
第5章	统计图表与链接功能	53
5.1	直方图	53
5.2	箱形图	57
5.3	散点图	62
5.4	3D 散点图	66
5.5	平行坐标图	71
5.6	条件图	76
5.7	条件地图	79
5.8	统计地图	81
第6章	空间关联性探索	84
6.1	邻接、距离与空间权重矩阵	84
6.2	用 GeoDa 处理空间权重	89
6.3	探索性空间数据分析	93
6.4	用 GeoDa 计算莫兰指数 I	96
第7章	空间回归模型分析	102
7.1	空间自回归模型	102
7.2	用 GeoDa 处理空间自回归模型	103
	参考文献	112

第1章 空间计量学概述

1.1 什么是空间计量学

尽管空间计量学已经发展了二三十年,不过至今尚未有一个统一的定义。这里为了教学方便,我们顾名思义地给它下一个定义。空间计量学就是一门运用计量,特别是概率统计和数字地图的方法对变量之间“空间”上的“因果”关系进行定量分析的学科。与“一般”计量学的不同之处,空间计量学尤其强调变量空间上的“自相关性”。比如,一个国家的经济发展水平,除了受本国的经济资源禀赋(如,劳动力、资本和技术等)外,还会受到周围国家经济发展水平的影响。一般计量学通常强调前者的“因果”关系,偶尔也将后者简化为等同于前者的一系列变量参与“因果”关系分析。这样得到的结果充其量也只是反映其他国家对该国经济发展的影响,而不能完整地刻画整个地区内所有国家之间空间关系因素对经济发展的“平均”相互影响关系。而空间计量学可以弥补一般计量学这种缺陷。前面一直把“因果”两个字打上引号就是要说明这里的“因果”只是在求证这些变量之间是否存在因果关系,而非确定这些变量事实上存在因果关系。

为了强调“空间计量”方法的普适性,本书取名中选用了“空间计量学”一词。熟悉计量经济学的读者会注意到将“*econometrics*”译成“计量经济学”更多的是因为“习惯”,而非“恰当”,或“准确”(事实上,在“计量经济学”译法出现后,也有人曾经企图模仿生物中将“*biometrics*”译成“生物计量学”,将“*econometrics*”改译成“经济计量学”,但是,在已经习惯化的“计量经济学”面前,只能是“沧海一粟”和“无能为力”)。“计量经济学”与其他“××经济学”(如,产业经济学、环境资源经济学等)有本质上的不同。其他“××经济学”是研究××领域的经济运行规律。相反,计量经济学并不是研究计量领域的经济运行规律,而是研究从经济领域的“经验数据资料”中,如何揭示和归纳出经济规律的“合理”技

术。因而,它研究的对象不是经济问题,而是数学问题,是概率统计在特定领域的深化和延续。“计量经济学”中的许多方法和模型的应用也绝非仅仅局限在经济领域,往往可以扩展到其他人文科学,甚至自然科学中去。

类似于有人将计量经济学分为理论计量经济学和应用计量经济学,我们这里也可以将空间计量学分为理论和应用两类。前者主要研究空间计量的理论和方法,其包括模型设立、估计、检验、相关算法分析,以及指标构建。后者主要将理论和方法通过相关软件应用到具体领域中去。为了满足国人“快学快用”的需要,本书重点放在应用计量学层面,而对理论部分只作简单介绍。

1.2 空间计量学的发展

空间计量学根据其研究内容可以分为广义(包括传统和现代)和狭义(或现代)两类。如果按广义内容来看,空间计量学的出现可以追溯到十七、十八世纪。它以概率论、统计学和数量地图(如等值图、点值图、分级统计图和图形统计图等)的出现为关键性标志。而从狭义内容来看,空间计量学的产生和发展只是20世纪80、90年代以后的事情。狭义(现代)空间计量学是以研究要素空间集聚(或扩散)和关联等现象(如集聚和扩散指标设立、空间“滞后”模型和空间误差模型等)为标杆性特征。

现代空间计量学的诞生主要受计算机,特别是PC机的快速发展的驱动。其实,今天我们应用的许多模型并非是在20世纪后叶才构建的,只是早前缺乏必要的计算手段,往往只能纸上谈兵,流于形式。20世纪80年代,为了计算一个十几个样本数据最小二乘法的三元一次线性回归方程,十几个人用算盘(尽管当时已有计算机和计算器,但是,这些“老”先生既不会计算机,又不喜欢计算器,且还特别习惯于算盘)计算,花了大约3个月才完成。这里所谓“完成”也只是确定了方程的系数,其中许多检验值还尚未涉及。一般的多元回归尚欠如此,那么,绝大多数人对需要循环迭代才能确定系数的非线性模型肯定只能束手无策了。

在20世纪80、90年代,与空间计量学密切相关的空间统计学和地理信息系统(GIS)已经发展到相当水平,为空间计量学提供了坚实的“原料”基础。随之,空间计量学也就应运而生了。在众多部门空间计量学中,发展最为突出的

是空间计量经济学。空间计量经济学现已发展出一系列适合空间经济计量的模型体系。

1.3 空间计量学的应用软件介绍

从前面空间计量学发展的历史中,我们可以看到,要“方便”地完成空间计量分析必须求助于计算机。然而,在很长一段时间里,标准统计和计量软件并不能提供空间计量的功能。这使得那些对空间计量感兴趣的研究者不得不需要自己编写程序。这无疑将相当一部分不会编程的研究者拒之门外。在最近十几年中,这种情况发生了很大改变。一些公司和个人一方面开发出专门应用于空间计量问题的软件;另一方面,在传统标准的统计和计量软件加入有关空间计量的工具箱或软件包。下面我们将“浏览”一下一些可用于空间计量和统计的主要软件。

□Matlab

Matlab 是一款由美国 The MathWorks 公司出品的商业数学软件。Matlab 是一种用于算法开发、数据可视化、数据分析以及数值计算的软件,主要应用于工程计算、控制设计、信号处理与通讯、图像处理、信号检测、金融建模设计与分析等领域。勒萨热和皮斯等人专门编写和开发了有关空间计量经济学模型运算的 Matlab 工具箱 Spatial Econometrics(可以免费从 <http://www.spatial-econometrics.com> 获取)。

□SAS

SAS 是美国 SAS Institue Inc. 公司出品的商业统计软件,广泛应用于金融、医药、卫生、生产、运输、通讯、政府和教育科研等方面的计量和统计工作。在数据处理和统计分析领域,SAS 系统被誉为国际上的标准统计软件系统。斯卡本伯格和果特威等人开发了有关空间数据分析的扩展集(可以从 <http://www.crcpress.com> 获取)。

□Stata

Stata 是 Statacorp 公司开发出来的统计程序,广泛应用于经济学、社会学、政治学及流行病学等领域的统计和计量工作。皮萨蒂编写了有关空间自回归

和误差模型的宏 `spatreg`。

□R 语言

R 语言是一种自由软件编程语言与操作环境,主要用于统计分析、绘图、数据挖掘。R 语言本来是由来自新西兰奥克兰大学的爱海克和杰德曼开发,现在由“R 开发核心团队”负责开发。R 语言中的 `spBayes` 软件包可以用于空间计量研究(在使用 R 语言时,直接下载)。

□ArcInfo

ArcInfo 是 Esri 公司开发的一个全功能 GIS 产品,包括很多可以用于空间数据集的统计分析工具,尤其是其中的 Statistical Analyst 工具包特别擅长计算邻里数据和分类分析。

□GeoDa

GeoDa 是由著名空间计量经济学家安塞林和他同事开发的自由软件,它是一种设计实现栅格数据探求性空间数据分析的软件。这种软件主要针对空间计量学的初学者,使用者不需要任何编程,通过鼠标点击就能完成所有运算操作。

1.4 GeoDa 使用简介

GeoDa 的操作相当容易,与大家日常用的 MS Office 相似,甚至更简单,只需点击主菜单和工具栏,加上鼠标左键和右键的交替使用就能完成所有操作。它的输出结果也可以用标准的 Windows 指令作进一步处理。下面我们就来介绍 GeoDa 软件的基本操作。

□安装软件

GeoDa 软件可以直接从 <http://www.geoda.uiuc.edu/downloadin.php> 免费下载,目前流行的有两个版本:GeoDa095i(Beta)和 OpenGeoDa。由于国内普遍使用的仍是 GeoDa095i(Beta),故本书主要讲述 GeoDa095i(Beta)。通过下载和解压缩后得到一个 `setup.exe` 文件,直接用左键双击该文件(以后照章办事)就能顺利安装。安装完毕后,在桌面会出现如图 1.1 所示的图标。



图 1.1 GeoDa 图标

□启动软件

双击桌面图标或通过 Windows Explorer(相应目录中)启动 GeoDa, 就会出现一个欢迎使用 GeoDa 的界面。欢迎界面很快消失后, 便会出现 GeoDa 窗口(图 1.2)。

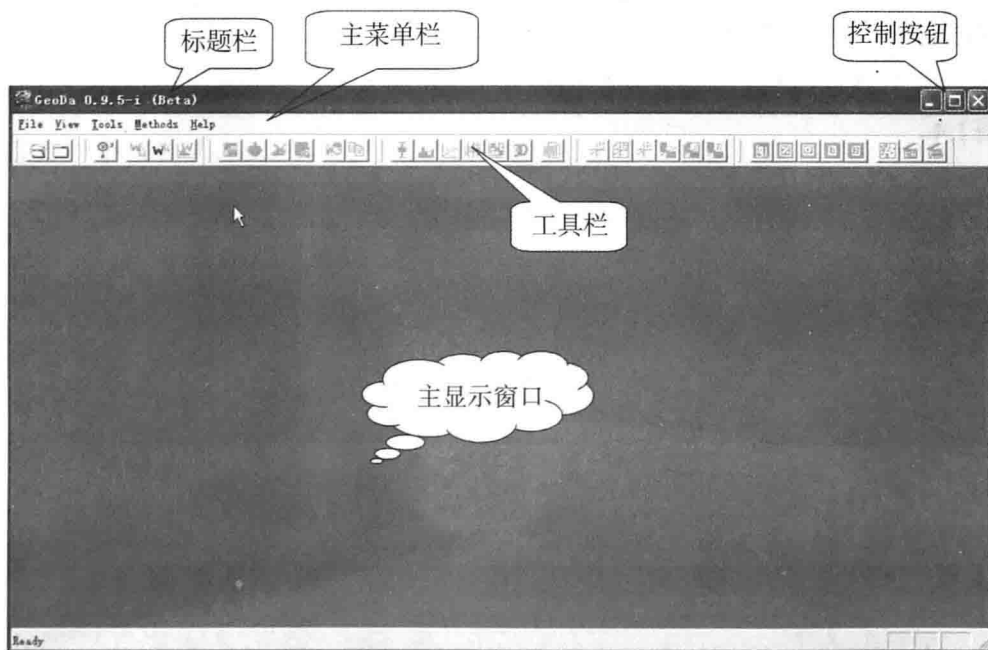


图 1.2 GeoDa 窗口

□GeoDa 窗口介绍

标题栏: 在图 1.2 中, 窗口的顶部是标题栏, 显示软件名称 GeoDa 0.9.5-i (Beta)。标题栏的右端有 3 个控制按钮: 最小化、最大化(或复原)和关闭, 点击这 3 个按钮可以控制窗口大小或关闭窗口。

主菜单栏: 标题栏下是主菜单栏。没有打开项目时, 标题栏只有 5 项:

“File(打开和关闭文件)”、“View(选择要显示的工具栏)”、“Tools(空间数据处理)”、“Methods(空间回归的方法)”和“Help(帮助项,至今还不能使用)”。打开项目后,“Methods(空间回归的方法)”消失,其他将增加“Edit(控制地图窗口和图层)”、“Table(数据表格处理)”、“Map(地图编制)”、“Explore(统计图表编制)”、“Space(空间自相关分析)”、“Regress(空间回归)”、“Options(特殊的应用选项)”和“Windows(选择或重新排列窗口)”等项目。

工具栏:它在主菜单下方,由6组按钮组成,从左到右分别是:打开和关闭项目;空间权重计算;地图编辑功能;探索性数据分析;空间自相关;等级平滑和制图。

单击工具栏中按钮与在菜单中选择相应的选项功能是相同的。工具栏可以放置于任何位置,也就是说,可以将它们移动到使用者喜欢的任何位置。选择一个工具栏,单击左侧分割栏的鼓起处,如图1.3,把它拖到不同的位置,如图1.4。

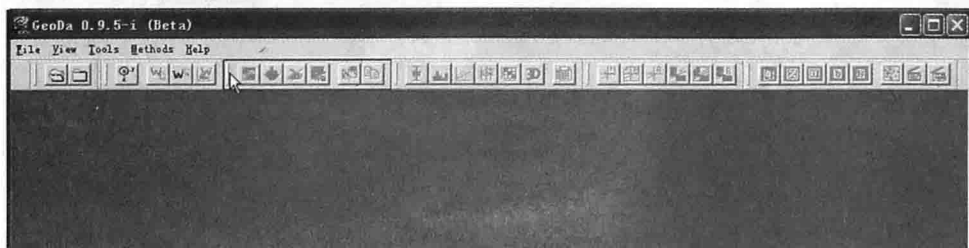


图 1.3 准备移动工具栏

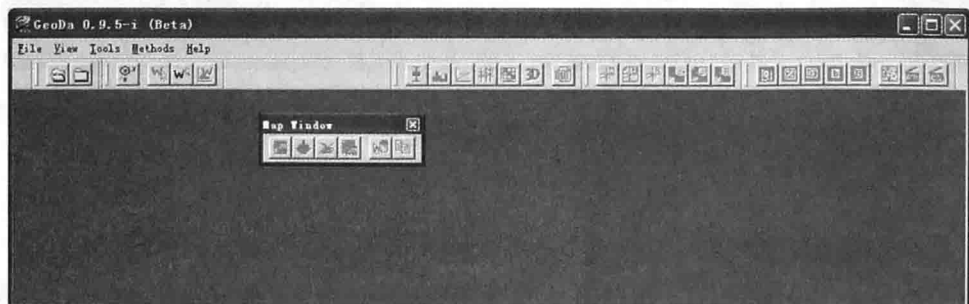
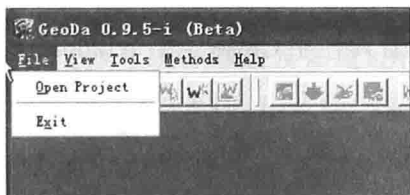


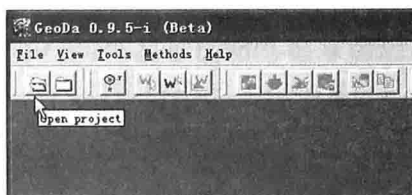
图 1.4 移动后的工具栏

□打开项目

在File菜单中选择Open Project(图1.5(a)),或单击Open Project 工具按钮(图1.5(b))后,会出现熟悉的Windows对话框(图1.6),要求输入GeoDa文档的路径和文件名及可用的关键字。关键字是每一记录(observation)的唯一标识,它是一个整数值(绝对不能有小数),如各个县的FIPS码。GeoDa文档有3个档:.shp、.shx和.dbf。前两者是地图或图形档;后者是资料(特别是数据)档,缺一不可。读入.shp档后,其余2档也会自动读入。比如,我们要打开一个美国密苏里州圣路易斯78个县项目(电子地图),它的文件名为stl_hom.shp,存放路径为C:\Program Files\GeoDa\Sample Data\stl_hom。点击Open Project后,出现一个Project Setting对话框,我们在Input Map中输入(直接用打开按钮更为方便)C:\Program Files\GeoDa\Sample Data\stl_hom\stl_hom.shp,再在Key对话框中输入(往往有一个默认变量存在)FIPSNO,再点击“OK”按钮就打开一个项目了。



(a)



(b)

图1.5 (a)从菜单中打开项目;(b)用工具按钮打开项目

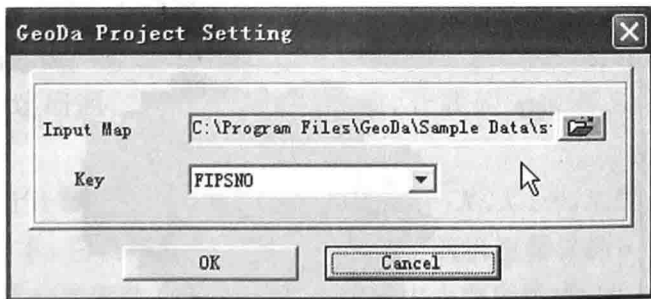


图1.6 选择输入shape文件

接着,打开一个地图窗口,显示用于分析的底图,图中显示了圣路易斯78个县(图1.7)。窗口左侧区域显示了图例。通过向左或向右拖拽两区(图例和地图)之间的分隔线可以改变其大小。

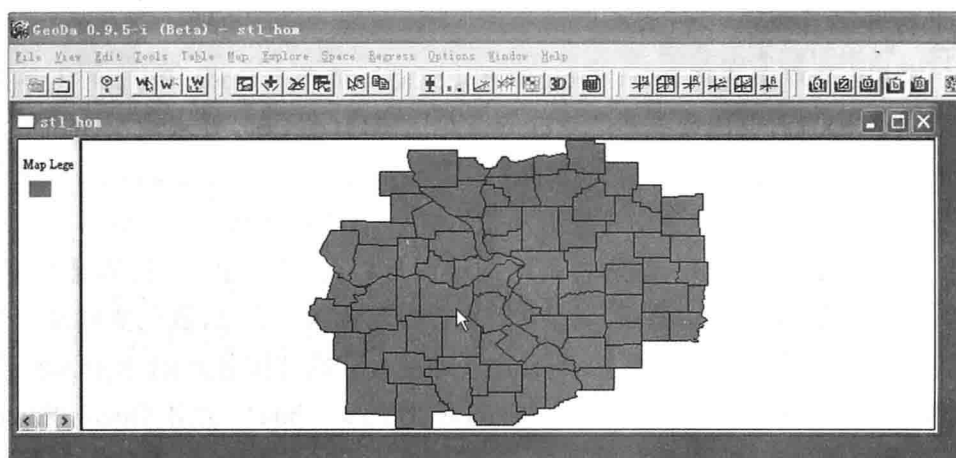


图 1.7 载入 stl_hom 样本数据后打开的窗口

□关闭窗口

点击菜单 Files 中的 Close All, 或点击工具栏中 Close All Windows 按钮 (图 1.8), 就可以关闭所有窗口。

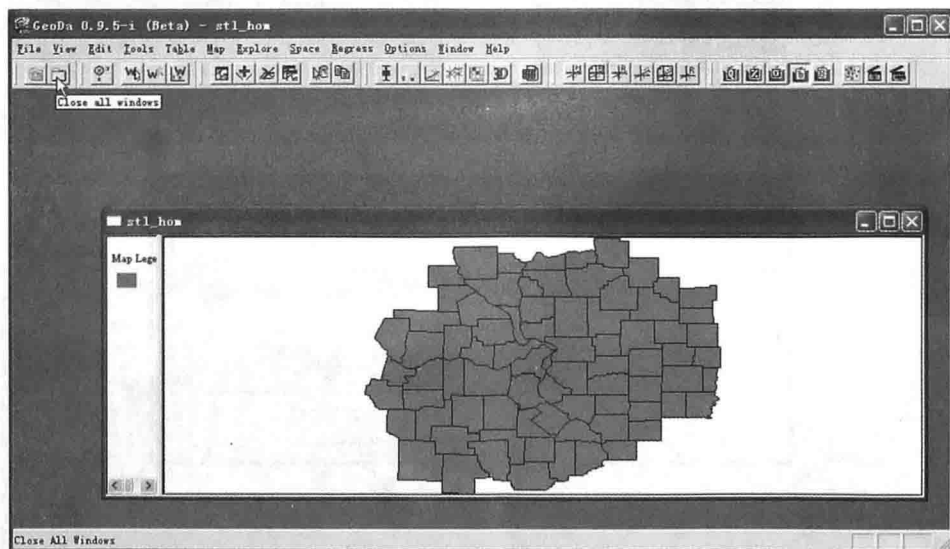


图 1.8 关闭项目

第2章 数字地图操作

空间计量的特色之一就是相关计量内容需要落实到地图上。GeoDa 能够完成的数量地图有四类:分位数地图(Quantile)、箱形地图(Box Map)、百分位地图(Percentile)和标准差地图(Std Dev)。

2.1 分位数地图

□分位数

学过数理统计的朋友知道,所谓分位数(或点)就是随机变量 X 累积概率(或分布函数)等于 $(0,1)$ 区间内某一固定值所对应的随机变量值。分位数包括上侧分位数、下侧分位数和双侧分位数。

$\forall \alpha \in (0,1)$,能满足 $P\{X > Z_{\alpha}\} = 1 - F(Z_{\alpha}) = \alpha$,就称 Z_{α} 为 α 的上侧分位数;能满足 $P\{X \leq Z_{\alpha}\} = F(Z_{\alpha}) = \alpha$,就称 Z_{α} 为 α 的下侧分位数;能满足 $P\{|X| > Z_{\alpha}\} = \alpha$,就称 Z_{α} 为 α 的双侧分位数。习惯上,或者说在没有特殊说明的情况下,我们所说的分位数都是指下侧分位数。然而,在统计学中,人们更习惯于应用 n 分位数的概念描述问题。在讨论 n 分位数前,让我们先来温习一下样本分布函数。

设 X 为具有分布函数 $F(X)$ 的随机变量,如果, $X_1, X_2, \dots, X_n$ 是具有相同分布函数 $F(X)$ 的样本,它们的观察值 $x_1, x_2, \dots, x_n$ 为随机变量 X 的 n 个独立的观察值。再假定有关分布函数 $F(X)$ 的观察值是按大小顺序排成,即

$$x_1 \leq x_2 \leq \dots \leq x_n$$

若 $x_k \leq x \leq x_{k+1}$,则不大于 x 的观察值频率为 k/n 。因而,函数

$$F_n(x) = \begin{cases} 0, & x < x_1 \\ \frac{k}{n}, & x_k \leq x < x_{k+1} \\ 1, & x_n \leq x \end{cases}$$

等于在 n 次重复独立试验中, 事件 $\{X \leq x\}$ 的频率, 我们称 $F_n(x)$ 为样本分布函数, 或经验分布函数。类似上面, 我们也可以给出样本分布函数的分位数定义。

n 分位数是指将一个样本分布函数分布范围分为 n 个等份的一系列数值点。比如, 4 分位数就是将概率分布均等分为 4 等份的一系列数值点, 即(累积)概率分别为 25%、50% 和 75% 时的分位数(或点), 这里将它们分别记为 Q_{25} 、 Q_{50} 和 Q_{75} 。5 分位数分别记为 Q_{20} 、 Q_{40} 、 Q_{60} 和 Q_{80} 。其中, 4 分位数最为常见和有用。 Q_{50} 也称中位数, Q_{75} 也称上四位数, Q_{25} 也称下四位数。下面我们以四分位数为例说明 n 分位数的算法。根据定义, 中位数的计算方法为

$$Q_{50} = \begin{cases} x_{\frac{n+1}{2}}, & n=2k-1 \\ \frac{x_{\frac{n}{2}} + x_{\frac{n+2}{2}}}{2}, & n=2k \end{cases}$$

下四位数和上四位数有两种略有差异的定义。按照第一种定义, 下四位数和上四位数计算方法分别为

$$Q_{25} = \begin{cases} x_{\frac{n+1}{4}}, & n=4k-1 \\ x_{\lceil \frac{n+1}{4} \rceil} + \left(x_{\lceil \frac{n+1}{4} \rceil + 1} - x_{\lceil \frac{n+1}{4} \rceil} \right) \left(\frac{n+1}{4} - \lceil \frac{n+1}{4} \rceil \right), & \text{other} \end{cases}$$

$$Q_{75} = \begin{cases} x_{\frac{3(n+1)}{4}}, & n=4k-1 \\ x_{\lceil \frac{3(n+1)}{4} \rceil} + \left(x_{\lceil \frac{3(n+1)}{4} \rceil + 1} - x_{\lceil \frac{3(n+1)}{4} \rceil} \right) \left(\frac{3(n+1)}{4} - \lceil \frac{3(n+1)}{4} \rceil \right), & \text{other} \end{cases}$$

按照第二种定义, 计算下四位数和上四位数计算方法分别为

$$Q_{25} = \begin{cases} x_{\frac{\lceil \frac{n}{2} \rceil + 1}{2}}, & \lceil \frac{n}{2} \rceil = 2k-1 \\ \frac{x_{\lceil \frac{\lceil \frac{n}{2} \rceil + 1}{2} \rceil} + x_{\lceil \frac{\lceil \frac{n}{2} \rceil + 1}{2} \rceil + 1}}{2}, & \text{other} \end{cases}$$

$$Q_{75} = \begin{cases} x_{n+1 - \frac{\lceil \frac{n}{2} \rceil + 1}{2}}, & \lceil \frac{n}{2} \rceil = 2k-1 \\ \frac{x_{n+1 - \lceil \frac{\lceil \frac{n}{2} \rceil + 1}{2} \rceil} + x_{n - \lceil \frac{\lceil \frac{n}{2} \rceil + 1}{2} \rceil}}{2}, & \text{other} \end{cases}$$