

在 ArcGIS 9 中进行空间统计

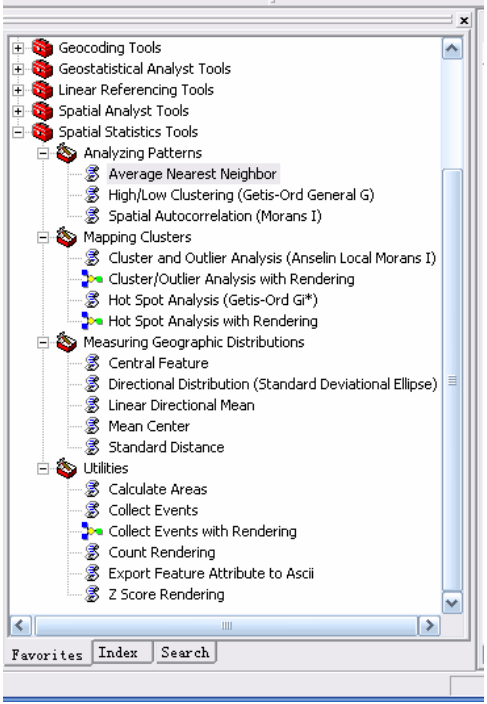
作者：ESRI 中国（北京）培训中心 姜云鹏

随着 GIS 在各个领域应用的不断扩展，有些特殊的行业，比如流行病学、生物学、气象、地质等行业，他们需要更深入的挖掘空间数据信息，这些信息的获得是与传统的 GIS 分析结果不尽相同的。比如：传统的 GIS 分析侧重于研究空间要素之间的关系，比如相邻、叠加、以及要素之间的距离、连通性等等。而这些特殊行业他们需要根据多种采样的数据来研究空间事物的变化信息，分布特征等信息，这些信息的获得，往往是一种统计分析的结果，而在空间上，事物的分布又是相互关联的。所以，空间统计应运而生。所谓空间统计，就是将空间信息与属性信息进行统一的考虑，研究特定属性或属性之间与空间位置的关系。

空间统计主要的工作是研究空间自相关性（Spatial Autocorrelation），分析空间分布的模式，例如聚类（cluster）或离散（dispersed）。通过使用 ArcGIS 9 中的空间统计工具，用户可以以一种非常直观而简单的方式获得这些信息。

ArcGIS 9 中的空间统计工具箱包括了一系列工具，用来分析地理要素的空间分布形态。传统的统计并不考虑地理要素的空间关系，而在空间统计中，要素的空间关系是分析中需要考虑的必要的，处于绝对重要地位的。

因此，对于空间数据分析的目的来说，使用 ArcGIS 9 中的空间统计工具比使用原来的不考虑空间信息而进行统计的工具要更为合适。通过使用这些工具，GIS 用户可以采用一种更高级的方法来解决空间数据分析中的问题。表一列出了主要的空间统计工具集以及它们的功能描述：



表一：空间统计工具集及其功能

工具集	功能
空间分布模式分析工具集 (Analyzing Patterns)	确定要素的某些属性值在一个区域中是聚集分布，均匀分布，或者是随机分布状态
聚集分布制图工具集 (Mapping Clusters)	确定统计量上重要的 hot spots(最受关注地区)，cold spots（不受关注地区）以及一些有特例的地区。
度量空间分布工具集 (Measuring Geographic Distributions)	确定数据的中心位置，数据分布的形状及方向性，离散数据的离散程度
辅助工具集 (Utilities)	对数据进行重新处理或符号化分析结果

说明：这些工具为 ArcGIS 9 中新增加的核心功能：

在 ArcGIS 8 版本中，空间统计中的许多工具已经以开发者例子程序提供给了用户。而在 9 版本中，这些工具都被包括进了核心的功能模块中，成为了 ArcGIS 平台的组成部分。而且，ArcGIS 9 的空间统计工具在

ArcGIS 各个 License 级别中均可使用。与 ArcGIS 9 的 Geoprocessing 工具一样，空间统计的工具也可以通过对话框、命令行以及模型（Model）等多种方式运行。通过与 Geoprocessing 框架下的其他工具进行组合，空间统计工具也可以很容易的进行扩展，用户可以创建自己的工具，也可以同第三方产品进行协同工作。

空间统计工具箱下的每个工具集都是按照功能进行分组的。下面会对每个工具集中的工具进行详细的描述：

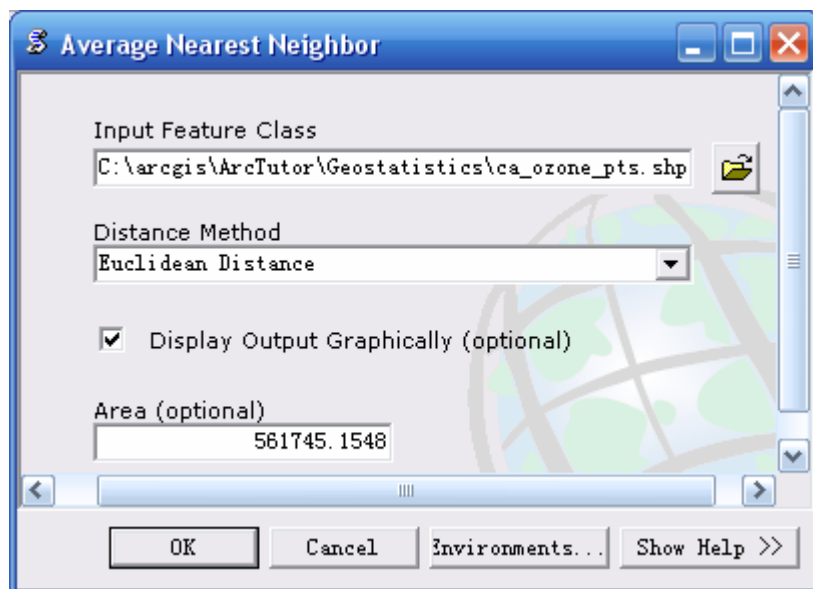
一、分析模式工具集（Analyzing Patterns）：

该工具集中的工具主要用来探讨数据的空间分布特征。包含三个工具：

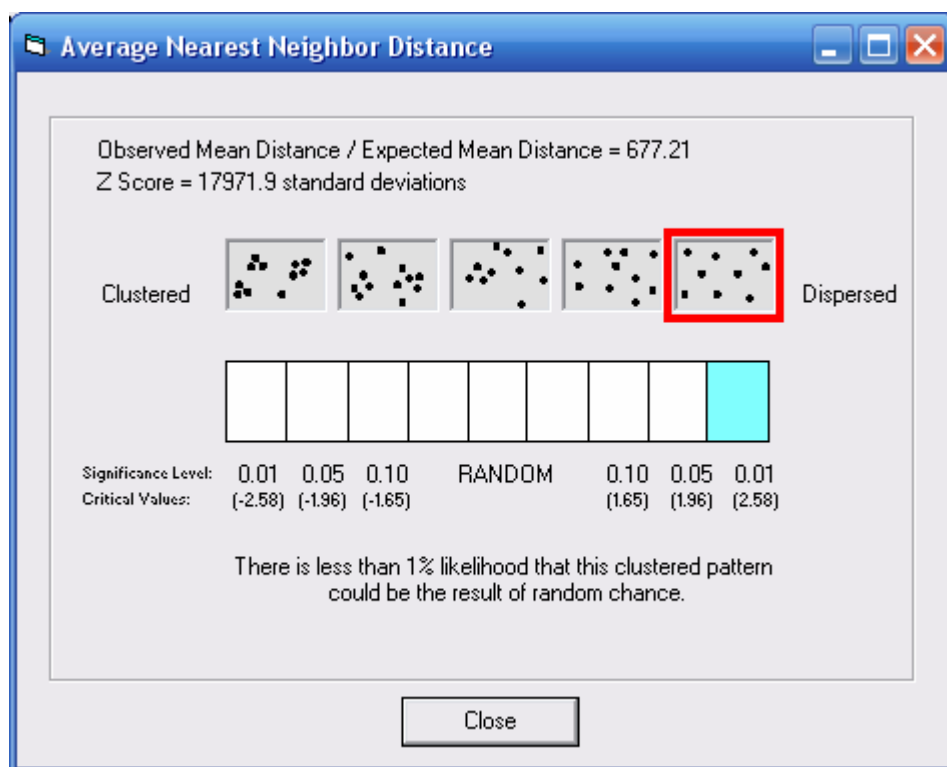
1、 平均最邻近距离工具（Average Nearest Neighbor）：

测量每个要素与之最邻近要素之间的距离，并计算平均值。再测量平均距离与假定为随机分布距离的相似程度。工具进行统计后返回 z score 值。Z score 值为负且越小，则要素分布越趋向于聚类分布，相反为离散分布。

该工具主要用于说明要素之间的接近程度以及它们之间的相互关系。



图一、平均最邻近距离工具面板



图二、平均最邻近距离工具图示面板



图三、平均最邻近距离工具运行结果窗口

提示：

- 用空间统计工具均以 Python 脚本或 Model 的方式提供给用户，所以必须要安装 Python 编译环境（在安装 ArcGIS 软件时，Python 默认情况下需要安装的）。

- 如果工具中提供了图形化显示输出结果 (Display Output Graphically) 选项, 推荐进行选中, 以便于观察形象的统计结果 (如图一、图二)。
- 对于一些 optional 的选项, 选择与否或是否进行设置对输出结果的影响是很大的, 推荐用户在对自己数据的充分认识的基础上, 进行合理的设置。
- 可以点击工具面板上的 “show help” 按钮, 显示该工具的简单描述信息。如果想获得更详细的信息, 点击描述信息上的 help 就会自动打开并跳转到 ArcGIS Desktop Help 中的该工具的详细信息。
- 请将运行结果面板的 Close this dialog when completed successfully 选项取消选中, 并激活 Details 按钮。这样, 统计的结果会在工具面板中进行显示 (如图三)。

2, 高/低聚类分析工具(High/Low Clustering ---Getis-Ord General G) :

测量特定区域的聚合程度。返回 General G Index 值和 Z Score 值。G Index 值越高, 越趋向于高聚类。相反为低聚类。Z 值为正且越大, 要素分布趋向高聚类分布。相反为低聚类分布。

应用案例: 查找城市中不同地区的人口聚集水平, 或人口随时间的聚集水平。

3, 空间自相关工具 (Spatial Autocorrelation --- Moran's I) :

Moran's I 方法是进行空间自相关统计的常用统计方法。该方法在 ArcGIS 9 中得到了实现。通过使用该工具, 会返回 Moran's I Index 值以及 Z Score 值。如果 z score 值小于 -1.96 或大于 1.96, 那么返回的统计结果就是可采信值。如果 z score 为正且大于 1.96, 则分布为聚集的。如果 z score 为负且小于 -1.96, 则分布为离散的。其他情况可以看作随机分布。

相关概念 :

- 空间相关性: 检测空间上两种现象 (统计量) 的变化是否存在相关性。比如, 水稻的产量往往与所处的土壤的肥沃程度相关。
- 空间自相关: 研究空间中, 某个空间单元与其周围的单元之间, 就某种特征, 通过统计方法, 进行空间相关性程度的计算, 以分析这些空间单元在空间上分布现象的特性。也就是说, 空间自相关研究的是不同观察对象的同一属性在空间上的相互关系。[1]

二、聚类分布制图工具集 (Mapping Clusters) :

1, 聚集及特例分析工具 (Cluster and Outlier Analysis --- Anselin Local Moran's I) :

使用该工具会输出一个新的要素类。该要素类在原要素类上添加了两个字段, 分别为 LMi<distance_method>和 LMz<distance_method>, 分别代表各个要素的索引值 I 和 z score 值。如果索引值 I 为正, 则要素值与其相邻的要素值相近, 如果索引 I 值为负值, 则与相邻要素值有很大的不同。如果 z score 为正且越大, 则要素越与相邻要素值相近, 相反, 如果 z score 值为负却越小, 则与相邻要素值差异越大 (也就是相关性不强)。

对于线类要素和多边形要素, 程序会计算要素的几何中心点, 然后对几何中心点进行分析。这就会造成有些要素的几何中心点不在几何要素内部, 如果想达到此目的, 请先用 Feature to Points 工具进行转换后再进行分析。

2, 聚集及特例分析并符号化 (Cluster and Outlier Analysis with Rendering) :

进行与聚集及边界分析工具相同的操作并对结果进行符号化。

3, 热点分析 (Hot Spot Analysis --- Getis-Ord G_i^*) :

对输入要素进行 Getis-Ord G_i^* 统计。并把统计结果作为新加字段 ($G_i^* < \text{distance band or threshold distance}$ 的输入值) 写入输出要素中。关于 Getis-Ord G_i^* 统计的方法请查阅相关文章。

该工具主要用于进行事件发生地区的预测或获取关注地区。比如通过对以往犯罪发生频率的统计, 推测可能再次发生的地点。

4, 热点分析并符号化工具 (Hot Spot Analysis with Rendering) :

进行热点分析工具相同的操作并对分析结果的 G_i 字段进行符号化, 生成一个存储了符号化方案的 layer 文件。

三、度量空间分布工具集

1, 中心要素 (Central Feature) :

查找距其所有要素距离最短的要素。使用该工具查找的是已存在的要素。

2, 方向性分布分析工具 (标准差椭圆) (Directional Distribution --- Standard Deviational Ellipse) :

标准差椭圆工具创建一个新的要素类。该要素类包含一个椭圆多边形要素。该要素包含以下属性: 椭圆的中心坐标。X、Y 方向上的标准距离, 也就是椭圆长轴、短轴的值。以及椭圆的旋转方向。使用椭圆可以概括要素的空间分布, 并识别方向的趋势。长轴为空间分布最多的方向, 短轴为空间分布最少的方向。

由于标准差椭圆工具会测量所有要素的 x、y 坐标与其坐标平均值之差的平均值, 所以分析结果是基于统计结果而不是目视理解。

该工具主要用于进行空间分布特征的方向性因素判定。比如通过对污染样点数据的分析, 找出污染扩散的主要方向。

3, 线性方向均值 (Linear Directional Mean) :

该工具要求输入要素为线要素。计算线要素的平均走向。所有的线都被看作从起点到终点两点构成的线, 线内结点都被忽略。方向均值的计算结果都是从正东开始, 以逆时针旋转的角度来表示的。还包括平均长度等统计信息。

该工具只要用来统计变化要素的主要发展趋势。比如统计一段时期的洋流运动方向, 来确定洋流的主要运动方向或趋势。

4, 均值中心 (Mean Center) :

该工具计算所有输入要素的平均 x、y 坐标, 生成一个新的要素。往往通过目视得出的结论是不准确的, 比如发生事件比较频繁的地区经常发生空间上的重叠, 而目视判断会忽略这些信息, 从而造成判断错误。而通过使用工具, 会对每个要素进行计算, 得出来的中心才是真正的均值中心, 是可信的结果。

5, 标准距离 (Standard Distance) :

对密集分布的测量可以提供一个表示中心周围要素离散度的值。这个值就是标准距离。通过计算要素的标准距离, 生成一个以标准距离为半径的圆, 来概括密集分布特征。该圆覆盖了要素的大部分。标准距离越大, 说明要素的分布越分散。如果数据中存在差异较大的极值, 会对结果产生比较大的影响。

四、辅助工具集 :

1, 计算面积 (Calculate Areas)。

计算多边形的面积。提供其他某些工具所需的面积参数。例如：平均最邻近距离工具对于统计区域的面积是敏感的。大多数情况需要指定分析区域的面积。即可从该工具获得该参数。

2, 事件收集 (Collect Events)。

收集同一空间位置时间发生的次数。比如同一地区犯罪发生次数，流行病发病次数等等。通常，在进行热点分析的时候，会有选项要求指定一个权重字段，而通过事件收集工具得到的 count 字段代表了时间发生的次数，可作为权重字段进行考虑。这种情况适用于对于事件发生情况的统计。比如流行病发病热点地区探索等等。

3, 收集事件并符号化 (Collect Events with Rendering)。

这是一个 Model 工具。进行与事件收集工具相同的工作并对其结果进行符号化，符号化字段为 count 字段。

4, 对 count 字段符号化工具 (Count Rendering)。

事件收集并符号化工具中已使用到该工具，只不过在 model 里进行了调用而已。该工具用于对 count 字段进行符号化并生成 Layer 文件。

5, 将要素属性值输出为 Ascii 码 (Export Feature Attribute to Ascii)。

该工具会将要素的 x, y 坐标及指定的属性值输出为 Ascii 码文件。

6, Z score 值符号化工具 (Z Score Rendering)。

对包含 z score 值字段进行符号化。以达到增强显示效果的目的。

总结：

空间统计工具集集合了常用的空间统计方法，这些方法都是基于统计学理论，同时充分考虑了事物分布的空间特性。可以说，空间统计工具架起了统计学与 GIS 的桥梁。通过使用 ArcGIS 9 中的空间统计工具，可以方便的进行常见的空间统计，还可以对其结果进行符号化，在 ArcMap 中进行观察。

空间统计理论及 ArcGIS 9 空间统计工具的具体应用，是 ESRI 中国（北京）培训中心所开设的《ArcGIS 高级分析教程》所探讨的重点内容。详细信息请登陆培训中心网站：<http://training.esrichina-bj.cn>。

参考文献：

- 1, 陈慈仁,林峰田,何灿群 空间自相关聚集分析 <http://www.bp.ntu.edu.tw>
- 2, ArcUser Online, Powerful Spatial Statistics Tools in ArcGIS 9, October-December 2004