“十二五”普通高等教育本科国家级规划教材

地理信息系统教程

Dili Xinxi Xitong Jiaocheng

（第二版）

主编：汤国安

编著：汤国安 刘学军 闾国年
盛业华 王　春 张海平

高等教育出版社

第九章：空间统计分析

本讲大纲

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.1 空间统计概述

当前大纲

9.1.1 基本概念

9.1.2 主要分析内容

9.1 空间统计概述

9.1.1 基本概念

❑ 空间统计分析

空间数据的统计分析着重于空间对象和现象的非空间特性的统计分析，中心议题是如何用数学统计模型来描述和模拟空间现象和过程。

数据的空间统计分析直接从空间对象的空间位置、联系等方面出发，研究具有随机性、结构性，或具有空间相关性和依赖性的自然现象。

9.1 空间统计概述

9.1.2 主要分析内容

空间统计分析与经典统计学的内容往往是交叉的。空间统计分析使用统计方法解释空间数据，分析数据在统计上是否是“典型”的，或“期望”的。同时，它又具有自己独有的空间自相关分析。主要分析内容包含以下几点：

- 基本统计量
- 探索性数据分析
- 空间插值
- 空间分类
- 空间回归

9.1 空间统计概述

9.1.2 主要分析内容

基本统计量：统计量是数据特征的反映，也是统计分析的基础。

探索性数据分析：探索性数据分析能让用户更深入了解数据，认识研究对象，从而对与其数据相关的问题做出更好的决策。探索性数据分析主要包括确定统计数据属性、探测数据分布、全局和局部异常值（过大值或过小值）、寻求全局的变化趋势、研究空间自相关和理解多种数据集之间相关性。

9.1 空间统计概述

9.1.2 主要分析内容

空间插值：基于探索性数据分析结果，选择合适的数据内插模型，由已知样点来创建表面并评估其不确定性，然后研究其空间分布。

空间分类：基于地图表达，采用与变量聚类分析相类似的方法来产生新的综合性或者简洁性专题地图。包括多变量统计分析，如主成分分析、层次分析，以及空间分类统计分析，如系统聚类分析、判别分析等。

空间回归：研究两个或两个以上的变量之间的统计关系，通过空间关系，包括考虑空间的自相关性，把属性数据与空间位置关系结合起来，更好地解释地理事物的空间关系。

当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.2 基本统计量

当前大纲

9.2.1 代表数据集中趋势的统计量

9.2.2 代表数据离散程度的统计量

9.2.3 代表数据形态的统计量

9.2.4 其它统计量

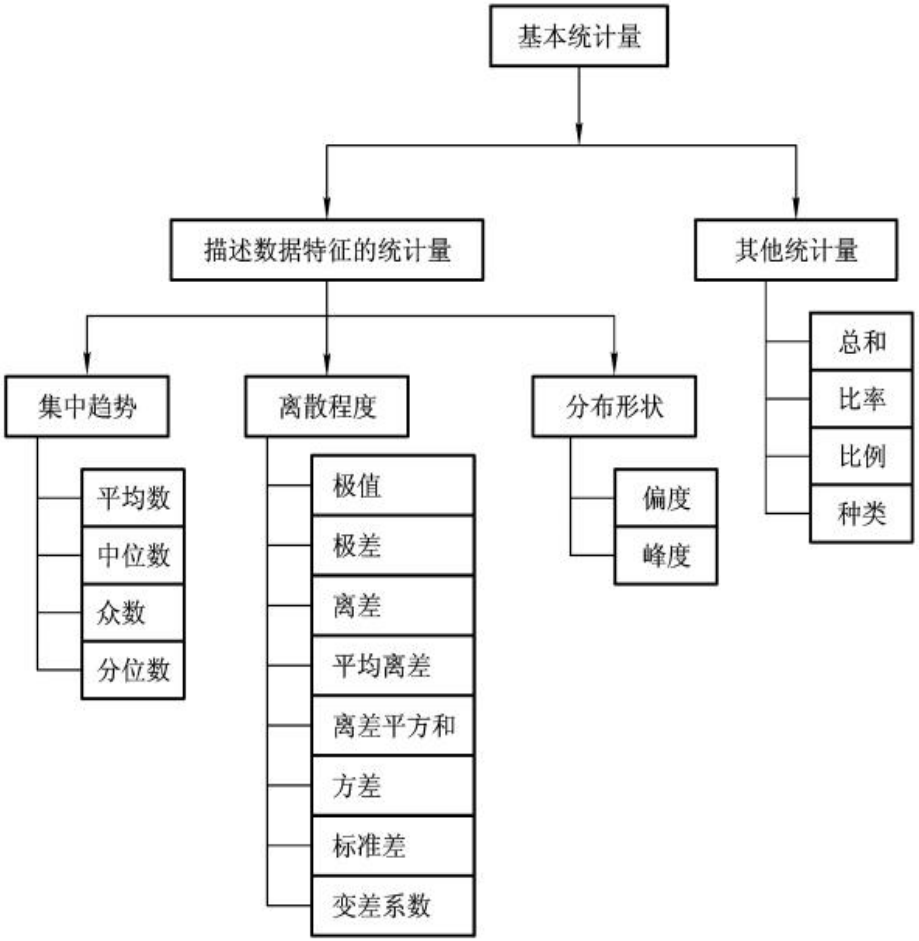
9.2 基本统计量

基本统计量概述

常用的基本统计量主要包括：

最大值、最小值、极差、均值、中值、总和、众数、种类、 离差、方差、标准差、变差系数、峰度和偏度等。

这些统计量反映了数据集的范围、集中情况、离散程度、空间分布等特征，对进一步的数据分析起着铺垫作用。



9.2 基本统计量

9.2.1 代表数据集中趋势的统计量

代表数据集中趋势的统计量包括平均数、中位数、众数，它们都可以用来表示数据的分布位置和一般水平。

平均数、中位数、众数在反映总体一般数量水平的同时，也掩盖了总体中各单位的数量差异。所以，只有这些统计量还不能充分说明一个数列中数值的分布情况和波动状态。

9.2 基本统计量

9.2.2 代表数据离散程度的统计量

代表数据离散程度的统计量包括最大值、最小值、分位数、极差、离差、平均离差、离差平方和、方差、标准差、变差系数等。

离散程度越大，数据波动性越大，以小样本数据代表数据总体的可靠性越低

离散程度越小，则数据波动性越小，以小样本数据代表数据总体的可靠性越高。

9.2 基本统计量

9.2.3 代表数据形态的统计量

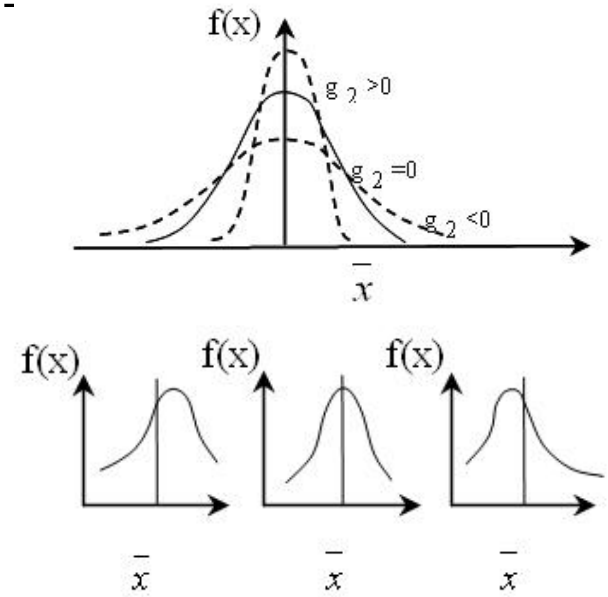
分布形态可以从两个角度考虑：

偏度或偏斜度：数据分布对称程度。

峰度：数据分布集中程度。

偏度和峰度是衡量数据分布特征的重要指标

对于标准正态分布
偏度=0
峰度=3



当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.3 探索性数据分析

当前大纲

9.3.1 基本分析工具

9.3.2 检验数据的分布

9.3.3 寻找数据的离群值

9.3.5 空间自相关与空间关系建模

9.3 探索性数据分析

探索性数据分析概述

探索性数据分析(Exploratory Spatial Data Analysis, ESDA)是以空间关联测度为核心，旨在描述空间数据的空间分布特征，发现离群值，揭示空间联系的结构，给出空间异质性的形式，从而引导建模。

首先分离出数据的模式和特点，再根据数据特点选择合适的模型。还可以用来揭示数据对于常见模型的意想不到的偏离。

探索性方法既要灵活适应数据的结构，也要对后续分析步骤揭示的模式灵活反应。

9.3 探索性数据分析

9.3.1 基本分析工具

❑ 直方图

对样本数据按一定的分级方案（等间隔分级、标准差等）进行分级，统计记录落入各个级别中的个数或占总样本数的百分比，然后用条带图或柱状图表现出来。

直方图可以直观反映采样数据分布特征、总体规律，可以用来检验数据分布和寻找数据离群值。

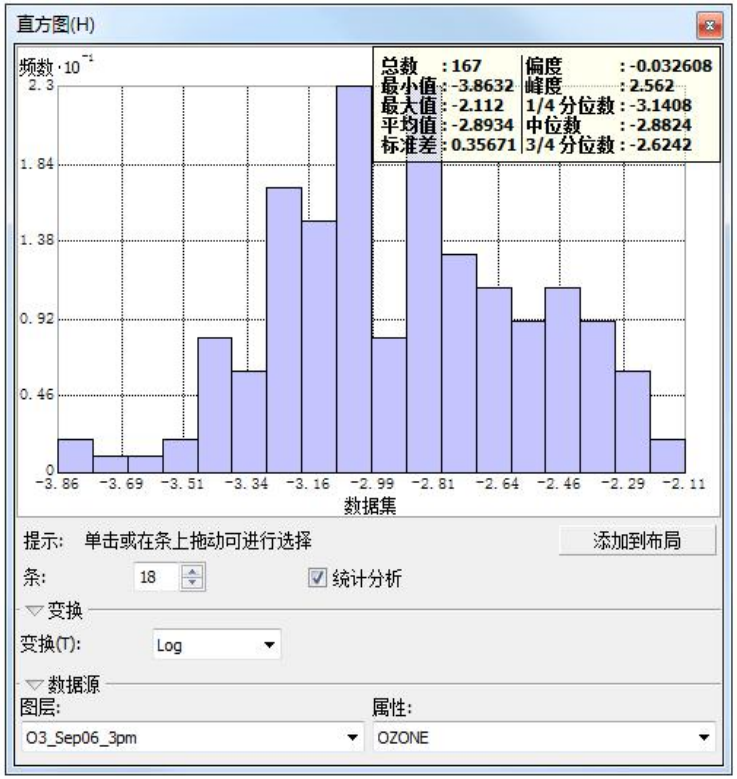
9.3 探索性数据分析

9.3.1 基本分析工具

- **直方图**

描述数据分布的重要特征包括中值、展布及对称性。

快速检验正态分布的方法：平均值与中值大致相等。

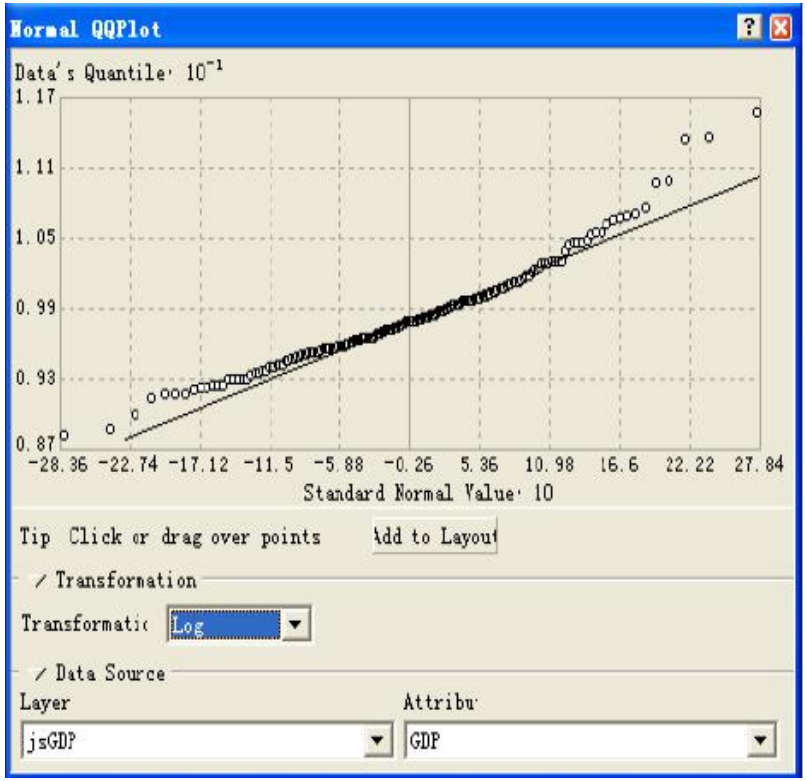


9.3 探索性数据分析

9.3.1 基本分析工具

❑ 正态QQ图

正态QQ Plot分布图：用来辅助判断样本数据是否服从正态分布。将数据的分布与标准正态分布对比，如果数据越接近一条直线，则越接近于服从正态分布。

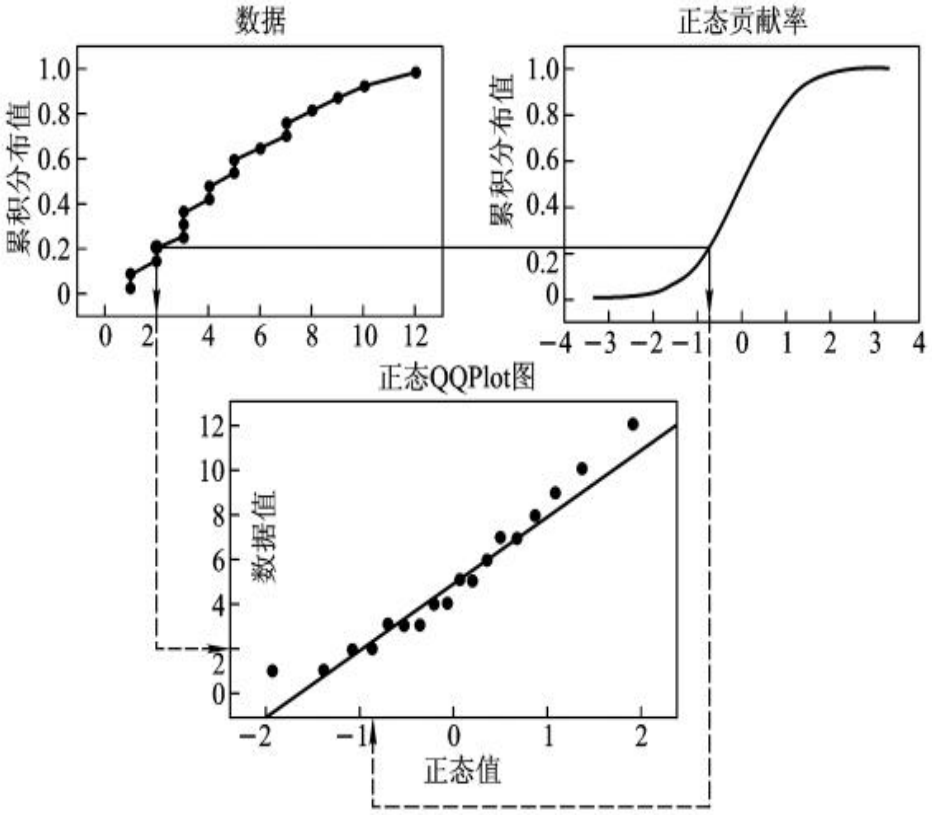


9.3 探索性数据分析

9.3.1 基本分析工具

- **正态QQ图**

构建QQ Plot分布图



数据排序

⬇

累计值
（低于该值的百分比）

⬇

累积值图

⬇

值之间线性内插

⬇

构建相同的正态分布

⬇

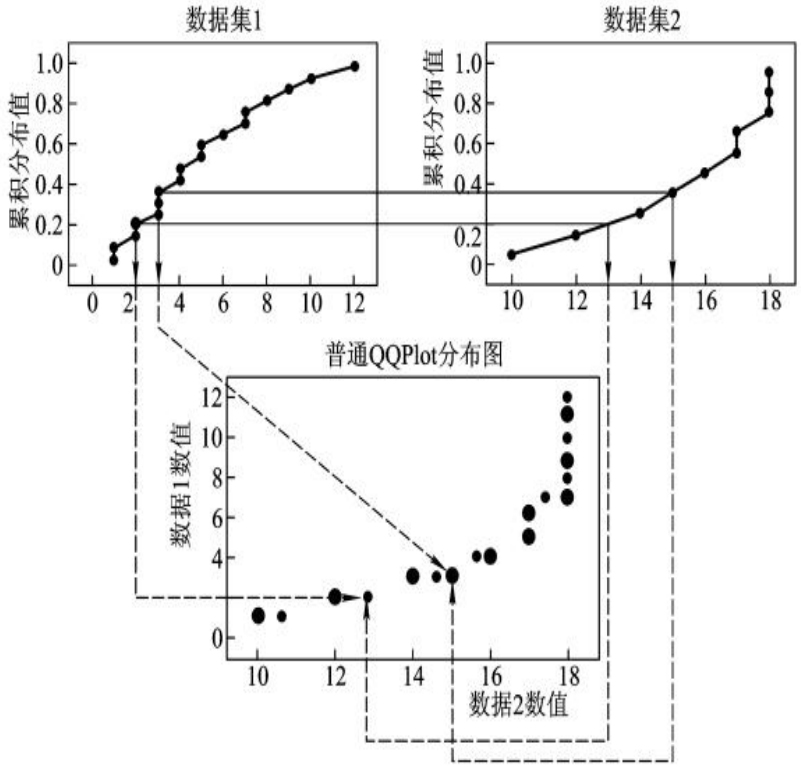
求出相同累积值对应的分布值

9.3 探索性数据分析

9.3.1 基本分析工具

❑ 正态QQ图

普通QQ Plot分布图：用来评估两个数据集的分布的相似性。普通QQ Plot分布图通过两个数据集中具有相同累积分布值作图来生成。



9.3 探索性数据分析

9.3.1 基本分析工具

❑ 方差变异分析工具

半变异函数和协方差函数把统计相关系数的大小作为一个距离的函数，是地理学相近相似定理的定量化。

协方差又称半方差，表示两随机变量之间的差异。

半变异函数又称半变差函数、半变异矩，是地统计分析的特有函数。

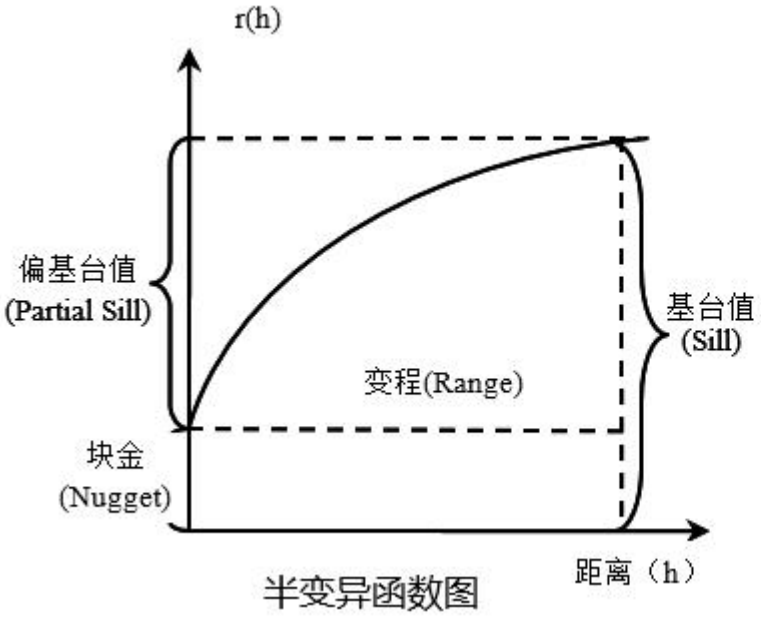
9.3 探索性数据分析

9.3.1 基本分析工具

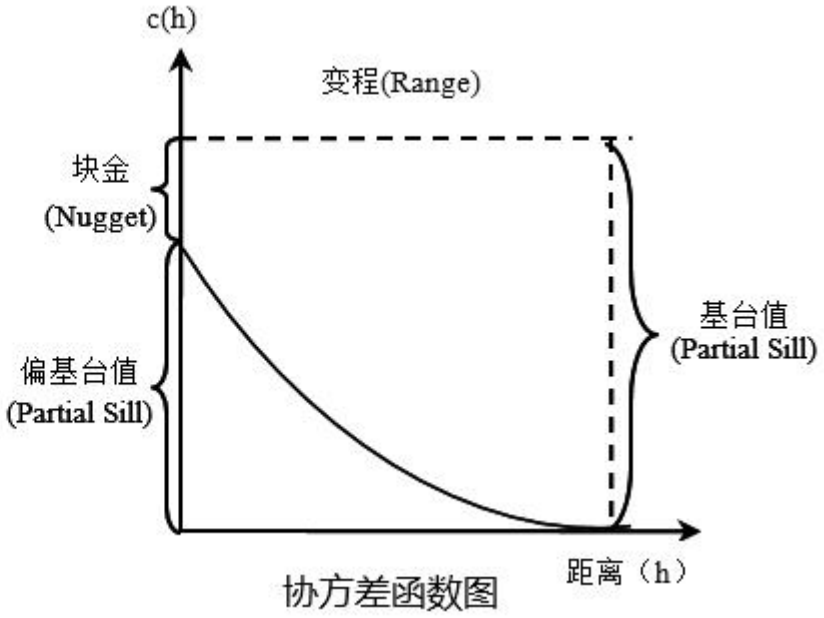
❑ 方差变异分析工具

半变异值的变化随着距离的加大而增加。

协方差随着距离的加大而减小。



半变异函数图



协方差函数图

9.3 探索性数据分析

9.3.1 基本分析工具

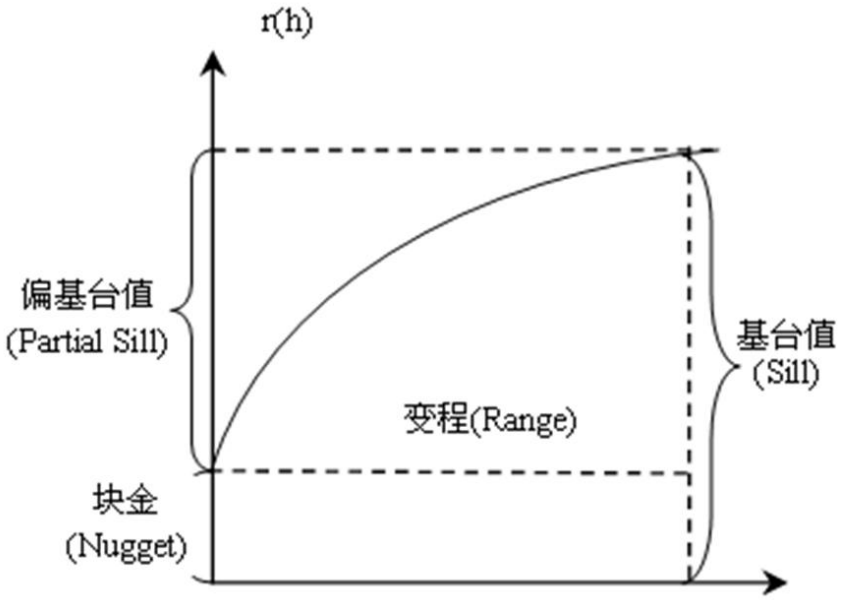
- **方差变异分析工具**

  半变异函数图

  间隔为0时的点

  半变异函数趋近平稳时的拐点

  由这两个点产生四个相应的参数：



- 块金值（Nugget）
- 变程（Range）
- 基台值（Sill）
- 偏基台值（Partial Sill）

9.3 探索性数据分析

9.3.1 基本分析工具

❑ 方差变异分析工具

块金值（Nugget）：理论上，当样点间的距离为0时，半变异函数值应为0；但由于存在观测误差和空间变异，使得两样点非常接近时，它们的半变异函数值不为0，即存在块金值

基台值（Sill）：当样点间的距离h增大时，变异函数r(h)从初始的块金值达到一个相对稳定的常数时，该常数值称为基台值

偏基台值（Partial Sill）：基台值与块金值的差值。

变程（Range）：当变异函数的取值由初始的块金值达到基台值时，样点的间隔距离称为变程

9.3 探索性数据分析

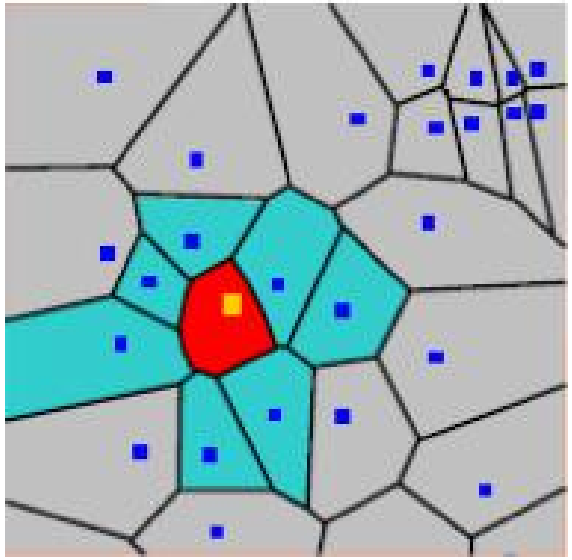
9.3.1 基本分析工具

❑ Voronoi 图

Voronoi地图是由在样点周围形成的一系列多边形组成的。

某一样点的Voronoi多边形的生成方法是：多边形内任何位置距这一样点的距离都比该多边形到其它样点的距离要近。

相邻点：和当前多边形相连的具有公共边的多边形内的点



9.3 探索性数据分析

9.3.1 基本分析工具

❑ Voronoi 图

Voronoi 图中多边形值可以采用多种分配和计算方法：

简化（Simple）：分配值为该多边形单元内样点的值

平均（Mean）：分配值为该单元与其相邻单元的平均值

模式（Mode）：所有的多边形单元被分成5级，分配值是此单元与相邻单元中出现频率最高的一级

聚类（Cluster）：所有的多边形单元被分成5级，若当前单元的级别与其相邻单元的级别都不同，则这个单元用灰色表示，以区别于其他单元

9.3 探索性数据分析

9.3.1 基本分析工具

❑ Voronoi 图

Voronoi 图中多边形值可以采用多种分配和计算方法：

熵（Entropy）：所有单元都根据数据值的自然分组分配到这五级中。分配到某个多边形单元的值是根据该单元和其相邻单元计算出来的熵。

中值（Median）：分配值为该多边形单元和其相邻单元的频率分布计算的中值。

标准差（StDev）：分配值为该多边形单元和其相邻单元的计算的标准差。

四分位数间间隔（IQR）：依据该多边形单元和其相邻单元的分布计算第一和第三分位数。分配值是第三分位数和第一分位数之差。

9.3 探索性数据分析

9.3.2 检验数据的分布

在空间统计的分析中，许多统计分析模型，都是建立在平稳假设的基础上，这种假设在一定程度上要求所有数据值具有相同的变异性。

另外，一些克立金插值（如普通克立金法、简单克立金法和泛克立金法等）都假设数据服从正态分布。如果数据不服从正态分布，需要进行一定的数据变换，从而使其服从正态分布。

在进行地统计分析前，检验数据分布特征，了解和认识数据具有非常重要的意义。

数据的检验可以通过直方图和正态QQ Plot分布图完成。如果数据服从正态分布，数据的直方图应该呈钟形曲线，在正态QQ Plot图中，数据的分布近似成为一条直线。

9.3 探索性数据分析

9.3.3 寻找数据的离群值

全局离群值指对于数据集中所有点来讲，具有很高或很低的值的观测样点。

局部离群值指对于整个数据集来讲，观测样点的值处于正常范围，但与其相邻测量点比较，它又偏高或偏低。

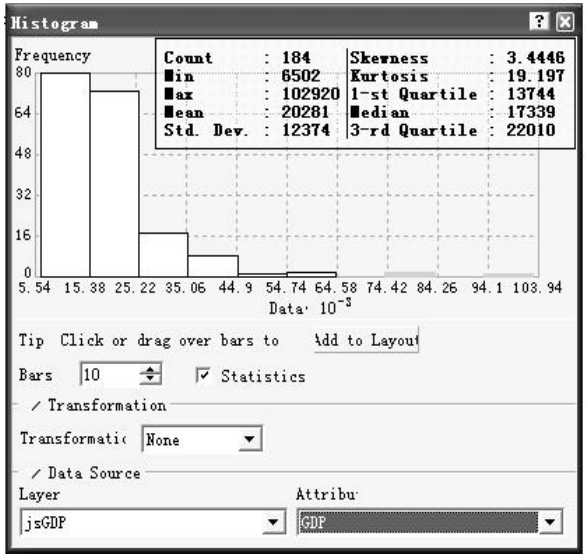
离群点的出现有可能就是真实异常值，也可能是由于不正确的测量或记录引起的。如果离群值是真实异常值，这个点可能就是研究和理解这个现象的最重要的点。反之，如果它是由于测量或数据输入的明显错误引起的，在生成表面之前，它们就需要改正或剔除。对于预测表面，离群值可能影响半变异建模和邻域分析的取值。

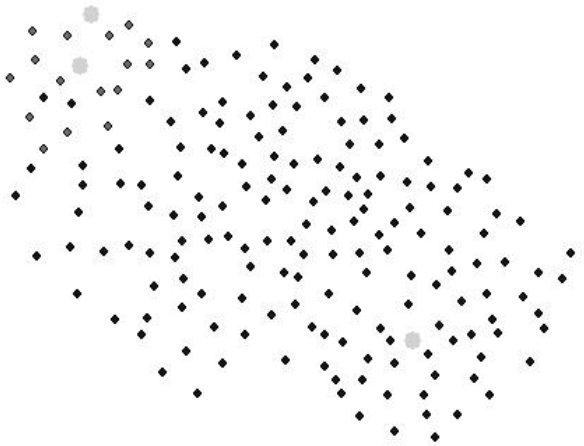
9.3 探索性数据分析

9.3.3 寻找数据的离群值

离群值的寻找方式：

- 利用直方图查找离群值
- 用半异/协方差函数云图识别离群值空间分类
- 用Voronoi图查找局部离群值





9.3 探索性数据分析

9.3.4 全局趋势分析

空间趋势反映了空间物体在空间区域上变化的主体特征，它主要揭示了空间物体的总体规律，而忽略局部的变异。

趋势面分析是根据空间抽样数据，拟合一个数学曲面，用该数学曲面来反映空间分布的变化情况。它可分为趋势面和偏差两大部分。

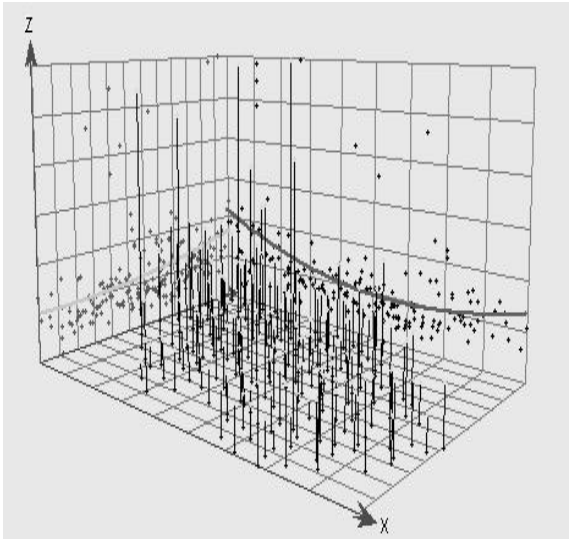
趋势面反映了空间数据总体的变化趋势，受全局性、大范围的因素影响。如果能够准确识别和量化全局趋势，在空间分析统计建模中就可以方便地剔除全局趋势，从而能更准确地模拟短程随机变异。

9.3 探索性数据分析

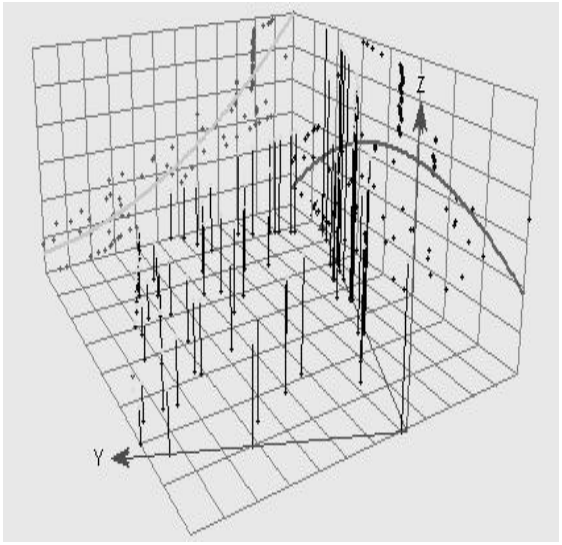
9.3.4 全局趋势分析

透视分析是探测全局趋势的常用方法，准确的判定趋势特征关键在于选择合适的透视角度。

同样的采样数据，透视角度不同，反映的趋势信息也不相同。



（a）



（b）

9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

空间自相关(spatial autocorrelation)分析包括全局空间自相关分析和局部空间自相关分析，自相关分析的结果可用来解释和寻找存在的空间聚集性或“焦点”。

空间自相关分析需要的空间数据类型是点或面数据，分析的对象是具有点(面)分布特征的特定属性。

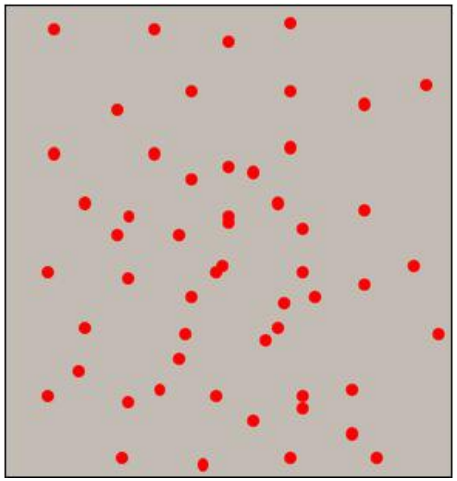
9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

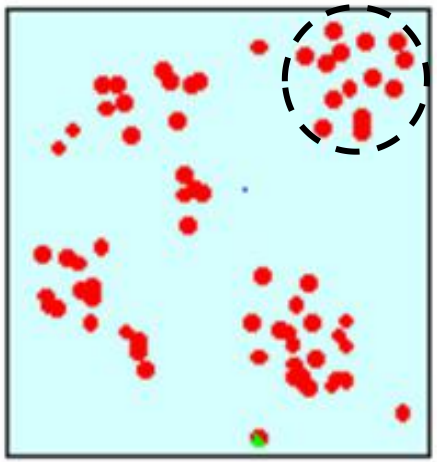
❑ 空间自相关的类型

全局空间自相关用来分析在整个研究范围内指定的属性是否具有自相关性。

局部空间自相关用来分析在特定的局部地点指定的属性是否具有自相关性。具有正自相关的属性，其相邻位置值与当前位置的值具有较高的相似性。



全局自相关



局部自相关

9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

❑ 空间权重矩阵

空间权重矩阵是研究空间自相关的基本前提之一。

空间数据中隐含的拓扑信息提供了空间邻近的基本度量，这通常可通过二元对称的空间权重矩阵 $W_{n\times n}$ 来表达:

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

9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

- **Moran’s I**

包括全程和局部两个参数，用来分析空间的相关性。

I值越大，表明正的空间相关性越强 。

对于全程空间自相关，Moran’s I定义是：

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

对于局部位置i的空间自相关，Moran’s I定义是：

$$I_i(d) = Z_i \sum_{j \neq i}^{n} w_{ij}' Z_j$$

9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

❑ Moran’s I

原假设是没有空间自相关。根据下面标准化统计量参照正态分布表可以进行假设检验。

$$Z_i = \frac{I - E(I)}{Var(I)}$$

Moran’sI如果是正的而且显著，表明具有正的空间相关性。 即在一定范围内各位置的值是相关的。

如果是负值而且显著的，则具有负的空间相关性，数据之间反相关。

接近于0则表明数据的空间分布是随机的，没有空间相关性。

9.3 探索性数据分析

9.3.5 空间自相关与空间关系建模

❑ Geray C系数

对于全局空间自相关：

$$C(d)=\frac{(n-1)\sum_{i}^{n}\sum_{j}^{n}w_{ij}(x_i-x_j)^2}{2nS^2\sum_{i}^{n}\sum_{j}^{n}w_{ij}}$$

对于局部位置i的空间自相关：

$$C_i(d)=\sum_{j\neq i}^{n}w_{ij}(x_i-x_j)^2$$

其中，Wij是空间权重矩阵。

C的值总是正的。假设检验是如果没有空间自相关，C的均值为1。显著性的低值（0和1之间）表明具有正的空间自相关，显著性的高值（大于1）表明具有负的空间自相关。

当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.4 空间数据常规统计与分析

当前大纲

9.4.1 空间数据分级统计分析

9.4.2 空间数据分区统计分析

9.4.3 样方统计与核密度估计

9.4 空间数据常规统计与分析

空间统计分析概述

空间统计分析是空间分析的核心内容之一。

在空间统计分析中，空间统计方法也有不同的类型划分。在对空间数据进行分析和可视化操作过程中，总会用到许多基本的空间统计方法。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

分级是对数据进行加工处理的一种重要方法，通过分级可以把数据划分成不同的级别，体现数据自身的特征，为应用研究及专题制图提供基础。

分级方法标准：

- 按使用分级方法的多少分级
- 按级差是否相等分级
- 按确定级差的方法分级

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按使用分级方法的多少可分为单一分级法和复合分级法

单一分级法是指对于一个数据集只用了一种分级方法。

复合分级法是指由于数据自身的特点，需要对一部分数据使用某种分级方法，对另一部分数据使用另外一种分级方法，才能更好地满足研究的需要。如一组坡度数据，一部分较小（坡面平缓），而另一部分很大（地势陡峭），对这两部分数据，就应选用两种不同的分级方法，才能更好地突出变化特征。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按级差是否相等可分为等值分级法和不等值分级法

等值分级法可以分为等面积分级、等间距分级、分位数分级等

不等值分级法可以分为自然裂点法分级、标准差分级、平均值嵌套分级等。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- **按确定级差的方法可分为自定义分级法和模式分级法**

自定义分级：对一个数据集，根据自己的应用目的设定各个级别的数值范围来实现分级的方法。适用于研究者对该数据集比较了解，能够找到合适的分级临界点。

模式分级：指按固定模式进行分级，在固定模式中，级差由特定的算法自动设定。

模式分级分为等间距分级、分位数分级、等面积分级、标准差分级、自然裂点法分级等。

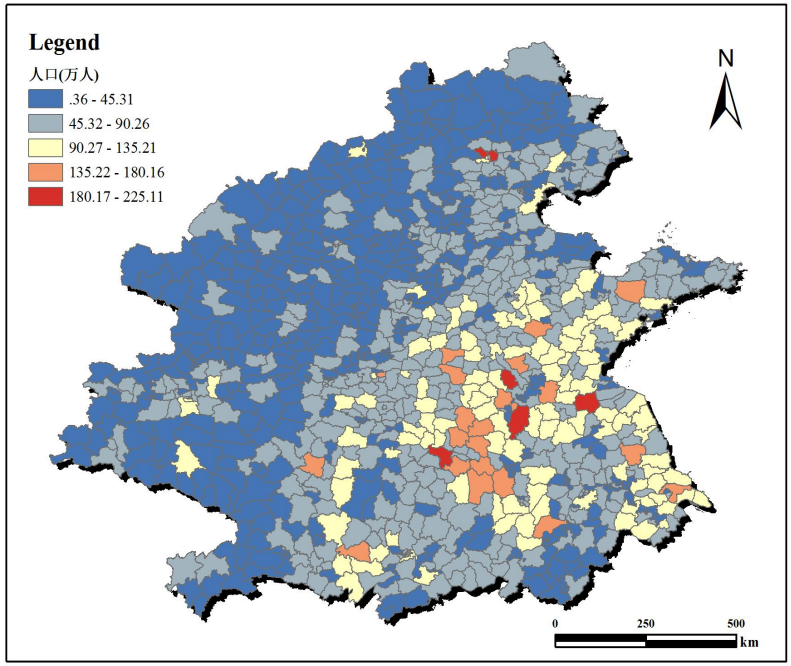
9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按确定级差的方法可分为自定义分级法和模式分级法

（1）等间距分级 一种最简单的分级方法，它按某个恒定间隔来对数据进行分级。假定数据集里有最大值和最小值，那么间距D=

$$\frac{\text{最大值}-\text{最小值}}{\text{分级数}}$$



原理简单、易操作，但当数据集中在某一小范围内时，各分级之间数据个数的差别太大会造成图面配置不均衡，影响了制图效果。

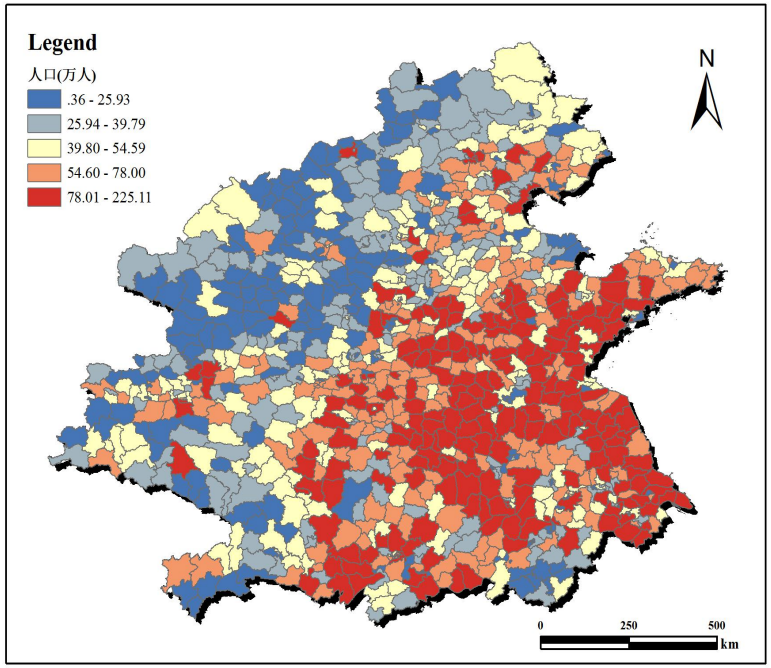
当数据具有均匀变化的分布特征时，等间距分级法就简明实用；若数据分布差异过大，将会以响制图与对统计结果的分析

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按确定级差的方法可分为自定义分级法和模式分级法

（2）分位数分级 把数列划分为相等个数的分段，根据实际需要选择四分位、五分位、六分位……十分位。为此，要先将数列按大小排列，从一端开始计算其分位数，把处于分位数上的那个值作为分级值。



分位数分级可以使每一级别的数据个数接近一致，往往能产生较好的制图效果。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- **按确定级差的方法可分为自定义分级法和模式分级法**

（3）等面积分级 使得每一级在图上占据的面积相等或大致相等。这种方法的特点是在图面上只反映各级占有相同的面积，制图效果好，但是没有充分利用图面表示级间的差异。

对于规则栅格数据而言，一定区域内的面积可由该区域内的栅格个数乘以栅格分辨率得到，所以按等面积分级只需考虑栅格个数即可。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- **按确定级差的方法可分为自定义分级法和模式分级法**

（4）标准差分级 标准差可以反映各数据间的离散程度，按标准差分级，首先要保证数据的分布具有正态分布的规律，才可计算平均值和标准差，然后根据数据波动情况划分等级。

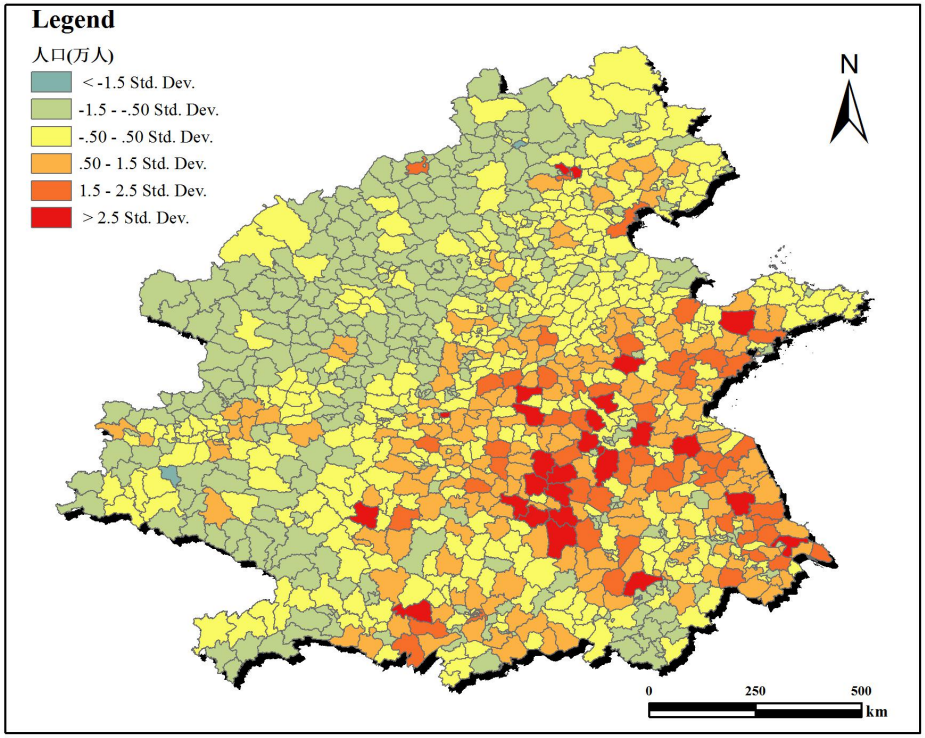
以算术平均值作为中间级别的一个分界点，以一倍或1/2倍标准差作为分界点。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按确定级差的方法可分为自定义分级法和模式分级法

(4) 标准差分级



9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- **按确定级差的方法可分为自定义分级法和模式分级法**

（5）自然裂点法分级 任何统计数列都存在着一些自然转折点、特征点，用这些点可以把研究的对象分成性质相似的群组，因此，裂点本身就是分级的良好界限。

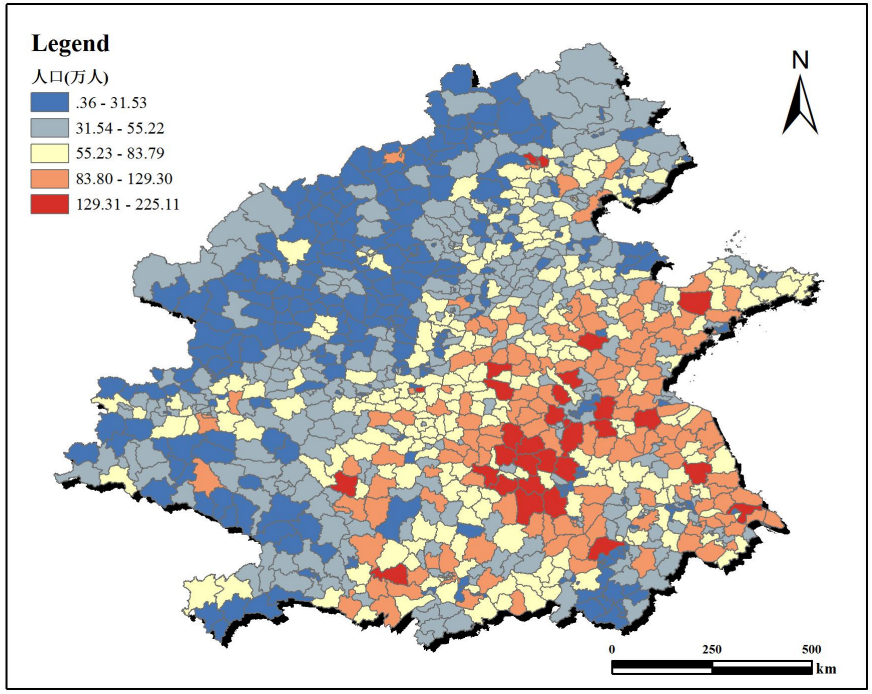
将统计数据制成频率直方图、坡度曲线图、积累频率直方图，有助于找出数据的自然裂点。如果频率最低点与峰值构成一个近似正态分布曲线，可以把任意两个正态分布曲线交点作为分级界线。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- 按确定级差的方法可分为自定义分级法和模式分级法

（5）自然裂点法分级



自然裂点法是基于让各级别中的变异总和达到最小的原则来选择分级断点的。

9.4 空间数据常规统计与分析

9.4.1 空间数据分级统计分析

- **按确定级差的方法可分为自定义分级法和模式分级法**

（6）其他分级方法 有规律的不等间距分级：这种方法与等间距分级法的区别在于它的间距是按一定规律变化的，而不是一个恒定的间隔。

该方法采用的间隔或级差有算术级数和几何级数两种，每种又都可通过以下6种变化方法来确定各级的分级间隔：按某一恒定速率递增、按某一加速度递增、按某一减速度递增、按某一恒定速率递减、按某一加速度递减、按某一减速度递减。

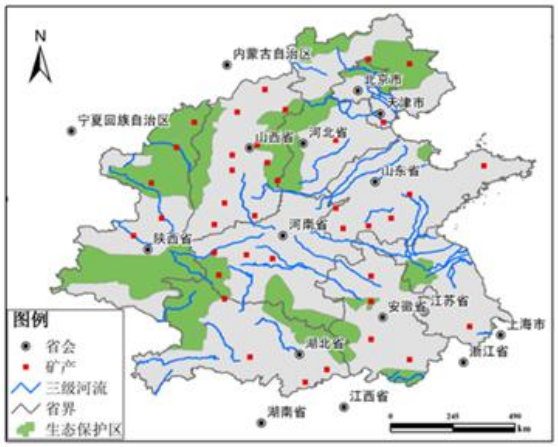
9.4 空间数据常规统计与分析

9.4.2 空间数据分区统计分析

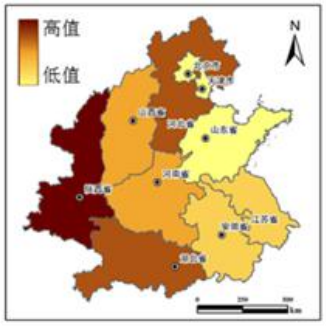
分区统计是将空间要素按照某种区域单元进行聚合的主要方法。

对同一主题的地理要素进行分区统计

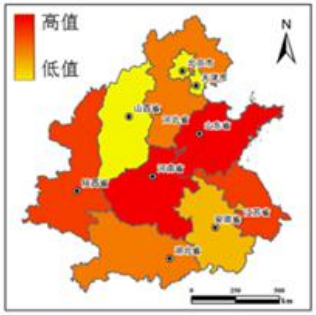
分区统计既可以用于统计区域单元内某种地理要素的数量特征、而且还可以统计其几何特征。



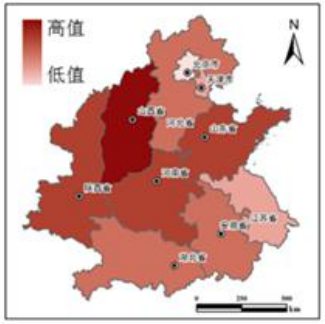
（a）地理要素分布图



（b）自然保护区面积分区统计图



（c）三级河流长度分区统计图



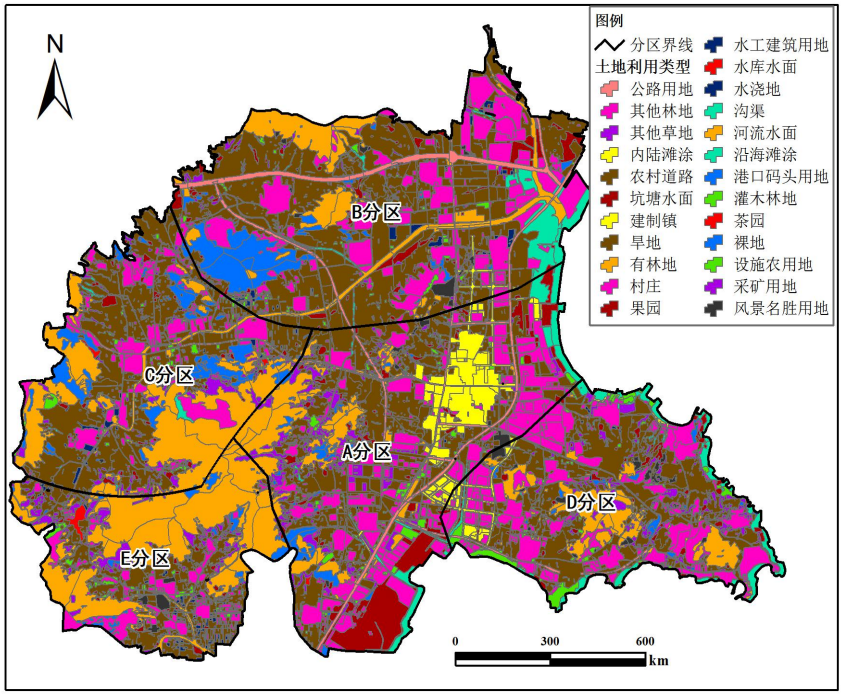
（d）矿产区数量分区统计图

9.4 空间数据常规统计与分析

9.4.2 空间数据分区统计分析

分区统计各个区域中不同主题要素的属性或几何特征。

例如某地区的土地利用图，研究区域共有5个分区，如果对每个区域中的各类用地进行面积汇总，则可以统计得到各个分区中各用地类型的面积汇总。



分区	类型	总占地面积
A分区	旱地	7563348.08
	林地	2534336.76
B分区	旱地	12152092.06
	林地	2012375.63
C分区	旱地	7192770.99
	林地	4493659.42
D分区	旱地	5702404.57
	林地	1848496.24
E分区	旱地	3078846.37
	林地	5682969.54

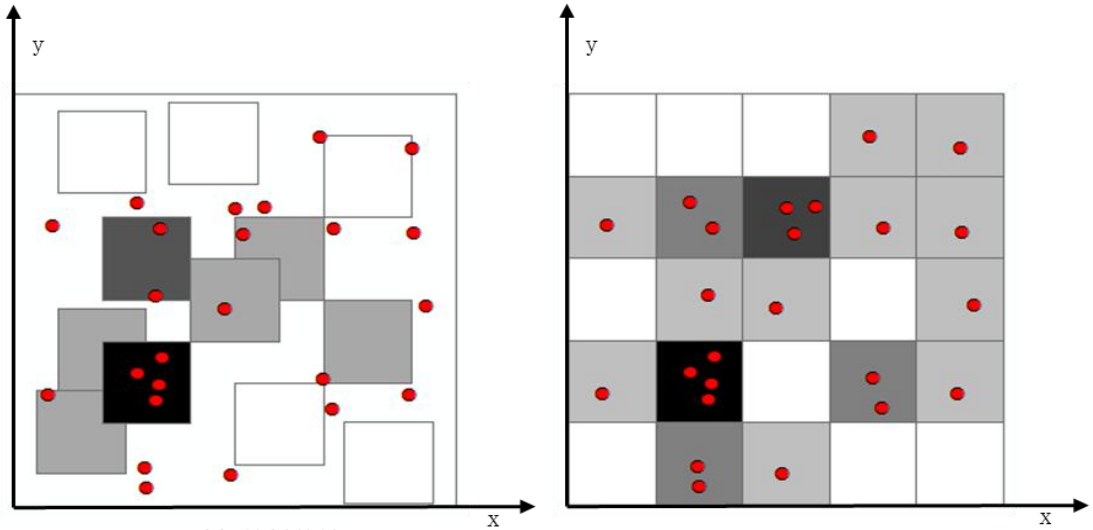
9.4 空间数据常规统计与分析

9.4.3样方统计与核密度估计

样方法(quadrat sampling method)是数据统计中应用较为广泛的方法。在非空间数据中，直方图就是一种典型的样方统计方法。

两种常见的样方空间统计方法：

- 随机抽样统计
- 利用所有值统计

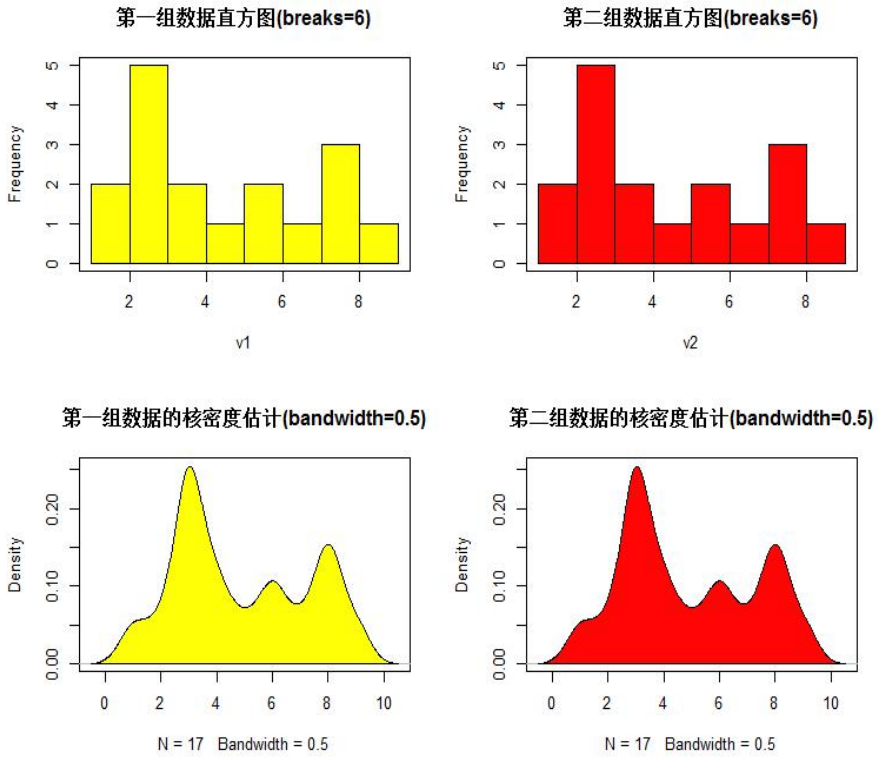


（a）随机抽样统计　　（b）利用所有值统计

9.4 空间数据常规统计与分析

9.4.3样方统计与核密度估计

基于概率密度函数的核平滑统计法为数据分布特征的描述提供了一种全新的方法。直方图和核密度图都较为准确地描述了两组数据的分布特征，但后者更为平滑，前者是离散的表达方式，而后者为连续的表达方式。



当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.5 空间插值

当前大纲

9.5.1 整体内插

9.5.2 局部分块内插

9.5.3 逐点内插法

空间插值类型

空间数据插值(spatial data interpolation)是进行数据外推的基本方法。空间内插的根本是对空间曲面特征的认识和理解，具体到方法上，即是内插点邻域范围的确定、权值确定方法（自相关程度）、内插函数的选择等3方面的问题。

分类依据	类型
是否考虑空间自相关	确定性插值和地统计插值
数据分布规律	基于规则分布数据的内插方法、基于不规则分布的内插方法和适合于等高线数据的内插方法等
内插函数与参考点的关系	曲面通过所有采样点的纯二维插值方法和曲面不通过参考点的曲面拟合插值方法
内插曲面的数学性质	多项式内插、样条内插、最小二乘配置内插等
对地形曲面理解	克立金法、多层曲面叠加法、加权平均法、分形内插等
内插点的分布范围	内插点的分布范围

9.5.1 整体内插

整体内插：在整个区域用一个数学函数来表达地形曲面。

整体内插函数通常是高次多项式，要求地形采样点的个数大于或等于多项式的系数数目。

当地形采样点的个数与多项式的系数相等时，这时能得到一个唯一的解，多项式通过所有的地形采样点，属纯二维插值；而当采样点个数多于多项式系数时，没有唯一解，一般采用最小二乘法求解，即要求多项式曲面与地形采样点之间差值的平方和为最小，属曲面拟合插值或趋势面插值。

9.5.1 整体内插

从数学角度讲，任何复杂的曲面都可用多项式在任意精度上逼近，但由于以下原因，在空间内插中整体内插并不常用。

- 整体内插函数保凸性较差。
- 不容易得到稳定的数值解。
- 多项式系数物理意义不明显。
- 解算速度慢且对计算机容量要求较高。
- 不能提供内插区域的局部地形特征。

9.5.2 局部分块内插

空间分块内插：将地形区域按一定的方法进行分块，对每一块根据地形曲面特征单独进行曲面拟合和高程内插。

区域分块简化了地形的曲面形态，每一块都可用不同的曲面进行表达，一般的可按地形结构线或规则区域进行分块，而分块大小取决于地形的复杂程度、地形采样点的密度和分布；为保证相邻分块之间的平滑连接，相邻分块之间要有一定宽度的重叠，另外一种分块之间的平滑连接是对内插曲面补充一定的连续性条件。

9.5.2 局部分块内插

不同的分块单元可用不同的内插函数，常用的内插数函数有：

- 线性内插
- 双线性内插
- 多项式内插
- 样条函数
- 多层曲面叠加法等

9.5.2 局部分块内插

❑ 线性内插和双线性内插

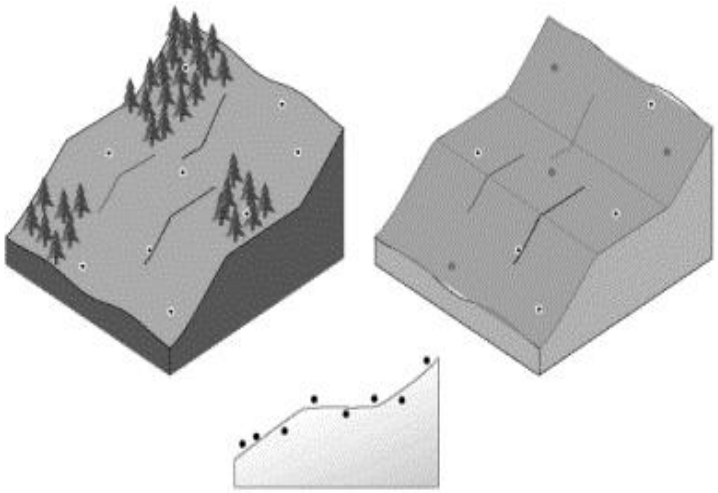
线性内插：如H=ax+by+c的多项式，它将分块单元内部的地形曲面视为平面。

双线性内插：在线性多项式中增加了交叉项xy，线性内插则变成双线性内插函数H=ax+by+cxy+d。

9.5.2 局部分块内插

❑ 线性内插和双线性内插

线性内插函数中有3个未知数，需要3个采样点才能惟一确定，而双线性内插函数中有4个未知数，需要4个已知点。



物理意义明确，计算简单，是基于TIN和基于正方形格网分布采样数据的DEM内插和分析应用的最常用的方法。

9.5.2 局部分块内插

❑ 二元样条函数内插

样条曲面可假想为将一张具有弹性的薄板压定在各个采样点上，而其他的地方自由弯曲。从数学上讲，就是一个分段的低次多项式，多项式的次数一般不超过三阶。通过样条函数，可以获取在各个采样点上具有最小曲率的拟合曲面。

与整体内插函数相比，样条函数不但保留了局部地形的细部特征，还能获取连续光滑的DEM。具有较好的保凸性和逼真性，同时也有良好的平滑性。

9.5.2 局部分块内插

❑ Coons曲面与Geomap曲面

Coons曲面是基于任意四边形的曲面拟合方法。

可用于由地形线围成的地貌形态单元。但Coons曲面仅考虑了曲边四边形的边界曲线，而没有考虑曲面内部的信息，对于恰当描述地貌形态有一定缺陷。

Geomap曲面是Bezier曲面在不规则格网划分上的推广形式。本质上，Coons和Geomap属于同一类曲面拟合问题Geomap曲面在地形上应用具有与Coons曲面类似的不足。

9.5.2 局部分块内插

❑ 多层曲面叠加内插

多层曲面叠加法认为任何一个规则或不规则的连续曲面都可看成由若干个简单的曲面来叠加逼近。

在每个数据点上建立一个曲面，然后在垂直方向上将各个曲面按一定比例进行叠加，形成一张整体连续的曲面，曲面严格通过每一个数据点。多层曲面叠加法的核心是简单曲面的设计，也称为核函数。

已经发展了许多种核函数的设计方法，如锥面、双曲面、三次曲面、高斯曲面（以高斯曲线为母线的旋转曲面）、Authur法、吕言法、Wild法等。

9.5.2 局部分块内插

❑ 最小二乘配置

最小二乘配置是一种基于统计的内插和测量数据处理方法，它认为一个测量数据一般由 3部分构成，即趋势、信号和误差。

包括最小二乘内插、最小二乘滤波和最小二乘推估。

核心问题是如何建立数据之间的协方差矩阵，换句话说，就是如何解决信号的相关性规律问题。在连续表面内插中，

最小二乘配置认为，数据点之间的相关规律仅与距离有关， 也就是说，距离越近，协方差越大，超过一定的距离，协方差趋于零。

9.5.2 局部分块内插

❑ 克立金法

克立金法（Kriging）与最小二乘配置比较类似，也是将变量的空间变化分为趋势、信号与误差3个部分，求解过程也比较相似。

不同之处在于所采用的相关性计算方法上，最小二乘采用协方差矩阵，而克立金法采用半方差，或者称为半变异函数。

克立金法的内蕴假设条件是区域变量的可变性和稳定性，也就说，一旦趋势确定后，变量在一定范围内的随机变化是同性变化，位置之间的差异仅仅是位置间距离的函数。通过不同数据点之间半方差的计算，可作出半方差随距离的变化的半方差图，从而用来估计未采样点和采样点之间的相关系数，进而取出内差点的高程。

9.5.2 局部分块内插

❑ 有限元内插

有限元内插是以离散方式处理连续变化量的数学方法，

其基本思路是将地形曲面分割成有限个单元的集合，单元形状可为三角形、正方形等。

节点：相邻单元边界的端点（顶点、几何中心、边的中心）。

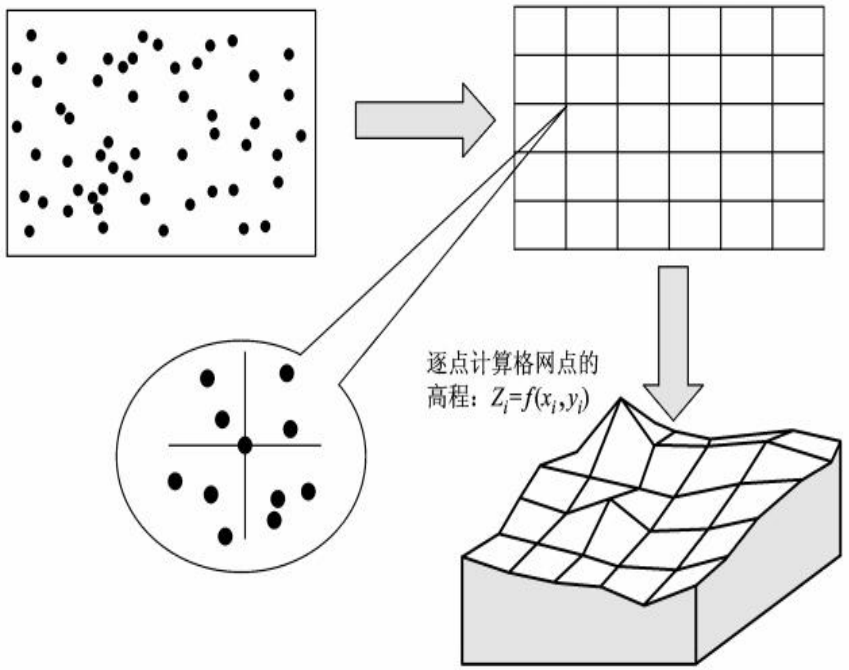
自由度：每个插值条件就是一个自由度。

有限元通常采用分片光滑的奇次样条函数作为单元的内插函数（也称为基函数），有限元的解是一系列基函数的线性组合。

9.5.3 逐点内插法

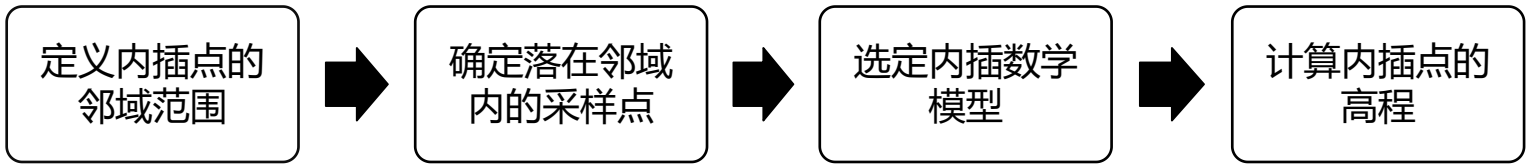
逐点内插是以内插点为中心，确定一个邻域范围，用落在邻域范围内的采样点计算内插点的高程值。

本质上是局部内插，但与局部分块内插有所不同，局部内插中的分块范围一经确定，在整个内插过程中其大小、形状和位置是不变的，凡是落在该块中的内插点，都用该块中的内插函数进行计算，而逐点内插法的邻域范围大小、形状、位置乃至采样点个数随内插点的位置而变动，一套数据只用来进行一个内插点的计算。



9.5.3 逐点内插法

逐点内插法的基本步骤为：



为实现上述步骤， 逐点内插法需要解决好以下几个问题：

- 内插函数
- 邻域内数据点的个数
- 采样点的分布
- 邻域大小和形状
- 采样点的权重
- 附加信息的考虑

当前内容

9.1 空间统计概述

9.2 基本统计量

9.3 探索性数据分析

9.4 空间数据常规统计与分析

9.5 空间插值

9.6 空间统计与空间关系建模

9.6 空间统计与空间关系建模

当前大纲

9.6.1 空间分布特征统计

9.6.2 空间分布模式挖掘

9.6.3 空间关系建模与探测

空间统计分析和关系建模概述

基于空间数据的空间统计分析与关系建模是GIS空间统计分析中除地统计分析外的另一重要组成部分。

包含了一系列用于分析空间分布、模式、过程和关系的统计工具。

9.6.1 空间分布特征统计

基于空间数据的空间统计，可通过度量一组要素的分布来计算各类用于表现分布特征的值，也可利用此特征值对一段时间内的分布变化进行追踪或对不同要素的分布进行比较。

中级分布特征统计能够帮助了解并定量地描述要素的地理分布特征。

常见的空间分布特征统计量包括一组地理要素的平均中心、中位数中心、中心要素、线性方向平均值、标准距离和方向分布等。

9.6.1 空间分布特征统计

❑ 平均中心

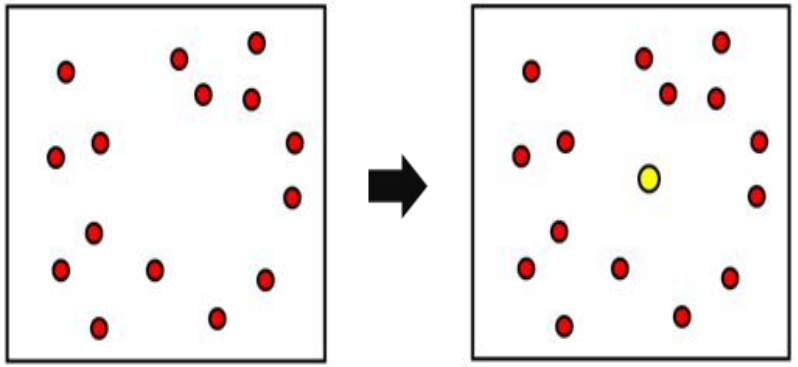
平均中心是研究区域中所有要素的平均 x 坐标和 y 坐标。

分析追踪分布的变化，以及比较不同类型要素的分布。

通过计算平均中心得到一组点、线、面或体的平均中心所在的位置。计算公式：

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\bar{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad , \quad \bar{Y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$



其中X¯和X¯分别为平均中心的坐标值，而¯Xw和¯Yw为加权后的平均中心坐标值。xi、yi和wi分别为要素i的坐标和权重值，n为要素总数。

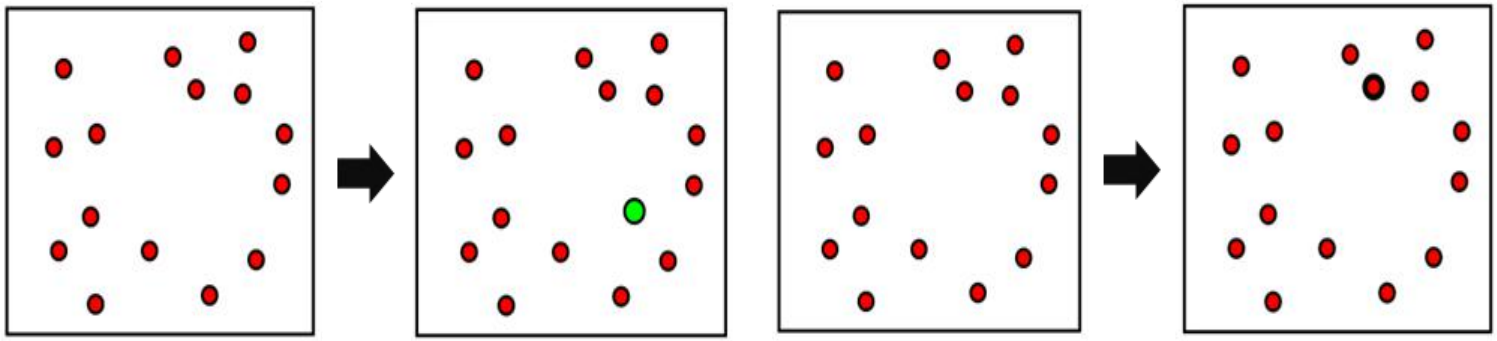
9.6.1 空间分布特征统计

❑ 中位数中心和中心要素

中位数中心是一种对异常值反应较为稳健的中心趋势的量度。可标识数据集中到其他所有要素的行程最小的位置点。

用于计算中位数中心的方法是一个迭代过程。

中心要素用于识别点、线或面输入要素类中处于最中央位置的要素。

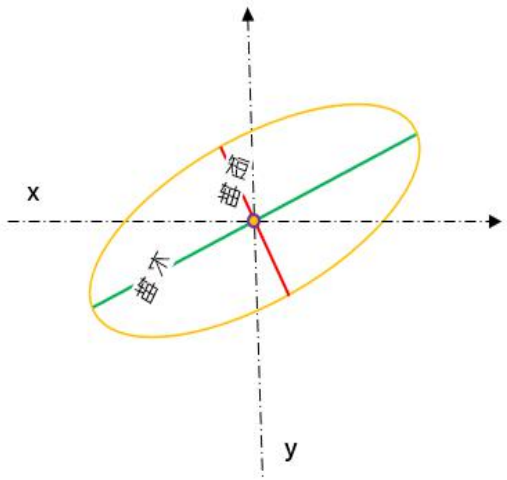


9.6.1 空间分布特征统计

❑ 标准差椭圆

标准差椭圆为一组数据的整体聚类（离散）趋势和方向分布特征的度量提供了有效的方式。

椭圆扁率体现了方向趋势的强弱程度，扁率越大，方向趋势越明显。椭圆大小体现数据的聚集或离散程度，椭圆越大，其分布越离散。长轴方向即为要素组的总体分布方向。



9.6.2 空间分布模式挖掘

识别地理模式对于理解地理现象非常重要。

尽管可以通过对要素制图来了解它们的总体模式及其关联值，但通过计算统计数据能够将模式量化。这样更便于比较不同分布方式或不同时段的模式。通常会先使用“分析模式”工具集中的工具进行初始分析，然后再进行更深入的分析。

分布模式包括全局和局部两个层面的模式挖掘：

- 全局模式统计
- 局部模式统计

9.6.2 空间分布模式挖掘

全局模式统计可提供对宏观空间模式进行量化的统计数据。解答“数据集中的要素或与数据集中要素关联的值是否发生空间聚类？”和“聚类程度是否会随时间变化？”之类的问题。

局部模式统计可通过执行聚类分析来识别具有统计显著性的热点、冷点和空间异常值的位置。

当根据一个或多个聚类的位置需要执行行动时，其用途特别明显。全局分析中的方法只对“是否存在空间聚类？”这样的问题回答“是”或“否”，与此不同的是，局部分析可以直观呈现聚类位置和范围。这些模型所解答的问题是“聚类（热点/冷点）的出现位置在哪里？”、“空间异常值的出现位置在哪里？”和“哪些要素十分相似？”。

9.6.2 空间分布模式挖掘

❑ 全局模式分析统计量

全局模式分析主要包括临近度、Moran’s I莫兰指数、Geary C、G-Statistics和格林系数等统计量，每种统计量能够识别全局模式的能力各不相同。

平均最邻近（Average Nearest Neighbor）只度量空间要素本身之间的邻近性，即只根据要素位置来度量空间邻近性。

全局莫兰空间自相关（Global Moran's I）指数根据要素位置和要素值来度量空间自相关。

G统计量高低聚类（Getis-Ord General G）可针对指定的研究区域测量高值或低值的聚集程度密度。

9.6.2 空间分布模式挖掘

❑ 局部分析统计量

在全局模式统计量中，可以通过空间自相关和高低值聚类对一组要素的全局模式进行分析和描述。

在局部分析统计量中，也包括用与之对应的用于局部统计分析的统计量。基于莫兰指数的聚类和异常值分析可识别具有高值或低值的要素的空间聚类.

9.6.2 空间分布模式挖掘

❑ 局部分析统计量

基于G-Statistics 格林系数的热点分析可对数据集中的每一个要素计算Getis-Ord Gi*统计量。通过得到的z得分和p值，我们可以知道高值或低值要素在空间上发生聚类的位置。

9.6.3 空间关系建模与探测

除了分析空间模式之外，GIS空间统计分析还可用于挖掘或量化要素间关系。

使用空间权重矩阵或利用回归分析可以建立空间关系模型。通常，空间关系模型通过回归模型实现。

在GIS中，较为常用的空间回归模型如地理加权回归，近年来，由我国学者开发的地理探测器（GeoDetector）的使用也越来越广泛。

9.6.3 空间关系建模与探测

❑ 地理加权回归

地理加权回归 (spatial weights matrix，GWR) 是若干空间回归技术中的一种。

通过局部区域建立使回归方程拟合适合数据集中的每个要素的不同变量之间的关系。

有助于对了解/预测的变量或过程提供局部模型。

GWR 构建这些独立方程的方法是：将落在每个目标要素的带宽范围内的要素的因变量和解释变量进行合并。

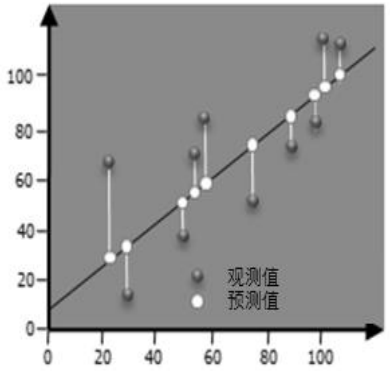
9.6.3 空间关系建模与探测

❑ 最小二乘法

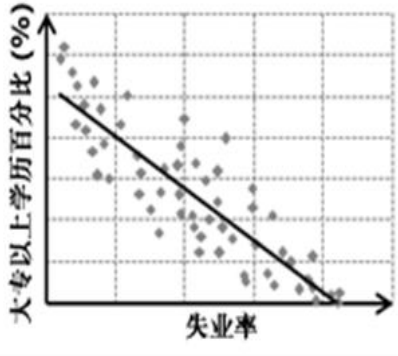
最小二乘法(OLS) 是所有空间回归分析的正确起点。

可创建一个回归方程来表示该过程。例如失业率与大专以上学历人数的关系。

地理加权回归使用OLS实现。若使用得当，这些方法可提供强大且可靠的统计数据，以对线性关系进行检查和估计



(a)



(b)

9.6.3 空间关系建模与探测

地理加权回归与最小二乘法比较

OLS属于全局空间回归模型。

GWR则属于局部空间回归模型。

对于空间问题，由于空间变量在局部区域的相似性和全局区域的异质性，很多情况下很难通过一个全局的OLS线性回归拟合出能够表示这些变量之间关系的线性模型，这就需要通过在不同的区域建立不同的线性回归模型对这些变量的关系进行建模，此时GWR便可以解决这类问题。

9.6.3 空间关系建模与探测

❑ 地理探测器

空间分层异质性简称空间分异性或区异性，是指层内方差小于层间方差的地理现象。

地理探测器（Geodetector）是探测空间分异性，以及揭示其背后驱动力的一组统计学方法。其核心思想是基于这样的假设：如果某个自变量对某个因变量有重要影响，那么自变量和因变量的空间分布应该具有相似性。既可以探测数值型数据，也可以探测定性数据，这正是地理探测器的一大优势。另一个独特优势是探测两因子交互作用于因变量。

9.6.3 空间关系建模与探测

❑ 地理探测器

地理探测器主要包括四个探测器，分别是分异及因子探测、交互作用探测、风险区探测和生态探测。

分异及因子探测用于探测要素属性Y的空间分异性；以及探测某因子X多大程度上解释了属性Y的空间分异）。用q值度量，表达式为：

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST}$$

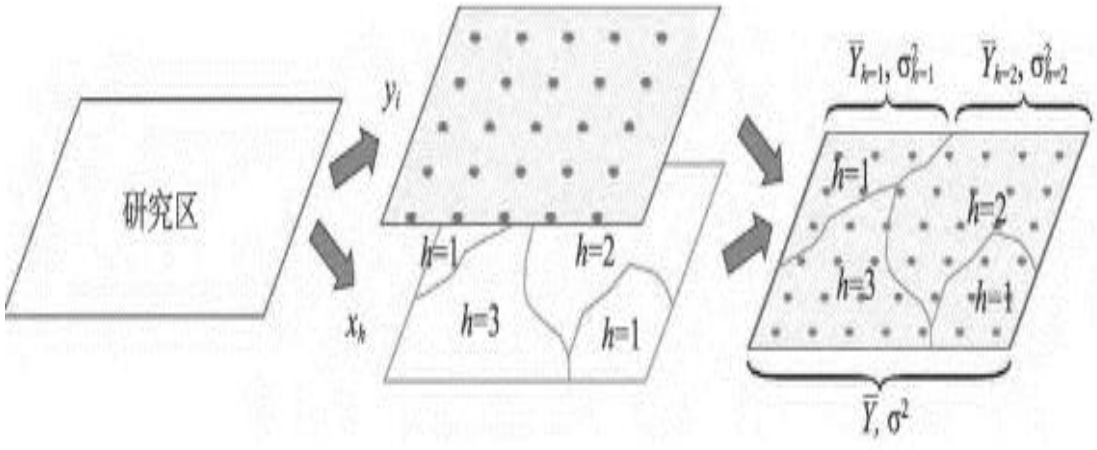
$$SSW = \sum_{h=1}^{L} N_h \sigma_h^2, \quad SST = N\sigma^2$$

式中：h= 1,...,L为变量Y或因子X的分层（Strata），即分类或分区；N_h和N分别为层h和全区的单元数；σ_h^2和σ^2分别是层h和全区的Y值的方差。SSW和SST分别为层内方差之和 （Within Sum of Squares）和全区总方差（Total Sum of Square）。

9.6.3 空间关系建模与探测

❑ 地理探测器

q的值域为[0, 1]，值越大说明Y的空间分异性越明显；如果分层是由自变量X生成的，则q值越大表示自变量X对属性Y的解释力越强，反之则越弱。极端情况下，q值为1表明因子X完全控制了Y的空间分布，q值为0则表明因子X与Y没有任何关系，q值表示X解释了100×q%的Y。其原理如下：



专业术语与思考题

专业术语

空间统计分析、直方图、空间数据探索性分析、协方差函数、半变异函数、Voronoi图、空间自相关、空间内插、回归分析、自然断裂法、核密度、样方统计、地理加权回归、地理探测器

专业术语与思考题

复习思考题

一、思考题（基础部分）

1、解释下列概念的含义：统计分析、空间统计分析、空间自相关、空间插值。

2、什么叫探索性数据分析？探索性数据分析的目的是什么？

3、空间自相关问题使用什么参数进行分析？其结果的地理解释是什么？

4、怎么解释变异函数图？

5、什么是空间回归，与经典的回归有什么差异？

专业术语与思考题

复习思考题

二、思考题（拓展部分）

1、探索性数据分析的内容有哪些？试以所在省（自治区、直辖市）各地级市的往年的GDP数据和人口数据为基本数据，对其作探索性数据分析，并给出分析结果的地理解释。

2、结合具体的数据比较距离倒数加权法、趋势面法、样条函数法、克立金法这 4 种插值的优缺点和各自的适用范围。

3、试以所在省市各地级市往年的GDP数据（离散点数据）和人口数据，请选择合适的内插模型，创建该地区GDP空间分布曲面。请绘制流程图说明你的分析思路和分析依据，并从内插模型、曲面的地理意义两方面给出自己的见解。