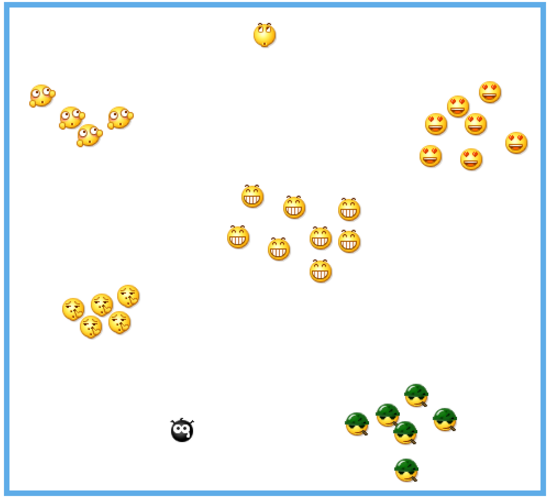
**空间统计博客**

**空间自相关**

**地理学第一定律**：任何事情都是有关系的，只不过靠的越近，关系越紧密。空间中的每一个事务，都是有联系的，近的事务之间的联系紧密程度，要高于距离远的事务之间的联系程度。所谓的联系紧密程度，自然也可以说，两个事务会在某一方面，有相似的地方。

**空间自相关性**：下了早操的学生



衡量空间自相关性：Moran’s （莫兰指数）

**莫兰指数**：Moran's I >0表示空间正相关性，其值越大，空间相关性越明显，Moran's I <0表示空间负相关性，其值越小，空间差异越大，否则，Moran's I = 0，空间呈随机性。

空间上面的正相关，就是指随着空间分布位置（距离）的聚集，相关性就也就越发显著。空间上的负相关就正好相反了，随着空间分布位置的离散，反而相关性变得显著了。

**空间异质性**

**空间异质性：**因为观察位置的不同，而产生的不同观测结果。

它无法用一个具体的的指数或指标来描述。更好的形容他的话，他是一种性质，一种现象，或者说是一种在我们的分析过程中需要充分考虑的因素，以及一种解释某些异常变化的原因。

一般来说，空间异质性，会用来解释，在不同的区域，某些类别数值相互之间的关系产生变化的原因，揭示这个变化的规律或者原因产生积极的作用。

**空间差异和空间异质性**是不同的概念

空间差异（spatialdisparity）是指不同地域范畴因为（社会、经济等）发展水平及其结构不同，而产生的差异。

而空间异质性（spatialheterogeneity）是指因为空间位置的不同而引发的获取到不同的数据。

**P值和Z值**

**零假设：**在空间统计中，零假设指的就是空间位置在一定区域里面呈现完全随机（均匀）分布（在自然现象里面，均匀分布是极小概率才会出现的，基本上都可以忽略了，所以一般谈的就是完全随机）。

**空间分布模式**：离散、随机、聚合

**p值（P-Value，Probability，Pr）**，代表的是概率。它是反映某一事件发生的可能性大小。在空间相关性的分析中，p值表示所观测到的空间模式是由某一随机过程创建而成的概率。比如我说，你计算出来的p值是1，那就表示你用于计算的这份数据，100%是随机生成的了（当然，不可能是1的，0.5以上就也不得了）。如果是0.1，就表示只有10%的可能性是随机生成的结果。**p值是越小越好**

**z得分（Zscores）表示标准差的倍数（standarddeviations）。**

**空间关系的概念化**

七种常见的空间关系：反距离、距离范围、无差别区域、面邻接、K最近邻要素、三角测量、空间时间窗。

**高/低值的聚类**

衡量高低值聚类：General G 指数 **对平均数非常敏感**

Z得分为正——观察General G指数大于期望GeneralG指数——数据在高值区域聚类。

Z得分为负——期望General G指数大于观察GeneralG指数——数据在低值区域聚类。

**地理加权回归**

**R，也就是判定系数** 这个系数在0-1之间，越接近1，表示这回归模型的效果越好，比如判定系数为1的话，就表示模型中的自变量能够100%的解释因变量的变化

**空间非平稳性（spatial nonstationarity)：**因为地理位置的变化，而引起的变量间关系或结构的变化，3个原因：

1、随机抽样的误差引起的。

2、是由于各地区不同的自然环境、人文环境等差异所引起的变量间的关系随着地理位置的变化而变化。

3、用于分析的模型与实际不符，

**空间非平稳性：**局部回归分析，移动窗口回归

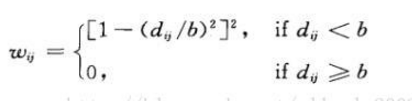
**常见的空间权重函数：**

1、高斯函数：

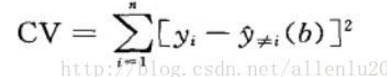


其中，b是带宽（窗口大小），dij是样本点i和j的距离（至于是哪种距离，就看选择了（欧式、曼哈顿、闵可夫斯基、球面、余弦等））。

2、双重平方函数（bi-square)



这两种距离函数都非常倚赖带宽b，那么这个带宽和确定呢？国际上最普遍的方法就是用Cleveland(1979)和Bowman(1984)提出的交叉确认（cross-validation,CV)方法来确定：



当GWR模型的AIC最小的时候，就是**最佳带宽。**

**AIC，**是衡量统计模型拟合优良性的一种标准，是由日本统计学家赤池弘次创立和发展的。赤池信息量准则建立在熵的概念基础上，可以权衡所估计模型的复杂度和此模型拟合数据的优良性。