**Homework\_09**

**学号：SA23008214**

**姓名：孙丹**

**Topic：学习莫兰指数/盖里指数以及过滤器的学习**

01-学习了解莫兰指数Moran's I与盖里指数Geary's C的概念，应用，区别及数据分析；

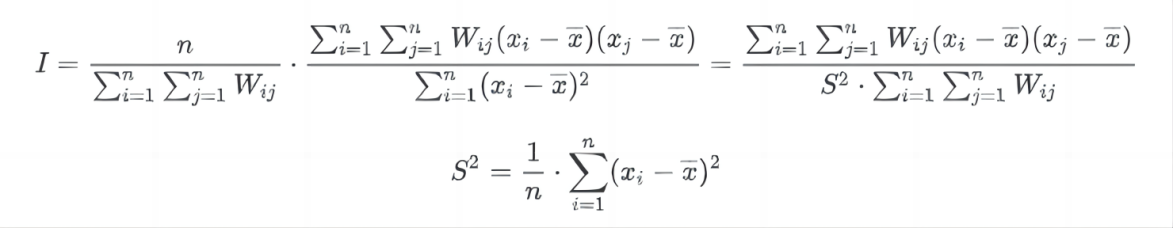
①概念：总体来看，Moran's I和Geary's C都是空间自相关的统计方法（或者说统计量）。看到比较通俗解释这两个统计方法的意思是：两个指数就是用来研究某事物“近朱者赤、近墨者黑”的程度，即分布模式是“物以类聚”型的、“随心所欲”型的还是“异性相吸”型的；

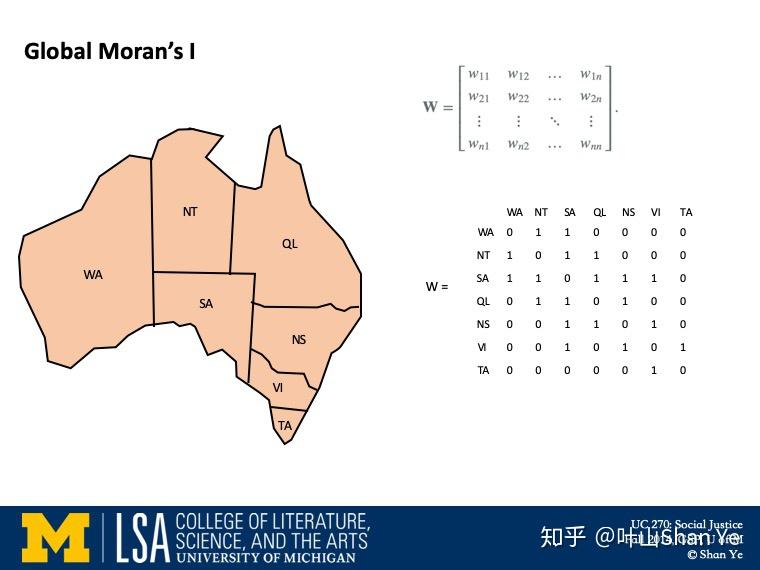
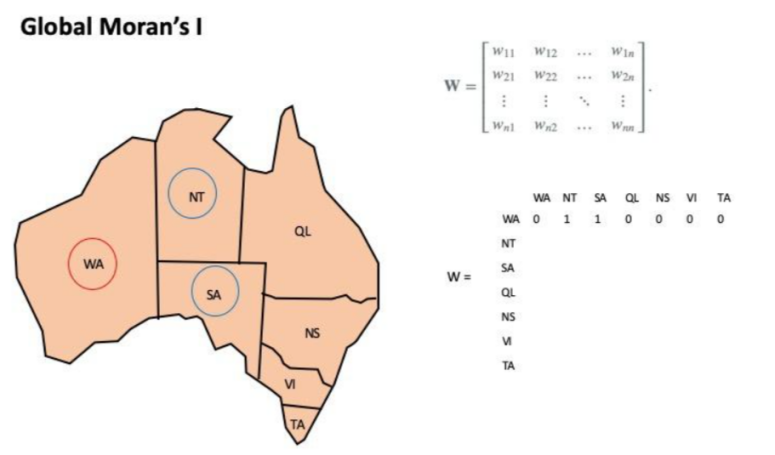
②莫兰指数与盖里指数的不同点

使用的计算原理不同：莫兰指数主要是进行相邻数据点和平均值之间关系的比较，而盖里指数则更加关注相邻数据点本身数据之间的比较。

比如在莫兰指数进行计算时，会找出一个数据群的平均值，让A与平均值相比，B也与平均值相比，分别计算AB与平均值的差异，然后根据分别获得的差异值进行数值相乘，根据乘积的正负就可以发现AB是相同的，都比平均值高还是都比平均值低，或者相对于平均值的差异不同，从而判断出AB的相似程度，再结合一些实际指标判断AB之间呈现的某种关系是不是与某种因素相关。（自我理解）

计算公式：



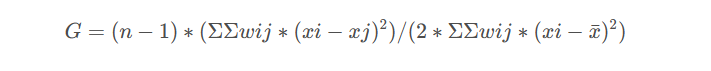


上面图片展现的这个例子就很好的说明了Wij的含义与获得，像WA与NT相邻，返回值为1，WA与QL不相邻，返回值为0，如此遍历所有的名称之间的数值，得到空间矩阵。

参考连接：https://www.zhihu.com/question/338420218

而在盖里指数中，是直接用A与B来比较，不再借助中间值，通过差异值平方，可以让所有的差异值都为正值，这样的话通过判断获得的值大小，再进行一些统计处理，可以获得AB之间的关系，再去探讨AB之间呈现的关系可能与哪种因素相关。

公式：

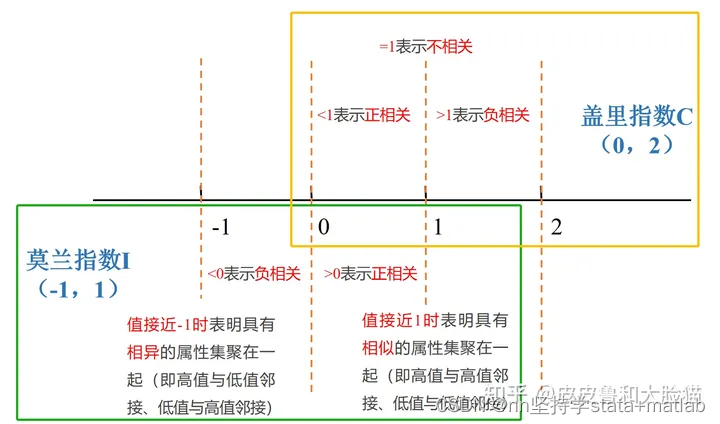


其中，G 是Geary’s C指数，n 是样本数量，wij 是地理邻接权重，xi 是第i个样本的值，xj 是第j个样本的值，x̄ 是所有样本的平均值。

③热点与异常值：对于题目中提到的热点与异常值，我的理解是热点即为“合群”值，异常值对应为“离群”值，即同类聚集在一起形成热点，在一个相似的数据集中，发现了一个差异很大的值即为异常值；

具体而言，对于莫兰指数，在四个象限中，高高聚集，低低聚集都是正常的，高低聚集，低高聚集则为异常由于其取值范围只能在-1到1之间，那么一般计算结果>0表示正相关,值接近1时表明具有相似的属性集聚在一起(即高值与高值邻接、低值与低值邻接)，为正常值，若<0表示负相关,值接近-1时表明具有相异的属性集聚在一起(即高值与低值邻接、低值与高值邻接)，若接近于0,则表示属性是随机分布的,或者不存在空间相关。

而对于盖里指数而言，其一般的取值范围在0到2间，同上若值 >1表示负相关，值=1表示不相关，值 <1表示正相关，即值越高越可能为异常值，值越低越正常。下图比较清晰的展示了两者之间的关系图。



02学习了解在建立空间权重矩阵邻接矩阵和距离矩阵时所应用的“过滤器”

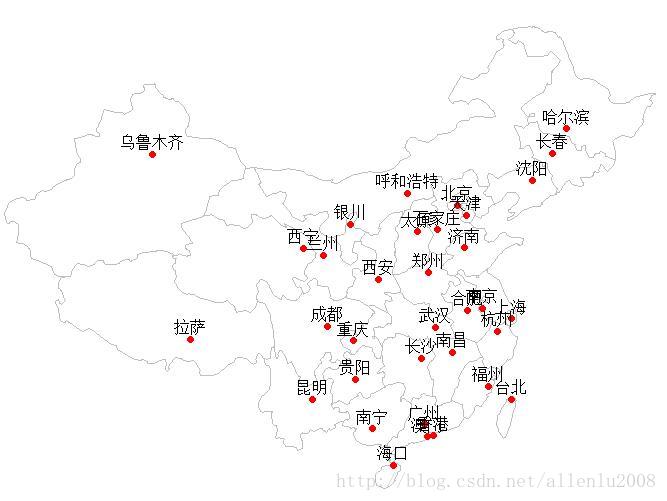
①概念：空间权重过滤器在地理数据分析、社交网络分析等领域中扮演着重要的角色。其中，k近邻过滤器和邻域过滤器是两种最为常见的空间权重过滤器。k近邻过滤器是一种基于每个单位的最近邻居来确定权重的过滤器。在构建空间权重矩阵时，每个单位都会被考虑其k个最近邻居。这些最近邻居的选择通常基于某种距离度量，如欧几里得距离、曼哈顿距离等。在矩阵中，每行对应一个单位，每列表示一个最近邻居，而元素值则反映了它们之间的连接强度。例如，两个单位之间的连接强度可以由它们之间的距离倒数表示，距离越近，连接强度越大。

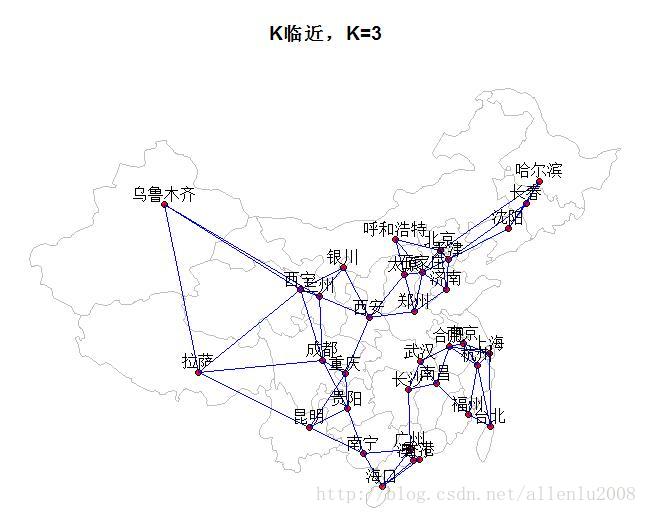
②对比：

k近邻过滤器的优点在于，它考虑了每个单位的局部环境，因此在某些情况下，它比其他类型的空间权重过滤器更为有效。例如，在分析人口迁移数据时，如果一个地区的人口迁移受到周边地区的影响较大，那么使用k近邻过滤器可能会得到更为准确的结果。与k近邻过滤器不同，邻域过滤器考虑的是每个单位的邻域范围内的其他单位。邻域可以通过距离阈值或者空间邻域（例如，多边形边界）来定义。在构建空间权重矩阵时，每行对应一个单位，每列表示邻域内的其他单位，而元素值则反映了它们之间的连接强度。邻域过滤器的优点在于，它考虑了每个单位的邻域范围内的所有单位，而不仅仅是k个最近邻居。因此，在某些情况下，邻域过滤器可能会得到更为全面的结果。例如，在分析城市之间的交通流量数据时，如果一个城市的交通流量受到周边多个城市的影响，那么使用邻域过滤器可能会得到更为准确的结果。

③具体内容：

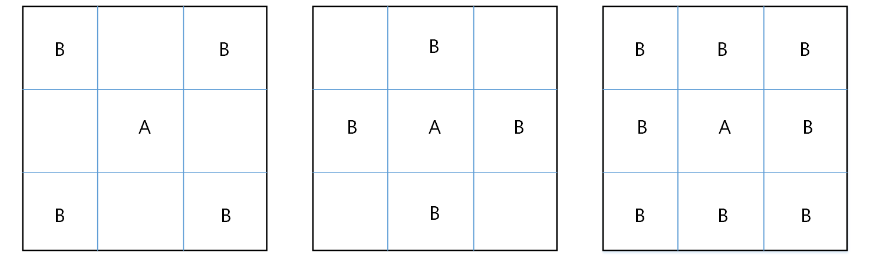
对于K邻近过滤器的理解，在网络上找到了一个较好的解释，下面这张图展现的就是中国主要城市的分布图，通过代码去定义K=3的临近关系，将获得的knn对象转换成nb对象，之后将nb对象转换为空间权重矩阵，可以获得下图这样的临近关系：





我们可以很清晰的看到，当我指定k=3时，会首先选择三个距离最近的“邻居”，即每个点都会至少是三条线段的交会点。

而对于邻域过滤器而言，通过了解发现其实一共存在着三种“邻接性”空间依赖关系：第一种是Bishop邻接，举例来看，它是指在一个含有9个，小矩形的大矩形中，相互共顶点的矩形关系，即处于对角线位置；第二种是Rock邻接，即在大矩形中共边的矩形之间的关系，处于竖直或水平直线上的矩形；第三种是Queen邻接，即在大矩形中AB型小矩形既共边又共顶点，类似于是Bishop邻接和Rock邻接的结合；但在实际研究过程中，由于地图形状多为不规则形，所以Bishop邻接，即公共顶点的空间相邻关系在现实情况中极其罕见。所以在大多数的时空统计分析中，选择以Queen连接设置的空间权重矩阵较为常见。下图分别显示的是Bishop邻接、Rock邻接、Queen邻接。



03创建一个基于Queen邻接的空间权重矩阵的R代码：

#加载软件包

library(spdep)

library(sp)

# 读取数据

data <- read.csv("F:/1/shuju/coordinate\_geo.csv")

# 提取经度和纬度列

coordinates <- data[, c("mapX", "mapY")]

#设置选择的邻居数

k <- 5

knn <- knn2nb(knearneigh(data, k = k))

# 创建空间权重矩阵

listw <- nb2listw(knn, style = "W")

# 检查权重矩阵的摘要

summary(listw)

# 可视化权重矩阵

plot(data, main = "邻近权重")

plot(knn, coordinates(data), add = TRUE)

#可视化结果

