

第七章 地理系统的聚类分析和判别分析



河北师范大学：胡引翠

主要内容



- ◆ § 1地理系统的聚类分析
- ◆ § 2地理系统的判别分析

§ 1地理系统的聚类分析



1.1地理系统分类概述

1.1.1地理系统分类的意义

地理分区和分类是一类重要的问题。

地理学的分类从靠经验和定性的知识进行分类转向应用数学和电子计算机的定量分析。这种定量分析方法称为“数值分类法”或“数量分类法”，亦称“聚类分析”（Cluster Analysis）”也被称为点群分析，或群分析，是研究多要素事物分类问题的数量方法。

聚类分析的职能是建立一种分类方法，它是将一批样品或变量，按照它们在性质上的亲疏程度进行分类。这种性质上的亲疏程度体现在特征属性值的大小上。

§ 1地理系统的聚类分析



1.1.2地理系统分类的原理

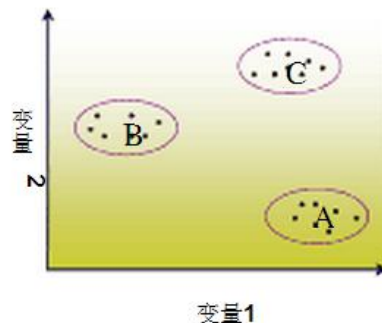
基本原理:

同一类中的个体有较大的相似性，不同类中的个体差异很大。

根据样本自身的属性，特征的相似性，用数学方法按照某种相似性或差异性指标，定量确定样本之间的亲疏关系，并按这种亲疏关系对样本进行聚类，得到一个分类系统。

基本特征:

事先无需知道分类对象的分类结构，而只需要知道一批批地理数据结构，然后选好分类统计量，并按一定的方法步骤进行计算，最后便能自然地客观地得出一张完整的分类系统图。



§ 1地理系统的聚类分析



物以类聚：

把它们分成2组，你怎么分？



§ 1地理系统的聚类分析



我国的县市进行分类:

-按照地理位置分类, 如北方、南方.....

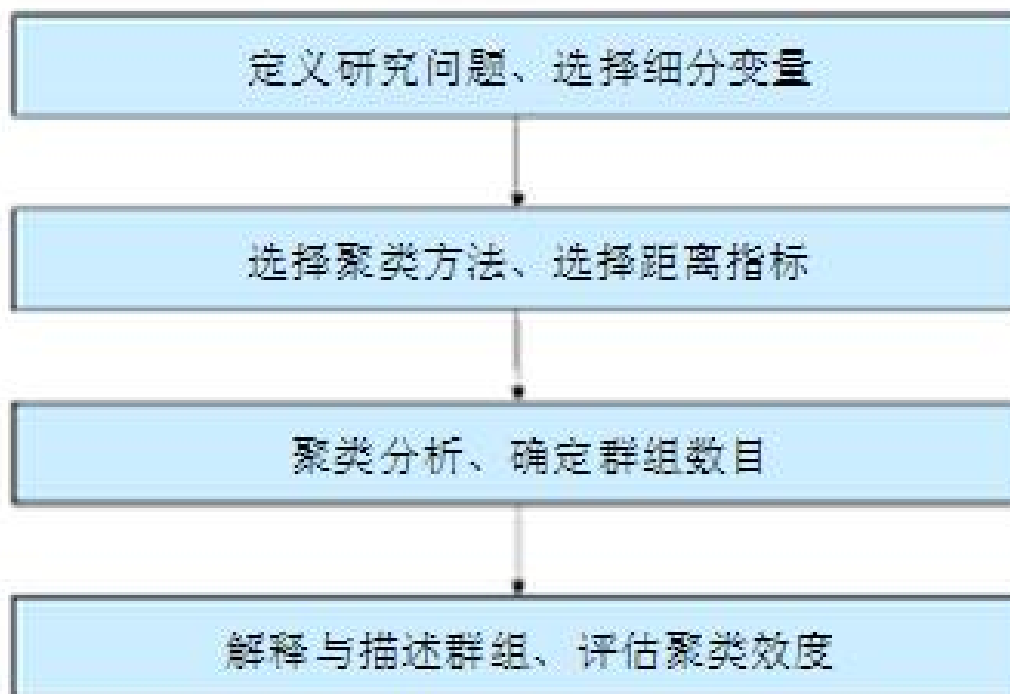
-按照自然条件分类, 如降水、土地、日照、湿度.....

-按照经济发展水平, 如收入、教育水准、医疗条件、基础设施.....

§ 1地理系统的聚类分析



1.1.3地理系统分类的步骤



§ 1地理系统的聚类分析



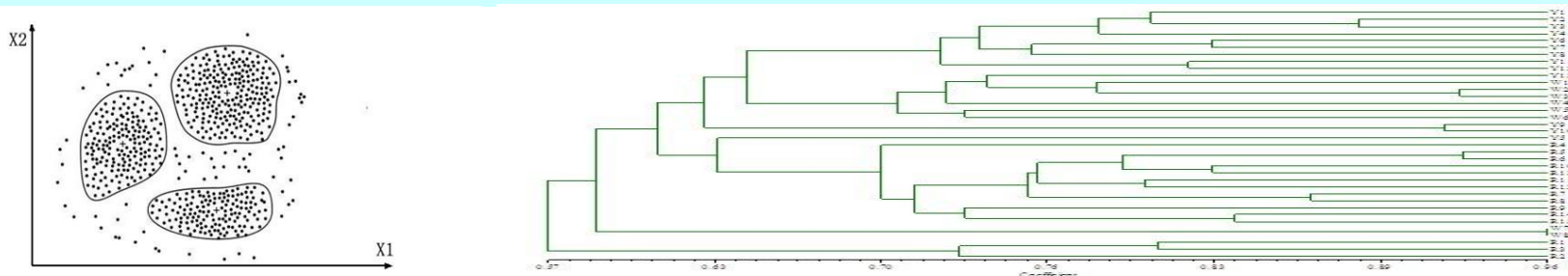
1.1.3地理系统分类的步骤

基本步骤:

首先：我们要根据一批批地理数据或指标找出能度量这些数据或指标之间相似程度的统计量。

然后，以统计量作为划分类型的依据，把一些相似程度较大的站点或样品（样本）先聚为一新类，新生成的类和其他站点求算距离，找出距离最近的类进行合并，再将合并的类型和其余站点（或样本）求算距离矩阵.....直到所有的站点（样品）都聚合完。

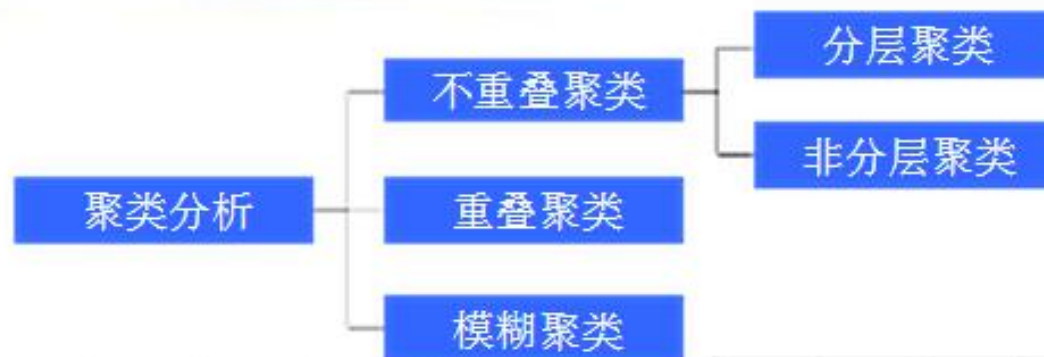
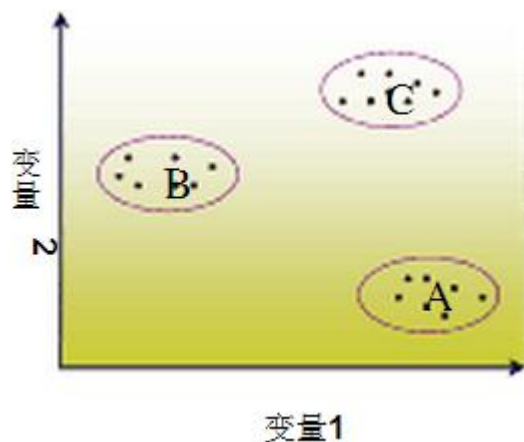
最后，便可根据各类之间的亲疏关系，逐步画成一张完整的分类系统图（谱系图）。



§ 1地理系统的聚类分析



1.1.3地理系统分类的方法



不重叠	市场1	市场2	市场3
顾客1	1		
顾客2			1
顾客3		1	

重叠	市场1	市场2	市场3
顾客1	1	1	
顾客2			1
顾客3		1	1

模糊	市场1	市场2	市场3
顾客1	0.1	0.1	0.8
顾客2	0.6	0.4	0.0
顾客3	0.2	0.3	0.5



1.1.4 地理系统分类的分类

“聚类分析” 方法有系统聚类法、动态聚类法和模糊聚类法等。

根据分类对象的不同，分为样品聚类（**Q**型聚类）和变量聚类（**R**型聚类）。

样品聚类（**Q**型聚类）：在**SPSS**中称对事件（**cases**）进行聚类，或是说对观测量进行聚类，根据被观测的对象的各种特征值进行分类。

变量聚类（**R**型聚类）：反映同一事物特点的变量有很多，变量之间往往具有一定的相关关系，需要找出彼此独立的有代表性的变量，就需要先进行变量聚类。（批量生产）



■ 1.2 聚类分析的数据处理

- ◆ Q型聚类分析就是根据地理区域（站点）的各属性变量的观测值予以分类的，它涉及到通过各种途径和手段所得到的有意义的地理数据。
- ◆ 为什么要对聚类分析的原始数据进行标准化处理？

由于要素(或指标)的量纲、数量级和数量变化幅度的差异，如用原始数据进行聚类分析，就会将不同性质、不同量纲、不同数量变化幅度的数值都统计在一起，这样就有可能突出某些数量级特别大的变量对分类的作用，而压低甚至排除了某些数量级很小的变量对分类的作用。
- ◆ 消除量纲的不同和使每一变量都统一在某种共同的、相对均匀化的数值范围内。为此，常对数据进行处理。
- ◆ 地理数据的处理有对数变换和标准化处理两个步骤。



区 代 号	人均 耕地 x_1 $/(\text{hm}^2 \cdot \text{人}^{-1})$	劳均 耕地 x_2 $/(\text{hm}^2 \cdot \text{个}^{-1})$	水田 比重 x_3 $\%$	复种 指数 x_4 $\%$	粮食 单产 x_5 $/(\text{kg} \cdot \text{hm}^{-2})$	人均粮食 x_6 $/(\text{kg} \cdot \text{人}^{-1})$	稻 谷 占 粮 食 比 重 $x_7\%$
G1	0.294	1.093	5.63	113.6	4 510.5	1 036.4	12.2
G2	0.315	0.971	0.39	95.1	2 773.5	683.7	0.85
G3	0.123	0.316	5.28	148.5	6 934.5	611.1	6.49
G4	0.179	0.527	0.39	111	4 458	632.6	0.92
G5	0.081	0.212	72.04	217.8	12 249	791.1	80.38
G6	0.082	0.211	43.78	179.6	8 973	636.5	48.17
G7	0.075	0.181	65.15	194.7	10 689	634.3	80.17
G8	0.293	0.666	5.35	94.9	3 679.5	771.7	7.8
G9	0.167	0.414	2.9	94.8	4 231.5	574.6	1.17



指标

地点（地区、
样品、样本）

聚 类 对 象	要 素					
	x_1	x_2	\cdots	x_j	\cdots	x_n
1	x_{11}	x_{12}	\cdots	x_{1j}	\cdots	x_{1n}
2	x_{21}	x_{22}	\cdots	x_{2j}	\cdots	x_{2n}
\vdots	\vdots	\vdots		\vdots		\vdots
i	x_{i1}	x_{i2}	\cdots	x_{ij}	\cdots	x_{in}
\vdots	\vdots	\vdots		\vdots		\vdots
m	x_{m1}	x_{m2}	\cdots	x_{mj}	\cdots	x_{mn}





■ 1.2 聚类分析的数据处理

(一) 数据的对数变换

各指标经对数变换后，数据变幅减小且变均匀。

$$x'_{ij} = \ln x_{ij}$$

(二) 数据的标准化

- 标准差标准化
- 级差标准化



◆ (二) 数据的标准化

1、标准差标准化

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

其中：

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

变换后，每一列均值为0，方差为1，且与指标的量纲无关。

$$S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$$

- ◆ 标准差标准化就是把变换后的数据 (x_{ij}) 减去其均值 (\bar{x}_j)，再除以其标准差 (S_j)



◆ （二）数据的标准化

II、极差标准化

所谓极差标准化，就是系列中的任一变量(x_{ij})与其第j列中的最小值 $x_j(\min)$ 之差和第j列的极差的比值。这样变换后的变量，也叫“极差正规化”。

变换后，数据的变化范围都在0到1之间，也消除了量纲的影响。

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$



III、线性比例变化法

对于正向指标，取

$$x'_{ij} = \frac{x_{ij}}{\max_{1 \leq i \leq m} x_{ij}}, (1 \leq i \leq m, 1 \leq j \leq n)$$

对于逆向指标，取

$$x'_{ij} = \frac{\min_{1 \leq i \leq m} x_{ij}}{x_{ij}}, (1 \leq i \leq m, 1 \leq j \leq n)$$

经过线性变换后，指标值均在0~1之间，正、逆向指标均化为正向指标，最优值为1，最劣值为0。



IV、平均值法

一组数 $x_1, x_2, x_3, \dots, x_n$, 其平均值为 \bar{x} ,
则各数的平均值法得到的标准化值为 $x_1/\bar{x}, x_2/\bar{x}, \dots$



V、定性指标量化处理方法

可以最优值为**10**分，最劣值为**0**分,其余相应给分如下：

指标	很低	低	一般	高	很高
正向 指标	1	3	5	7	9
逆向 指标	9	7	5	3	1



举例：表6.2给出了某地区9个农业区的7项指标。

表 6.3 极差标准化处理后的数据

区 代 号	人均 耕地 X_1 / (hm ² ·人 ⁻¹)	劳 耕		x_1	x_2	x_3	x_4	x_5	x_6	x_7
G ₁	0.294	1.		0.91	1.00	0.07	0.15	0.18	1.00	0.14
G ₂	0.315	0.		1.00	0.87	0.00	0.00	0.00	0.24	0.00
G ₃	0.123	0.		0.20	0.15	0.07	0.44	0.44	0.08	0.07
G ₄	0.179	0.		0.44	0.38	0.00	0.13	0.18	0.13	0.00
G ₅	0.081	0.		0.03	0.03	1.00	1.00	1.00	0.45	1.00
G ₆	0.082	0.		0.03	0.03	0.61	0.69	0.65	0.13	0.59
G ₇	0.075	0.		0.00	0.00	0.90	0.81	0.84	0.13	1.00
G ₈	0.293	0.		0.91	0.53	0.07	0.00	0.10	0.43	0.09
G ₉	0.167	0.		0.38	0.26	0.04	0.00	0.15	0.00	0.00



例2: 某航空公司在国际市场上购买飞机，按照6个决策指标对不同型号的飞机进行综合评价。有4种型号飞机可供选择。具体指标如下表：写出决策矩阵，进行标准化处理。

	最大速度 (马赫)	最大范围 (公里)	最大负载 (千克)	费用 (10^6 美元)	可靠性	灵敏度
A	2.0	1500	20000	5.5	一般	很高
B	2.5	2700	18000	6.5	低	一般
C	1.8	2000	21000	4.5	高	高
D	2.2	1800	20000	5.0	一般	一般

解：前三个指标是正向指标，第四个指标是逆向指标，第五、六指标是定性指标。先将第五、六指标量化，写出决策矩阵，再分别用前三种方法归一化。



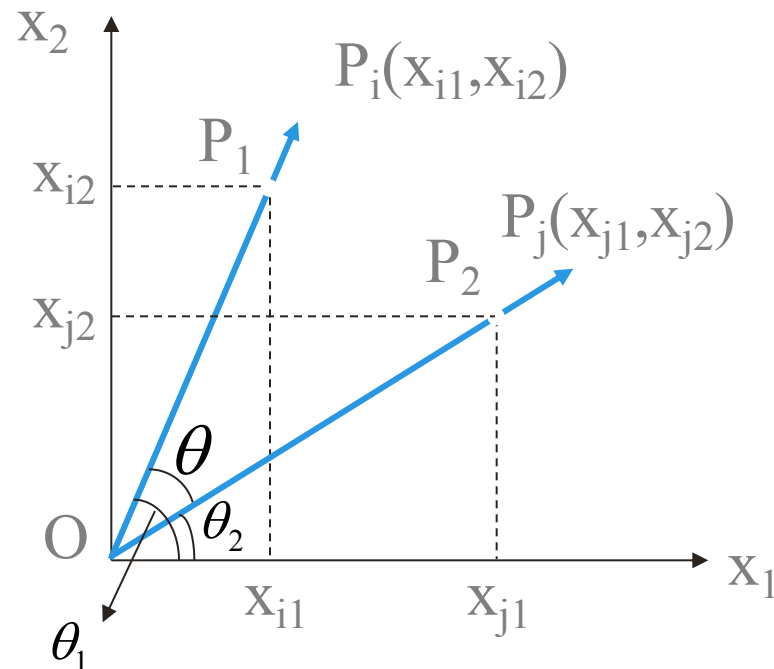
■ 1. 3聚类分析的统计量

(一) 相似系数

1. 夹角余弦 ($\cos\theta$)

- 二维空间:

$$\begin{aligned}\cos \theta &= \cos(\theta_1 - \theta_2) \\ &= \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 \\ &= \frac{x_{i1}}{P_1} \cdot \frac{x_{j1}}{P_2} + \frac{x_{i2}}{P_1} \cdot \frac{x_{j2}}{P_2}\end{aligned}$$



$$P_1 = \sqrt{x_{i1}^2 + x_{i2}^2} \quad P_2 = \sqrt{x_{j1}^2 + x_{j2}^2}$$



二维：

$$\cos \theta = \frac{x_{i1}x_{j1} + x_{i2}x_{j2}}{\sqrt{(x_{i1}^2 + x_{i2}^2)(x_{j1}^2 + x_{j2}^2)}}$$

三维：

$$\cos \theta = \frac{x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}}{\sqrt{(x_{i1}^2 + x_{i2}^2 + x_{i3}^2)(x_{j1}^2 + x_{j2}^2 + x_{j3}^2)}}$$

n 维：

$$\cos \theta_{ij} = \frac{\sum_{k=1}^m x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2 \cdot \sum_{k=1}^m x_{jk}^2}}$$

式中：i和j代表两个地点（样品）， x_{ik}, x_{jk} 分别代表第i个地点和第j个地点的第k个地理指标值。



◆ 相似系数矩阵

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

◆ 性质：

对角线元素为1

对称方阵

取值范围：-1~1

绝对值越大，相似程度越强。



■ 1. 3聚类分析的统计量

(一) 相似系数

2. 相关系数 (r)

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \cdot \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}}$$

在数据标准化后，相关系数公式简化为：

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m x_{ik} \cdot x_{jk}$$

式中：i和j代表两个地点（样品）， x_{ik} , x_{jk} 分别代表第i个地点和第j个地点的第k个地理指标值。



■ 1.3 聚类分析的统计量

◆ 相关阵

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

◆ 性质：

对角线元素为1

对称方阵

取值范围：0~1

取值越大，相关程度越强。



■ 1.3 聚类分析的统计量

(二) 距离系数

- ◆ 假设把我们研究的对象（地点或样品）视为 m 维空间的点，所谓距离，就是用各种方法计算出各点间的相互距离（ d_{ij} ）
- ◆ 并用它来刻画各点间的相似性或亲疏程度。
- ◆ 距离（距离系数）有多种，主要介绍两种距离，绝对距离值距离 $d_{ij}(1)$ 和欧氏距离 $d_{ij}(2)$ 。



■ 1.3 聚类分析的统计量

(二) 距离系数

1. 绝对距离 ($d_{ij}(1)$)

$$d_{ij}(1) = \sum_{k=1}^m |x_{ik} - x_{jk}|$$

式中： i 和 j 代表两个地点（样品）， x_{ik}, x_{jk} 分别代表第 i 个地点和第 j 个地点的第 k 个地理指标值。



■绝对距离 d_{ij} (1)

表 6.3 极差标准化处理后的数据

	x_1	x_2	x_3	x_4	x_5	x_6	x_7			
G_1	0.91	1.00	0.87	0.45	0.40	1.00	0.44	$D = (d_{ij})_{9 \times 9} =$		
G_2	1.00									
G_3	0.20									
G_4	0.44									
G_5	0.00									
G_6	0.00									
G_7	0.00									
G_8	0.91									
G_9	0.30									

0										
1.52	0									
3.10	2.70	0								
2.19	1.47	1.23	0							
5.86	6.02	3.64	4.77	0						
4.72	4.46	1.86	2.99	1.78	0					
5.79	5.53	2.93	4.06	0.83	1.07	0				
1.32	0.88	2.24	1.29	5.14	3.96	5.03	0			
2.62	1.66	1.20	0.51	4.84	3.06	3.32	1.40	0		



■ 1. 3聚类分析的统计量

(二) 距离系数

2. 欧氏距离 ($d_{ij}(2)$)

- 空间中两点的距离:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

称为三维空间中的欧几里德距离，简称欧氏距离。

距离小表示两点接近，反之则远离。



n维空间中： $i(x_{i1}, x_{i2}, \dots, x_{im}), j(x_{j1}, x_{j2}, \dots, x_{jm})$

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

式中： i 和 j 代表两个地点（样品）， x_{ik}, x_{jk} 分别代表第 i 个地点和第 j 个地点的第 k 个地理指标值。

欧氏距离系数： $d_{ij}(2) = \sqrt{\frac{1}{m} \sum_{k=1}^m (x_{ik} - x_{jk})^2}$



■ 欧氏距离 $d_{ij}(2)$

表 6.3 极差标准化处理后的数据

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
G_1	0.91	1.00	0.07	0.00	0.00	0.24	0.00
G_2	1.00	0.87	0.00	0.00	0.00	0.24	0.00
G_3	0.20	0.15	0.07	0.44	0.44	0.08	0.07
G_4	0.44	0.38	0.00	0.13	0.18	0.13	0.00
G_5	0.03	0.03	1.00	1.00	1.00	0.45	1.00
G_6	0.03	0.03	0.61	0.69	0.65	0.13	0.59
G_7	0.00	0.00	0.90	0.81	0.84	0.13	1.00
G_8	0.91	0.53	0.07	0.00	0.10	0.43	0.09
G_9	0.38	0.26	0.04	0.00	0.15	0.00	0.00

$$d_{ij}^2(2) = (x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + (x_{13} - x_{23})^2 + (x_{14} - x_{24})^2 + (x_{15} - x_{25})^2 + (x_{16} - x_{26})^2 + (x_{17} - x_{27})^2$$



◆ 距离系数矩阵

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

◆ 性质：

对角线元素为0

对称方阵

取值越小，相似程度越强。



■ 1.4 地理系统的聚类方法

聚类分析的分类：

- Q型聚类分析
- R型聚类分析

基本做法：

- 将 n 个地点各自看成一类，定义类与类间的距离；
- 选择距离最小的一对合成一新类；
- 计算新类与其它类间的距离，重复第一步。



■ 1.4 地理系统的聚类方法

- ◆ 聚类方法有多种，以系统聚类法应用最广。

原理

- 先把各个分类对象单独视为一类(原始类)；求出距离矩阵。
- 定义样品间、类与类之间的距离，进而选择距离最小的样品和样品（类和类）形成新的类型，然后计算新类和其他类型之间的距离，
- 再将距离最近的两类合并，这样每合并一次都缩小一类，直到所有的样品都为一类为止。
- 这样就可以根据归并的先后顺序和距离作出聚类谱系图。

关键：

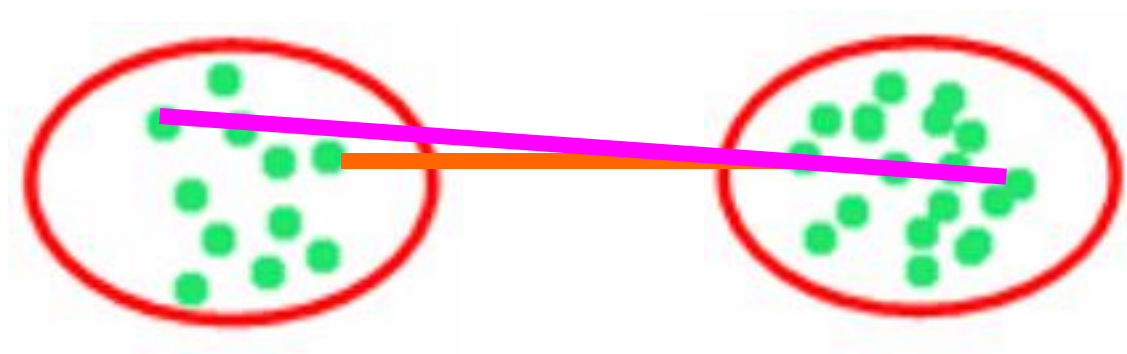
在于合并成新类之后，新类和新类、新类和尚未归类的原有各类之间的距离的计算。



■ 1.4 地理系统的聚类方法

定义类与类之间距离的方法：

- 最短距离法
- 最长距离法
- 类平均法
- 重心法
- 离差平方和法





■ 1.4 地理系统的聚类方法

(一) 最短距离法

- 应用广泛；
- 定义两类之间的距离：两类间所有样本中最近的两个样本距离；
- d_{ij} 表示地点*i*与*j*的距离；
- G_1, G_2, \dots 表示类别。

$$d_{rk} = \min \{d_{pk}, d_{qk}\}$$

↓ ↓ ↓

新类的记号 旧类的记号



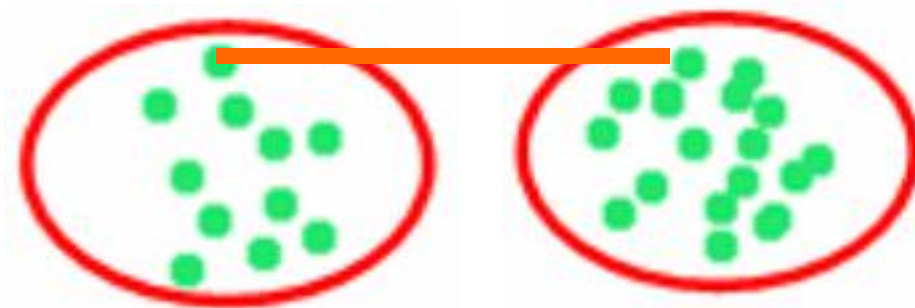
系统聚类的距离计算——最短距离法

- ◆ 定义两类之间的距离以两类间所有样本中最近的两个样本间的距离来表示。

- ◆ 表示地点（或样品） i 与地点（或样品） j 的距离，用 d_{ij} 表示， G_1, G_2, \dots 表示类别，而 G_p 和 G_q 两类间的距离用 D_{pq} 表示，则

$$D_{pq} = \min d_{ij}$$

$$i \in G_p, j \in G_q$$





- ◆ 例：有一组5个区域，2个指标的地理资料矩阵如表所示。试对这5个区域用最短距离法进行聚类分析。

区域	指标 x_1	指标 x_2
1	1	3
2	2	4
3	4	5
4	6	10
5	8	12
均值	4.2	6.8
标准差	2.86	3.96



- 1.数据的标准差标准化处理。
- 2.计算初始距离系数矩阵 $D(0)$ 。

$$D(0) = \begin{matrix} & G_1 & G_2 & G_3 & G_4 & G_5 \\ \begin{matrix} G_1 \\ G_2 \\ G_3 \\ G_4 \\ G_5 \end{matrix} & \begin{bmatrix} 0 & & & & \\ 0.043 & 0 & & & \\ 1.167 & 0.723 & 0 & & \\ 2.484 & 2.060 & 1.442 & 0 & \\ 3.337 & 2.910 & 2.252 & 0.862 & 0 \end{bmatrix} \end{matrix}$$

区域	指标 x_1	指标 x_2
1	-1.12	-0.96
2	-0.77	-0.71
3	-0.07	-0.45
4	0.63	0.81
5	1.33	1.31



1. 在 $D(0)$ 中，选出距离系数最小的，即 $d_{12}=0.043$ ，将 G_1 和 G_2 合并成新类 G_6 ，记为 $G_6=\{G_1, G_2\}$ ，并记入联结表中。

再利用最短距离公式计算 G_6 与其它各类之间的距离。

$$d_{63} = \min\{d_{13}, d_{23}\} = \min\{1.167, 0.723\} = 0.723$$

$$d_{64} = \min\{d_{14}, d_{24}\} = \min\{2.484, 2.060\} = 2.060$$

$$d_{65} = \min\{d_{15}, d_{25}\} = \min\{3.337, 2.910\} = 2.910$$

从而形成距离系数矩阵 $D(1)$ 。



$$D(1) = \begin{matrix} & G_6 & G_3 & G_4 & G_5 \\ \begin{matrix} G_6 \\ G_3 \\ G_4 \\ G_5 \end{matrix} & \begin{bmatrix} 0 \\ 0.723 & 0 \\ 2.060 & 1.442 & 0 \\ 2.910 & 2.252 & 0.862 & 0 \end{bmatrix} \end{matrix}$$

2.在D(1)中，选出距离系数最小的，即 $d_{63}=0.723$ ，将 G_6 和 G_3 合并成新类 G_7 ，记为 $G_7=\{G_6, G_3\}$ ，并记入[联结表](#)

计算 G_7 与其它各类之间的距离。



$$d_{74} = \min \{d_{64}, d_{34}\} = \min \{2.060, 1.442\} = 1.442$$

$$d_{75} = \min \{d_{65}, d_{35}\} = \min \{2.910, 2.252\} = 2.252$$

从而形成距离系数矩阵D(2)。

$$D(2) \begin{matrix} & \begin{matrix} G_7 & G_4 & G_5 \end{matrix} \\ \begin{matrix} G_7 \\ G_4 \\ G_5 \end{matrix} & \begin{matrix} 0 & & \\ 1.442 & 0 & \\ 2.252 & 0.862 & 0 \end{matrix} \end{matrix}$$



3.在D(2)中，选出距离系数最小的，即 $d_{45}=0.862$ ，将 G_4 和 G_5 合并成新类 G_8 ，记为 $G_8=\{G_4, G_5\}$ ，并记入[联结表](#)

计算 G_8 与其它各类之间的距离。

$$d_{87} = \min\{d_{47}, d_{57}\} = \min\{1.442, 2.252\} = 1.442$$

从而形成距离系数矩阵D(2)。

$$D(2) = \begin{matrix} & \begin{matrix} G_7 & G_8 \end{matrix} \\ \begin{matrix} G_7 \\ G_8 \end{matrix} & \begin{bmatrix} 0 & \\ 1.442 & 0 \end{bmatrix} \end{matrix}$$



4. $d_{78}=1.442$, 将 G_7 和 G_8 合并成新类 G_9 , 记为 $G_9=\{G_7, G_8\}$, 并记入 联结表

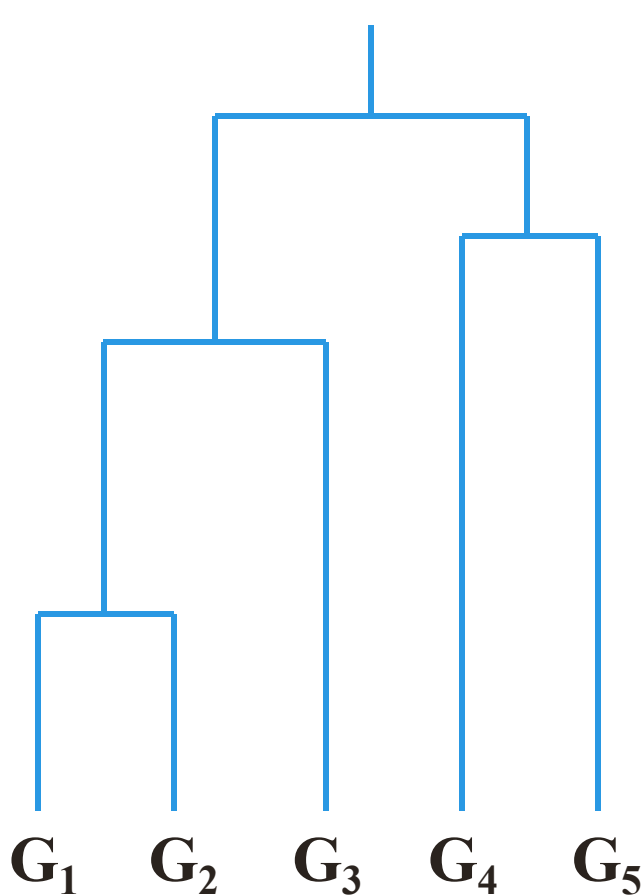
5. 作出分类谱系图。

1.442

0.862

0.723

0.043





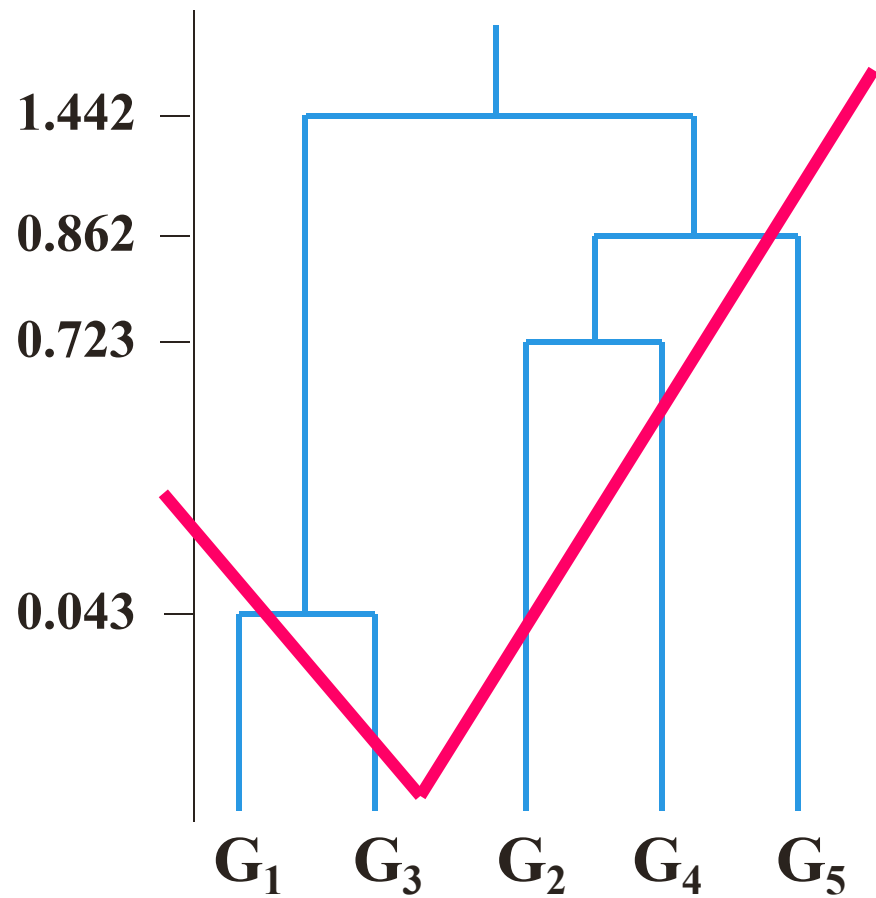
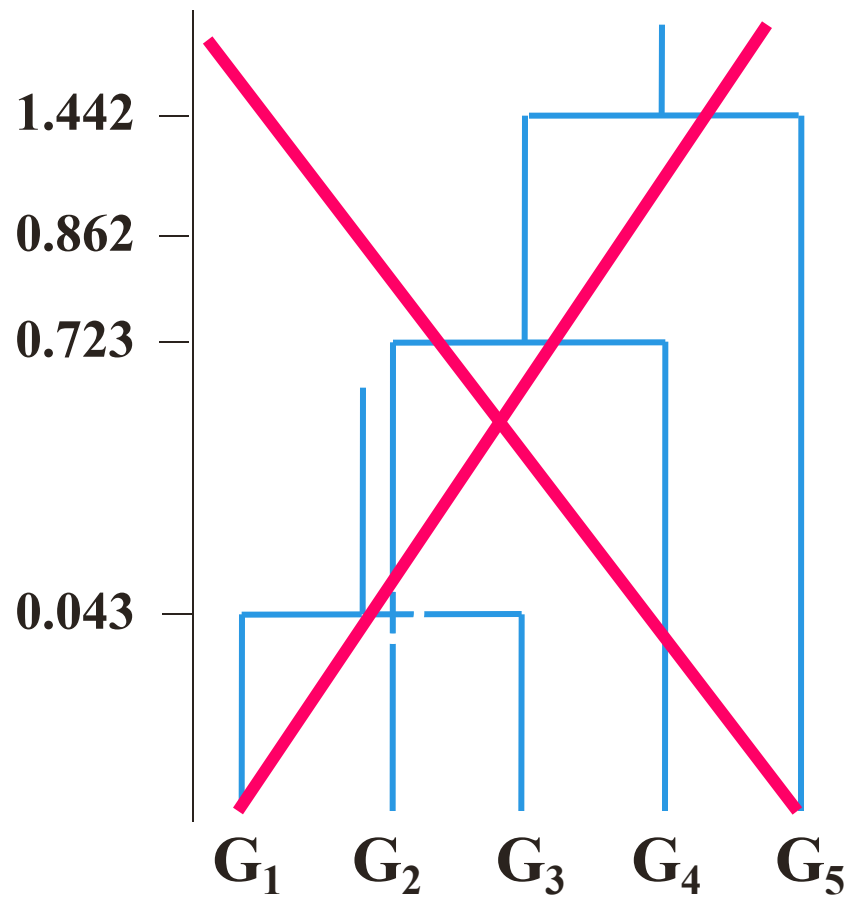
联 结 表

联结顺序	联结法		距离系数
	新类	类别	
1	G_6	$G_1、G_2$	0.043
2	G_7	$G_3、G_1、G_2$	0.723
3	G_8	$G_4、G_5$	0.862
4	G_9	$G_3、G_1、G_2、G_4、G_5$	1.442



注意：分类谱系图不能交叉。例如：

联结顺序	联结法		距离系数
	新类	类别	
1	G_6	$G_1、G_3$	0.043
2	G_7	$G_2、G_4$	0.723
3	G_8	$G_5、G_2、G_4$	0.862
4	G_9	$G_1、G_3、G_5、G_2、G_4$	1.442





(二) 最长距离法

- 定义新类与其它类距离的原理不同；
- 其余相同。

$$d_{rk} = \max \{d_{pk}, d_{qk}\}$$

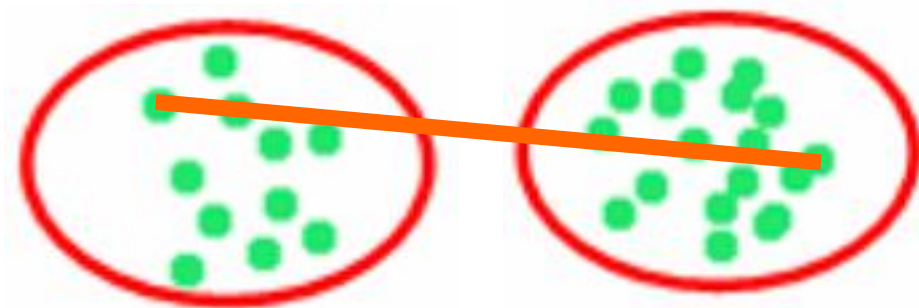


系统聚类的距离计算——最长距离法

- ◆ 定义两类之间的距离以两类间所有样本中最远的两个样本间的距离来表示。
- ◆ 表示地点（或样品） i （ $i \in G_p$ ）与地点（或样品） j （ $j \in G_q$ ）的距离，用 G_1, G_2, \dots 表示类别，而 G_p 和 G_q 两类间的距离用 D_{pq} 表示，则

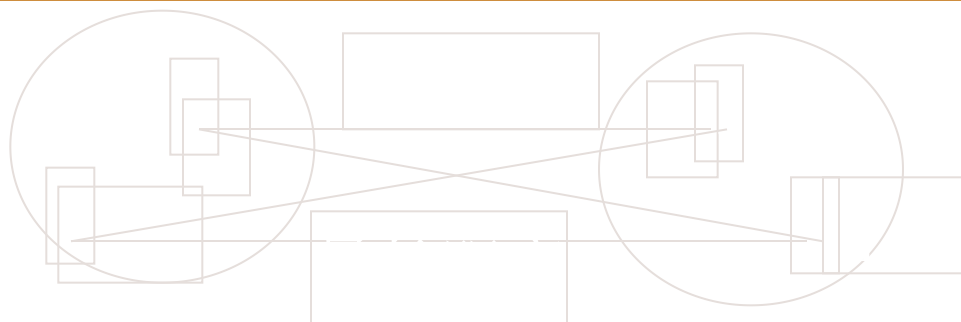
$$D_{pq} = \max d_{ij}$$

$$i \in G_p, j \in G_q$$





系统聚类法计算类之间距离的统一公式



从图中可以看出，最短距离聚类法具有空间压缩性，而最大距离聚类法具有空间扩张性。