1. **实验原理**

聚类分析是将样本个体或指标变量按其具有的特性进行分类的一种统计分析方法。采用谱系聚类方法，首先视各种样品自成一类，然后把最相似(距离最小或相似系数最大)的样品聚为小类，再将已聚合的小类按其相似性再聚合，随着相似性的减弱，最后将一切样品都聚合成一个大类，从而得到一个按相似性大小聚合起来的一个谱系图。其中对于相似程度，样品之间定义距离，变量之间定义相似系数。（本文采用欧式距离作为样品间距离）

1. **预备知识**
2. **聚类分析的方法**

聚类分析的方法有很多，在这里主要介绍常用的系统聚类法和逐步聚类法，系统聚类适用于小样本的样本聚类或指标聚类，逐步聚类适用于大样本的样本聚类。对于小样本的样本聚类，如果采用逐步聚类，聚类结果将与样本的顺序有关。

1. **聚类指标**

一般用系统聚类法来聚类指标，它的基本思想是先把k个指标看成一类，然后用主成分分析法将它分解成若干类，分类的原则是使得每一类的类内指标总变异尽可能多地被该类的类成分所解释。如果每一类所占比例满足事先给出的要求，则停止分类，反之将继续分类。

系统聚类分析的方法决定于类之间的距离的定义,由此有8种方法：

1) 最短距离法(single linkage)

2) 最长距离法(complete method)

3) 中间距离法(median method)

4) 重心法(centroid method)

5) 类平均法(averange linkage)

6) 可变类平均法（flexible-bete method)

7) 可变法及mcquitty 相似分析法(MCQ)

8) 离差平方和法(WARD)

实际操作中，只要在SAS语句’method=’等号后输入相应的方法的英文代码即可。

1. **聚类样本**

要聚类样本，可以用系统聚类也可以用逐步聚类法。

* 系统聚类法聚类样本

基本思想是先把n个聚类样本看成n类，然后按类间距离将相似程度最大的两个类合并为一类，再将所有的类（包括合并形成的新类）中相似程度最大的两个类合并为一类。重复此过程，直到所有的类间距离达到一定的要求为止，或直到所有的样本被合并为一类为止，然后根据类间距离的要求以及实际意义选择一个合适的分类。

* 逐步聚类法聚类样本

基本思想是先选择若干个初始凝聚点，可以是所有样本中的任意几个样本，也可以是随意确定的几个新样本观测值。然后把每个样本按距离大小归入到与该样本最近的凝聚点所代表的初始类中，再以这些初始类的“重心”（类内各样本观测值的平均值）作为新的凝聚点重新将样本归类。重复以上步骤，直到分成的类没有什么变化为止。

1. **聚类统计量（程序输出结果）**

 统计量:，取分类个数G使得足够大，但G本身比较小，而且不再大幅度增加；

 半偏相关统计量,是上一步的与该步值的差值,变偏相关越大，说明这两个类越不应该合并；所以如果G+1类合并为G类时很大，则分类应取G+1类；

 伪F统计量(PSF)， PSF值越大表示这些观测可显著地分为NCL个类，所以应取伪F统计量较大而分类数G较小的聚类水平；

* + 伪统计量(PSF2)该统计量用以评价此步骤合并类和的效果。如该值越大,说明不应该合并这两个类，应取合并前的聚类水平。

1. **实验举例**

**（1）绘制树状图—TREE过程**

TREE语句的格式：proc tree options；

By variables；

Id variable；

Copy variables；

Run；

其中“options”中可以包括以下内容

1. Data=sas data set：指定要画图的SAS数据集名称；
2. Horizontal：表示谱系图水平放置；如果省略，则为垂直放置
3. Vertical：表示谱系图垂直放置；
4. Graphics：要求画高分辨率的聚类谱系图。
5. Spaces= n：表示各ID变量值的间隔单位，这里n是正整数；
6. Nclusters= n；指定输出各类结果数据集中所分成的类的个数。
7. Out=sas data set：输出分类结果。

Tree过程并不是一个独立的过程，它是利用cluster和varclus过程建立的数据集显示树图，使聚类的结果更加形象。

1. **CLUSTER过程（小样本聚类）**

Cluster语句格式：

Proc cluster options;

Var variables;

Id variable;

Copy variable;

Run;

Optins中可填入以下内容；

1. Data=数据集名称；
2. Outtree=sas数据集：命名一个TREE过程所使用的输出数据集；
3. Method=方法名称，一般有一下几种方法（前面介绍的八种）；
4. Nosquare：阻止过程在MEHOD= AVERAGE、CENTROID、MEDIAN或WARD方法中将输入的距离平方；
5. Nonorm：表示类间距离不进行标准化；
6. Pseudo：要求打印伪F统计量及伪统计量的值。只有当输入数据是原始观测数据（但“METHOD”选项不能为SIN）或者“METHOD”选项为AVE、CEN或VAR时，才可选此项；



1. Rsquare（或rsq）：要求打印出统计量和偏相关统计量SPRSQ的值，当输入数据是原始观测数据且“METHOD”选项是AVE 或CEN时，才可选此项。



其中，类平均法和离差平方和法使用最广泛。

1. VAR语句

此语句指定用来聚类数值型变量。默认情况下使用没有列在其他语句中的数据值型变量。

1. ID 语句

ID语句中的变量用以区分聚类过程打印和OUTTREE=数据集中的观测，若缺省，则用OBN表示。

1. COPY 语句

此语句把指定的变量复制到OUTTREE=数据集中，以备后用。

**例1：**

1、调查了27名沥青工和焦炉工的年龄、工龄、吸烟情况，检测了血清P21、P53、外周血淋巴细胞SVE、染色体畸变数、染色体畸变细胞数等。（据见附件），试将27名工人归类。

|  |
| --- |
| 编号 年龄 工龄 吸烟 血清P21 P21倍数 P53 SCE 染色体畸变数 染色体畸变细数 |
| C1 46 25 5 2138 1.68 0.53 8.11 4 4 |
| C2 35 12 20 3510 2.76 1.43 6.84 3 3 |
| C3 52 25 20 2784 2.19 0.54 4.11 3 3 |
| C4 32 7 20 2451 1.93 0.47 11.45 9 6 |
| C5 38 22 0 3247 2.56 0.8 11.68 5 5 |
| C6 51 31 30 3710 2.92 0.37 11.6 2 2 |
| C7 40 9 10 3194 2.51 0.4 11.4 5 5 |
| C8 34 17 20 4658 3.67 0.46 11.35 3 3 |
| C9 50 29 0 5019 3.95 0.47 13.45 10 8 |
| C10 42 20 20 7482 5.89 0.12 13.11 0 0 |
| C11 57 30 15 3800 2.99 0.19 10.76 2 2 |
| C12 36 15 20 2478 1.95 0.25 10 0 0 |
| C13 37 12 0 3827 3.01 0.82 10.5 4 4 |
| C14 52 32 0 2984 2.35 0.16 11.15 3 3 |
| C15 52 32 10 3749 2.95 0.72 11.45 11 10 |
| C16 42 27 30 4941 3.89 0.73 13.8 7 6 |
| C17 44 27 20 3948 3.11 0.33 13.65 16 14 |
| C18 40 21 5 3360 2.64 0.37 11.4 0 0 |
| C19 38 21 5 2936 2.31 0.69 11.4 1 1 |
| C20 44 27 20 6851 5.39 0.99 12.28 7 6 |
| C21 43 27 0 3926 3.09 0.47 11.95 0 0 |
| C22 26 10 3 4381 3.45 0.52 11.8 7 5 |
| C23 37 18 20 7142 5.62 0.85 11.81 5 5 |
| C24 28 9 20 2612 2.06 0.37 11.65 1 1 |
| C25 25 9 30 2638 2.08 0.78 12.25 1 1 |
| C26 34 14 20 4322 3.4 0.41 15 5 1 |
| C27 50 32 20 2862 2.25 0.69 8.8 2 2 |

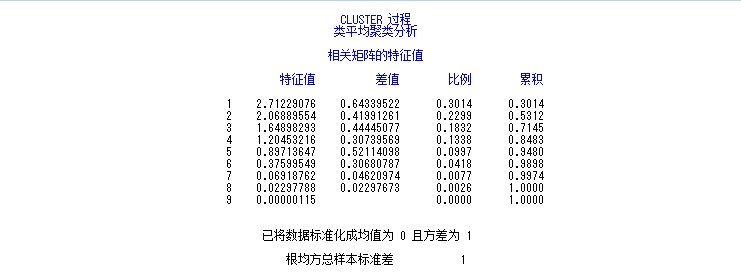


图1 九个聚类变量的相关阵的特征值等信息

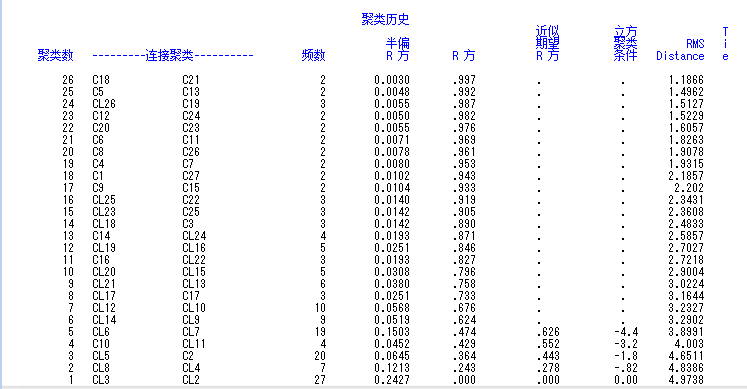


图2 类平均聚类的并类历史过程

**对图2的输出结果进行说明：**

* 1. 连接聚类为每次聚成1个新类的2个样品（标有OB）或旧类（标有CL）；FREQ为新类中所含的样品数；SPRSQ半偏R方表示每次合并对信息的损失程度；RSQ为R方，反映累计聚类结果；ERSQ为在一致无效假设下近似期望的R方；
  2. 根据指标，我们可以看到在倒数第五行，半偏R方、R方、RMS值突然波动，可以决定将26个样本分成6类较为合适；半偏R方表示每次合并对信息的损失程度，可以看到当6类合并成5类和3类合并成2类时，信息损失量最多，故此统计量表明适合聚成6类和3类；R方统计量中6类归成5类和3类归为2类时，R2减少量很多，此项也表明支持聚成6类或3类；
  3. 综上根据各个统计量，运用类平均法可以将样本归为6类或3类。

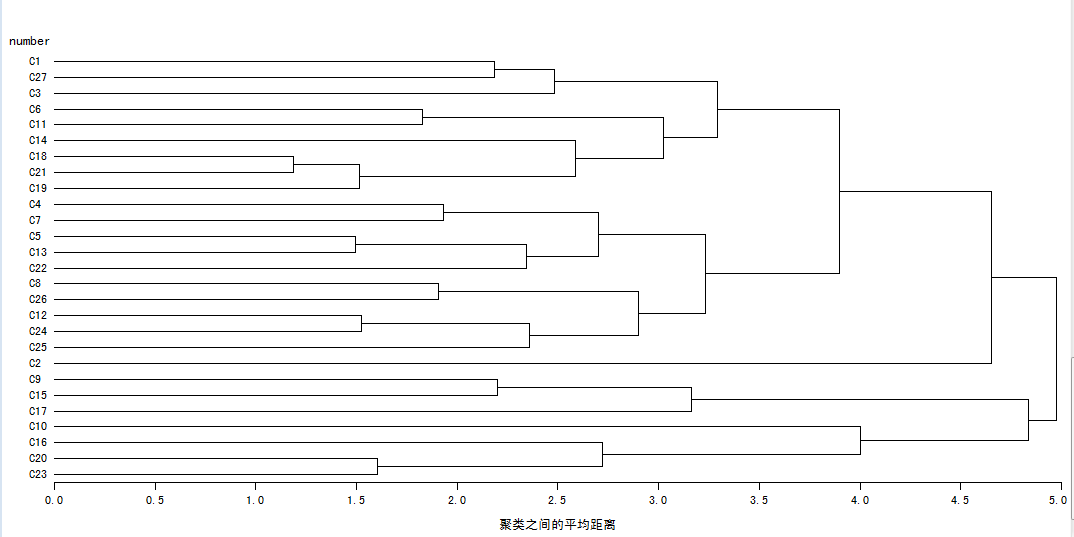


图3 聚类树形图

从图3可知，假设分成3类，则第一类为c1、c27、…c24、c25等19个样本，第二类只有一个样本c2，第三类有c9、c15、…c23等7个样本。如果分为6类，则第一类为c1、c27、 …c19等9个样本 第二类为c4、c7、…c22等5个样本，第三类为c8、c26、…c25等5个样本 ,第四类为c2一个样本，第五类为c9、c15、c17,第六类为c10、…c23等4个样本。

1. **VARCLUS过程（指标聚类）**

# varclus语句格式：

Proc varclus options；

Var variables；

Id variable；

By variable；

Run；

其中“OPTIONS”中可以包括以下内容：

1. CENTRIOD：指定用重心分量作为聚类的类分量；
2. Maxc=n（或maxclusters= n）/minc=n：指定结果中最多(最少)可形成的类个数；
3. HIERARCHY（或HI）：保留不同层级的聚类结果，以保持聚类过程的树状结构。
4. Proportion=n（可以是0-1或者0-100的数，表达意思是一样的），表示聚类的结果必须满足所有类中指标变量的变异至少为n%被类成分所解释，如果一个类的比例小于此值，则继续分解成两类。

VAR语句和ID 语句和上面是相同的，在此不再重复。

**例2**：城镇居民消费水平通常用8项指标来描述：人均粮食支出（x1），人均副食支出（x2），人均烟、酒、茶支出（x3），人均其他副食支出（x4），人均衣着商品支出（x5），人均日用品支出（x6），人均燃料支出（x7），人均非商品支出（x8）。这8项指标间存在着一定的相关性。为了研究城镇居民的消费结构，需将相关性强的指标归并到一起。

**data** test;

input number area $ x1-x6;

cards;

1 北京 190.33 43.77 9.73 60.54 49.01 9.04

2 天津 135.2 36.4 10.47 44.16 36.49 3.94

3 河北 95.21 22.83 9.3 22.44 22.81 2.8

4 山西 104.78 25.11 6.4 9.89 18.17 3.25

5 内蒙 128.41 27.63 8.94 12.58 23.99 3.27

6 辽宁 145.68 32.83 17.79 27.29 39.09 3.47

7 吉林 159.37 33.38 18.37 11.81 25.29 5.22

8 黑龙江 116.22 29.57 13.24 13.76 21.75 6.04

9 上海 221.11 38.64 12.53 115.65 50.82 5.89

10 江苏 144.98 29.12 11.67 42.6 27.3 5.74

11 浙江 169.92 32.75 12.72 47.12 34.35 5

12 安徽 153.11 23.09 15.62 23.54 18.18 6.39

13 福建 144.92 21.26 16.96 19.52 21.75 6.73

14 江西 140.54 21.5 17.64 19.19 15.97 4.94

15 山东 115.84 30.26 12.2 33.61 33.77 3.85

16 河南 101.18 23.26 8.46 20.5 20.5 4.3

17 湖北 140.64 18.47 12.35 18.53 20.95 6.23

18 湖南 164.02 24.74 13.63 22.2 18.06 6.04

19 广东 182.55 20.52 18.32 42.4 36.97 11.68

20 广西 139.08 18.47 14.68 13.41 20.66 3.85

21 四川 137.8 20.74 11.07 17.74 16.49 4.39

22 贵州 121.67 21.53 12.58 14.49 12.18 4.57

23 云南 124.27 19.81 8.89 14.22 15.53 3.03

24 陕西 106.02 20.56 10.94 10.11 18 3.29

25 甘肃 95.65 16.82 5.7 6.03 12.36 4.49

26 青海 107.12 16.45 8.98 5.4 8.78 5.93

27 宁夏 113.74 24.11 6.46 9.61 22.92 2.53

28 新疆 123.24 38 13.72 4.64 17.77 5.75

;

**run**;

**proc** **varclus** data= test outtree=result proportion=**0.75**;

\*其聚类方法为主成分聚类法;

\*proportion表示当能解释方差占总方差的75%时，停止分类，数值可以根据题意或实际自己定义。;

var x1-x6;

**run**;

**proc** **tree** data=result horizontal page=**1** space=**1**;

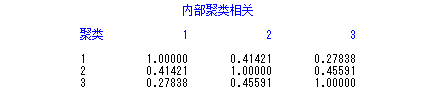
**run**;

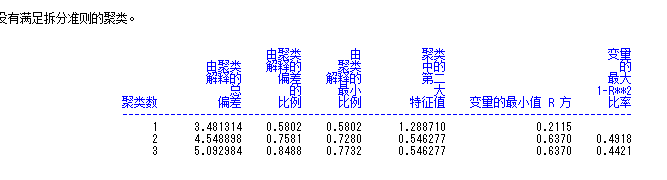
输出结果：



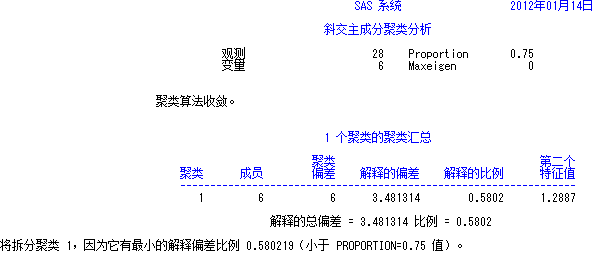
标准化得分系数，是由主成分分析或因子分析得到的类成分关于标准化数据的回归模型系数，用来预测类成分得分。由上图可以得到类成分关于标准化指标变量的线性组合表达式:



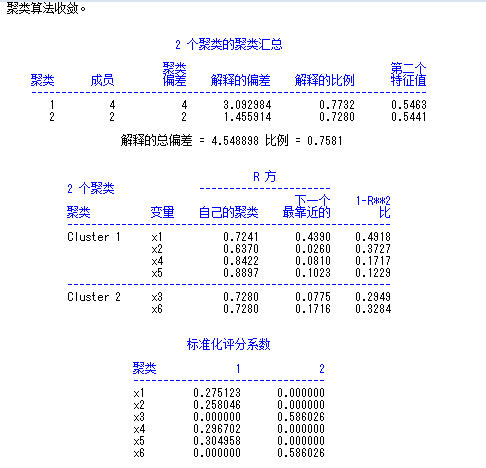


内部聚类相关就是类成分之间的相关系数。此时以达到隐含的停止分裂的临界值，停止分裂。

第二列表示分成一类、二类和三类时分别能解释的总方差量，第三列表示分成一类、二类和三类时分别能解释的方差占全部6个变量指标的的总方差百分比。第四列表示分成一类、二类和三类时有一个类成分（第一主成分）占全部6个变量指标的总方差的最小百分比。第五列为各类中1个变量与其所在类成分的最小相关系数的平方。

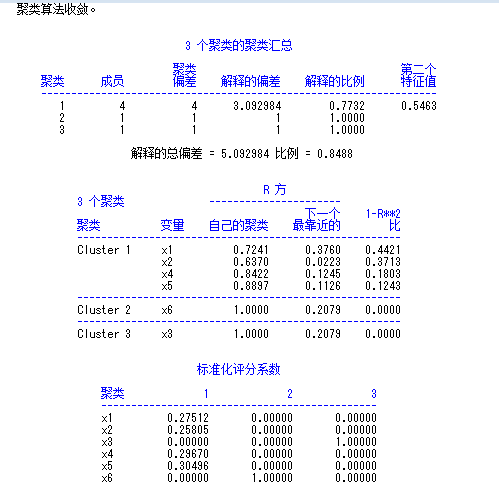
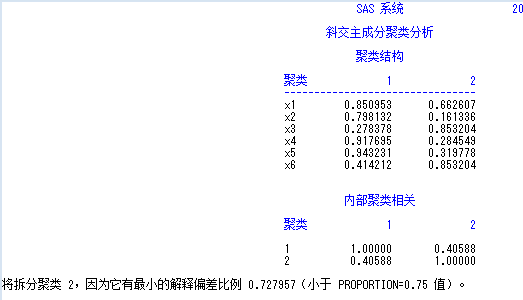


当分为一类时能解释的总方差量为3.481314，占6个变量指标的总方差量的58.02%，没有达到75%，故需要继续分类。

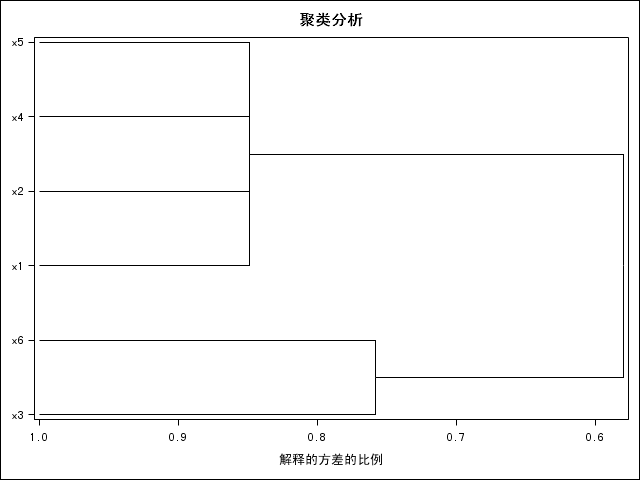
,

聚成两类，分为两类时的方差总方差量占6个指标变量总方差量的75.81%，但是第二个类中指标变量总体变异的72.8%被类解释，小于所定义的75%，虽然第一类的77.32%被解释，大于所定义的75%，故需要再继续分类。

“自己的聚类”这一列，给出了每个指标变量的自身相似系数的平方R2，该值越大，说明该变量越应该分到此类；“下一个靠近的”给出每一个指标变量的相邻相似系数的平方R2，该值越小，说明分类越合理；第三列的值越大，说明分类越合理。



由上图可知，6个指标变量总体方差的84.88%可以被类解释，其中每一类的变量总体变异都大于75%，说明分成三类是合理的。而又由下面的聚类合理性分析可以看出，分成分类是可行的。从中也可以看成每一类的具体指标变量。

****

由树形图可以更见明显地看出分成三类，以及每一类的指标变量。

1. **fastclus进行大样本样品聚类**

FASTCLUS过程叫动态聚类，也叫快速聚类。它是在一个变量或者几个变量的欧氏距离基础之上对数据进行分类，这些类之间互不相交。此过程称为大数据集的聚类，其观测数应在100~100000之间。对于小数据集，此过程对于观测的排序比较敏感。方法支持在过程结束后打印出简单的分类结果。如果想对类进行广泛的检查，则要求指定一个含有类之间关系的变量。

FASTCLUS语句格式：

proc standard data= out= mean=0 std=1;

var 变量;

run;

proc fastclus data= maxc= radius= maxiter= list distance out= outstat= vardef=;

var 变量;

run;

其中：

第一个过程步是将数据标准化，均值为0，标准差为1；（可以不进行标准化）

第二个过程步是用来聚类的，maxc= 用来指定聚类数， radius= 给出确定新凝聚点的准则r，maxiter=给出递推运算次数， list 列出每个样品的归类结果，distance 输出类内样品均值间的距离。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学校 | 声誉得分 | 学术资源得分 | 学术成果得分 | 学生情况得分 | 教师资源得分 | 物资资源得分 |
| 湖南大学 | 66.7 | 29.1 | 28.9 | 87.3 | 54.5 | 44.6 |
| 南京航空航天大学 | 70.7 | 25.3 | 42.1 | 78.9 | 46.2 | 41.4 |
| 浙江大学 | 88.7 | 56.2 | 100.0 | 87.4 | 83.3 | 97.8 |
| 中国科学技术大学 | 91.9 | 100.0 | 71.6 | 100.0 | 65.4 | 51.7 |
| 南京大学 | 88.4 | 72.6 | 64.3 | 94.8 | 88.6 | 63.4 |
| 北京化工大学 | 59.6 | 26.9 | 34.9 | 66.7 | 38.7 | 37.7 |
| 南京理工大学 | 65.3 | 16.2 | 28.4 | 66.5 | 44.4 | 44.8 |
| 华中农业大学 | 58.7 | 34.7 | 17.6 | 68.9 | 42.5 | 38.4 |
| 中国医科大学 | 62.8 | 37.0 | 15.6 | 77.8 | 47.7 | 14.5 |
| 北京航空航天大学 | 81.1 | 44.5 | 55.5 | 93.2 | 58.5 | 74.6 |
| 南开大学 | 84.7 | 59.4 | 41.6 | 90.4 | 69.1 | 56.3 |
| 北京师范大学 | 82.6 | 43.3 | 69.3 | 70.4 | 56.4 | 52.2 |
| 四川大学 | 76.5 | 42.7 | 51.1 | 78.8 | 57.0 | 78.5 |
| 武汉大学 | 82.0 | 41.5 | 47.3 | 77.0 | 60.0 | 76.2 |
| 西安交通大学 | 79.7 | 37.1 | 58.6 | 89.5 | 58.1 | 55.8 |
| 华中科技大学 | 81.1 | 30.8 | 54.9 | 79.0 | 59.3 | 79.0 |
| 厦门大学 | 80.8 | 56.9 | 35.8 | 88.3 | 53.5 | 63.4 |
| 中山大学 | 79.8 | 40.9 | 49.8 | 80.5 | 55.4 | 64.8 |
| 天津大学 | 80.4 | 50.0 | 42.9 | 82.3 | 57.6 | 52.5 |
| 吉林大学 | 75.0 | 40.8 | 43.5 | 78.9 | 53.1 | 71.6 |
| 华东师范大学 | 75.1 | 44.8 | 35.7 | 81.3 | 51.1 | 74.0 |
| 西北工业大学 | 70.7 | 31.4 | 50.2 | 88.3 | 47.8 | 67.9 |
| 大连理工大学 | 71.2 | 35.3 | 56.0 | 89.6 | 48.3 | 47.4 |
| 中南大学 | 71.0 | 37.7 | 43.3 | 81.4 | 59.2 | 56.7 |
| 东南大学 | 77.5 | 22.2 | 58.5 | 87.5 | 52.2 | 50.2 |
| 山东大学 | 75.3 | 35.7 | 43.5 | 71.4 | 56.4 | 64.3 |
| 中国农业大学 | 74.5 | 34.6 | 34.5 | 70.0 | 67.3 | 56.7 |
| 北京理工大学 | 72.0 | 29.0 | 35.3 | 83.9 | 51.1 | 67.6 |
| 同济大学 | 82.2 | 24.6 | 39.2 | 83.3 | 46.8 | 59.2 |
| 华南理工大学 | 69.8 | 29.4 | 38.0 | 77.3 | 45.8 | 64.1 |
| 北京科技大学 | 68.1 | 33.7 | 29.2 | 82.0 | 51.0 | 49.2 |
| 北京邮电大学 | 68.4 | 16.5 | 22.6 | 83.4 | 39.8 | 33.2 |
| 首都医科大学 | 61.5 | 57.4 | 13.9 | 68.8 | 30.9 | 15.0 |
| 北京交通大学 | 65.6 | 38.5 | 28.6 | 78.4 | 45.9 | 50.1 |
| 兰州大学 | 74.2 | 20.5 | 39.5 | 78.2 | 43.2 | 41.1 |
| 东北大学 | 68.9 | 37.7 | 30.0 | 74.9 | 40.3 | 43.0 |
| 电子科技大学 | 67.9 | 24.3 | 34.8 | 85.0 | 41.1 | 42.8 |
| 中国政法大学 | 70.1 | 51.6 | 11.2 | 77.8 | 44.1 | 39.0 |
| 清华大学 | 100.0 | 82.7 | 98.4 | 98.8 | 100.0 | 100.0 |
| 北京大学 | 99.8 | 89.5 | 77.9 | 96.2 | 72.9 | 86.4 |
| 上海交通大学 | 89.4 | 46.5 | 96.9 | 86.3 | 70.8 | 73.4 |
| 复旦大学 | 90.8 | 80.0 | 63.6 | 86.3 | 77.9 | 64.6 |
| 中国人民大学 | 86.7 | 83.6 | 63.7 | 86.1 | 55.5 | 77.2 |
| 哈尔滨工业大学 | 82.8 | 33.2 | 85.5 | 89.5 | 72.0 | 92.5 |
| 北京工业大学 | 59.8 | 20.4 | 21.0 | 68.4 | 45.8 | 47.2 |
| 华东理工大学 | 67.8 | 25.7 | 34.3 | 71.5 | 48.2 | 41.5 |
| 重庆大学 | 65.9 | 22.5 | 30.6 | 76.8 | 47.4 | 50.0 |
| 东北师范大学 | 64.6 | 26.4 | 22.9 | 80.6 | 45.4 | 49.3 |
| 华中师范大学 | 66.4 | 29.6 | 24.6 | 66.8 | 41.1 | 54.8 |
| 中央音乐学院 | 68.5 | 44.5 | 3.1 | 73.8 | 48.4 | 47.9 |
| 西安电子科技大学 | 64.2 | 21.2 | 30.8 | 80.1 | 38.8 | 45.9 |
| 哈尔滨工程大学 | 64.9 | 15.9 | 17.2 | 84.2 | 42.2 | 45.4 |

利用聚类分析的方法，对以上高校进行分类。突出各高校的强势，加强各高校间的交流，提出合理的建议。

**data** test;

input school $ x1 x2 x3 x4 x5 x6;

cards;

湖南大学 66.7 29.1 28.9 87.3 54.5 44.6

南京航空航天大学 70.7 25.3 42.1 78.9 46.2 41.4

浙江大学 88.7 56.2 100.0 87.4 83.3 97.8

中国科学技术大学 91.9 100.0 71.6 100.0 65.4 51.7

南京大学 88.4 72.6 64.3 94.8 88.6 63.4

北京化工大学 59.6 26.9 34.9 66.7 38.7 37.7

南京理工大学 65.3 16.2 28.4 66.5 44.4 44.8

华中农业大学 58.7 34.7 17.6 68.9 42.5 38.4

中国医科大学 62.8 37.0 15.6 77.8 47.7 14.5

北京航空航天大学 81.1 44.5 55.5 93.2 58.5 74.6

南开大学 84.7 59.4 41.6 90.4 69.1 56.3

北京师范大学 82.6 43.3 69.3 70.4 56.4 52.2

四川大学 76.5 42.7 51.1 78.8 57.0 78.5

武汉大学 82.0 41.5 47.3 77.0 60.0 76.2

西安交通大学 79.7 37.1 58.6 89.5 58.1 55.8

华中科技大学 81.1 30.8 54.9 79.0 59.3 79.0

厦门大学 80.8 56.9 35.8 88.3 53.5 63.4

中山大学 79.8 40.9 49.8 80.5 55.4 64.8

天津大学 80.4 50.0 42.9 82.3 57.6 52.5

吉林大学 75.0 40.8 43.5 78.9 53.1 71.6

华东师范大学 75.1 44.8 35.7 81.3 51.1 74.0

西北工业大学 70.7 31.4 50.2 88.3 47.8 67.9

大连理工大学 71.2 35.3 56.0 89.6 48.3 47.4

中南大学 71.0 37.7 43.3 81.4 59.2 56.7

东南大学 77.5 22.2 58.5 87.5 52.2 50.2

山东大学 75.3 35.7 43.5 71.4 56.4 64.3

中国农业大学 74.5 34.6 34.5 70.0 67.3 56.7

北京理工大学 72.0 29.0 35.3 83.9 51.1 67.6

同济大学 82.2 24.6 39.2 83.3 46.8 59.2

华南理工大学 69.8 29.4 38.0 77.3 45.8 64.1

北京科技大学 68.1 33.7 29.2 82.0 51.0 49.2

北京邮电大学 68.4 16.5 22.6 83.4 39.8 33.2

首都医科大学 61.5 57.4 13.9 68.8 30.9 15.0

北京交通大学 65.6 38.5 28.6 78.4 45.9 50.1

兰州大学 74.2 20.5 39.5 78.2 43.2 41.1

东北大学 68.9 37.7 30.0 74.9 40.3 43.0

电子科技大学 67.9 24.3 34.8 85.0 41.1 42.8

中国政法大学 70.1 51.6 11.2 77.8 44.1 39.0

清华大学 100.0 82.7 98.4 98.8 100.0 100.0

北京大学 99.8 89.5 77.9 96.2 72.9 86.4

上海交通大学 89.4 46.5 96.9 86.3 70.8 73.4

复旦大学 90.8 80.0 63.6 86.3 77.9 64.6

中国人民大学 86.7 83.6 63.7 86.1 55.5 77.2

哈尔滨工业大学 82.8 33.2 85.5 89.5 72.0 92.5

北京工业大学 59.8 20.4 21.0 68.4 45.8 47.2

华东理工大学 67.8 25.7 34.3 71.5 48.2 41.5

重庆大学 65.9 22.5 30.6 76.8 47.4 50.0

东北师范大学 64.6 26.4 22.9 80.6 45.4 49.3

华中师范大学 66.4 29.6 24.6 66.8 41.1 54.8

中央音乐学院 68.5 44.5 3.1 73.8 48.4 47.9

西安电子科技大学 64.2 21.2 30.8 80.1 38.8 45.9

哈尔滨工程大学 64.9 15.9 17.2 84.2 42.2 45.4

;

**run**;

**proc** **fastclus** data=test radius=**1** out=result distance maxc=**6**;

\*将聚类结果放到数据集result中;

id school;

var x1-x6;

**run**;

**quit**;

**proc** **sort** data=result out=result;

by cluster;

**proc** **print**;

**run**;

\*对分类结果数据集按分类排序并输出;