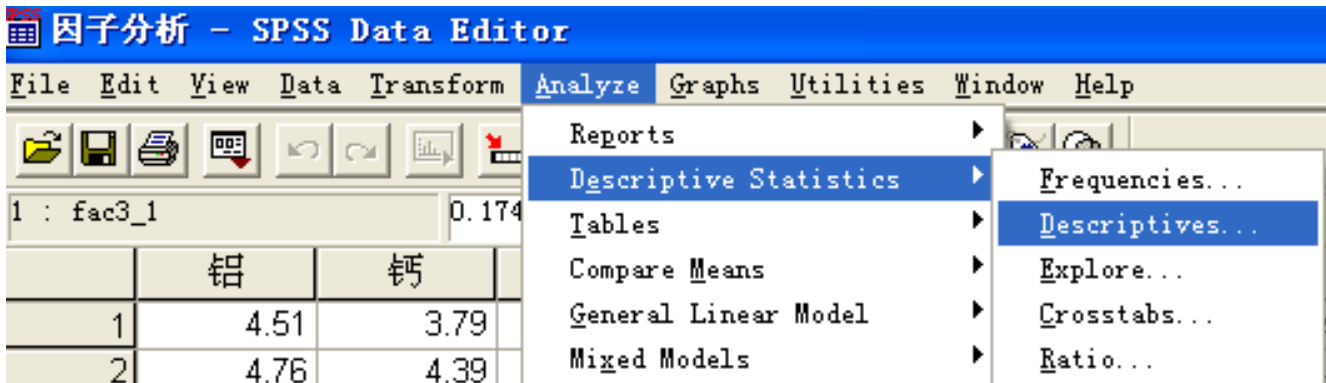


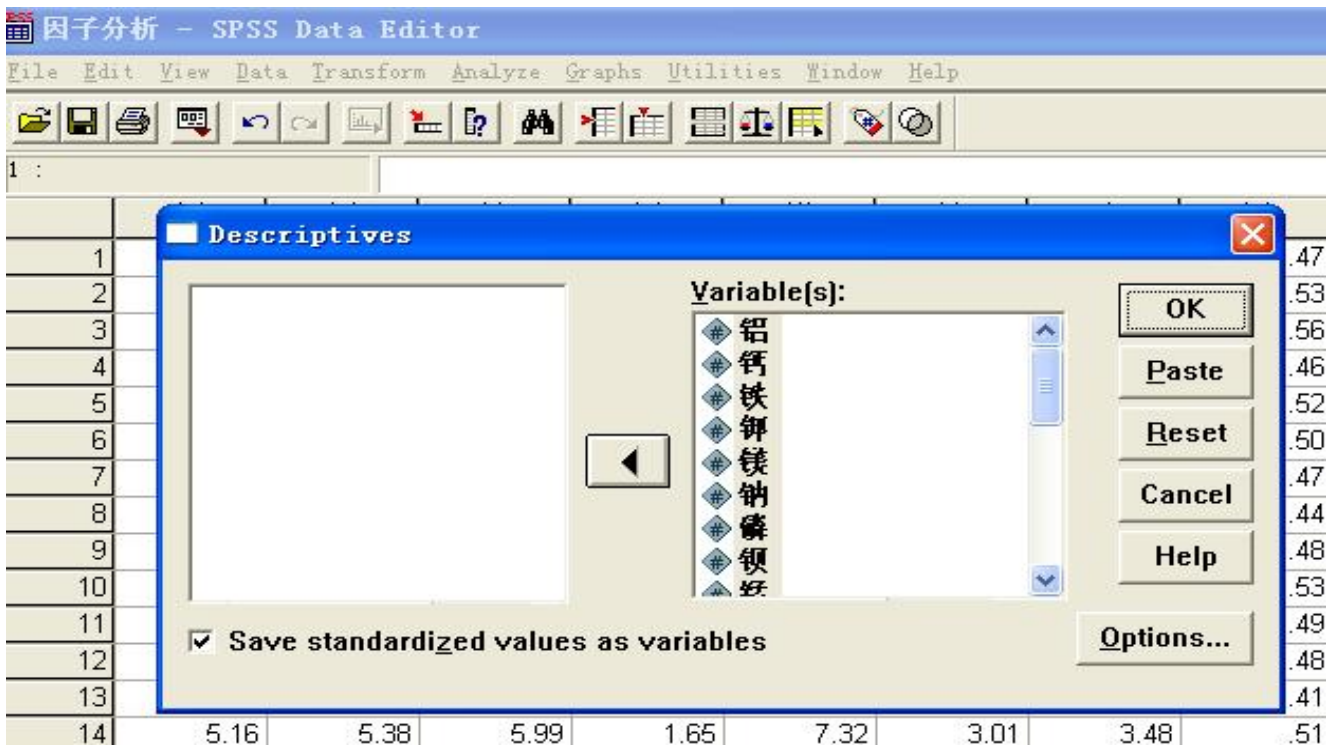
## 1.主因子分析(刘红杰原创)QQ498236930

### 第一步：矩阵标准化

内蒙古第三地质矿产勘查开发院



出现如下对话框：



### 第二步：对标准化过的矩阵分析

**因子分析 - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : fac3\_1 0.174

	铝	钙
1	4.51	3.79
2	4.76	4.39
3	5.49	4.83
4	4.55	4.39
5	5.11	4.16
6	4.97	4.26
7	5.02	4.23
8	4.41	4.07
9	4.69	5.48
10	5.00	5.01

Reports  
Descriptive Statistics  
Tables  
Compare Means  
General Linear Model  
Mixed Models  
Correlate  
Regression  
Loglinear  
Classify  
**Data Reduction**  
Scale  
Nonparametric Tests  
Time Series

钠 磷 钒

	钠	磷	钒
3	3.19	5.72	
10	3.09	4.96	
6	3.33	3.68	
7	2.89	5.35	
8	3.00	4.81	
2	3.05	4.84	
17	3.06	4.30	

Factor...  
Correspondence Analysis...  
Optimal Scaling...

**因子分析 - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Window Help

	镉	钴	铬	铜	镓	镍	铅	锌	z铝
1	1.60	41.02	128.54	264.67	134.33	142.47	6.79	89.92	-1.1
2									-1.1
3									-1.1
4									-1.1
5									-1.1
6									-1.1
7									-1.1
8									-1.1
9									-1.1
10									-1.1
11									-1.1
12									-1.1
13									-1.1
14									-1.1
15									-1.1
16									-1.1
17									-1.1
18									-1.1
19									-1.1
20									-1.1
21	.69	41.37	16						1.63
22	.80	46.51	19						1.08
23	.80	42.20	20						1.00
24									

**Factor Analysis**

Variables:

- Zscore(铝) [z铝]
- Zscore(钙) [z钙]
- Zscore(铁) [z铁]

Descriptives...

**Factor Analysis: Descriptives**

Statistics

- ☐ Univariate descriptives
- ☒ Initial solution

Correlation Matrix

- ☒ Coefficients
- ☒ Inverse
- ☒ Significance levels
- ☐ Reproduced
- ☐ Determinant
- ☐ Anti-image
- ☒ KMO and Bartlett's test of sphericity

Continue  
Cancel  
Help

因子分析 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 :

	镉	钴	铬	铜	镓	镍	铅	锌	z铅
1	1.60	41.02	128.54	264.67	134.33	142.47	6.79	89.92	-1.
2									-1.
3									-1.
4									-1.
5									-1.
6									-1.
7									-1.
8									-1.
9									-1.
10									-1.
11									-1.
12									-1.
13									-1.
14									-1.
15									-1.
16									-1.
17									-1.
18									-1.
19									-1.
20									-1.
21									-1.
22									-1.
23									-1.

**Factor Analysis**

Variables:

- # Zscore(铅) [z铅]
- # Zscore(镉) [z镉]

**Factor Analysis: Extraction**

Method: Principal components

Analyze

- ☒ Correlation matrix
- ☐ Covariance matrix

Display

- ☒ Unrotated factor solution
- ☐ Scree plot

Extract

- ☒ Eigenvalues over: 1
- ☐ Number of factors:

Maximum Iterations for Convergence: 25

OK

Paste

Continue

Cancel

Help

因子分析 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 :

	镉	钴	铬	铜	镓	镍	铅	锌	z铅
1	1.60	41.02	128.54	264.67	134.33	142.47	6.79	89.92	-1.6
2									1.5
3									-1.8
4									1.6
5									1.2
6									1.3
7									1.3
8									1.9
9									1.6
10									1.1
11									1.3
12									1.5
13									2.1
14									1.1
15									-1.8
16									-1.0
17									-1.8
18									1.5
19									1.6
20									1.8
21	.69							73.63	1.5
22	.80	46.51	199.10	133.76	357.47	133.52	7.07	79.08	-1.1

Factor Analysis

Variables:

- # Zscore(铅) [z铅]
- # Zscore(镉) [z镉]

Factor Analysis: Rotation

Method

- ☐ None
- ☒ Varimax
- ☐ Direct Oblimin
- ☐ Quartimax
- ☐ Equamax
- ☐ Promax

Delta: 0 Kappa: 4

Display

- ☒ Rotated solution
- ☐ Loading plot(s)

Maximum Iterations for Convergence: 25

Options...

Continue

Cancel

Help

OK

Paste

Reset

Cancel

Help



因子分析 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 :

	镉	钴	铬	铜	镓	镍	铅	锌	z铅
1	1.60	41.02	128.54	264.67	134.33	142.47	6.79	89.92	-1
2									1
3									1
4									1
5									1
6									1
7									1
8									1
9									1
10									1
11									1
12									1
13									1
14									2
15									1
16									-
17									-
18									-
19									-
20									-

**Factor Analysis**

Variables:

- Zscore(铅) [z铅]
- Zscore(镉) [z镉]

**Factor Analysis: Factor Scores**

☒ Save as variables

Method

- ☒ Regression
- ☐ Bartlett
- ☐ Anderson-Rubin

☒ Display factor score coefficient matrix

Continue Cancel Help

Descriptives... Extraction... Rotation... Scores... Options...

因子分析 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 :

	镉	钴	铬	铜	镓	镍	铅	锌	z铅
1	1.60	41.02	128.54	264.67	134.33	142.47	6.79	89.92	-1
2									1
3									1
4									1
5									1
6									1
7									1
8									1
9									1
10									1
11									1
12									1
13									1
14									2
15									1
16									-
17									-
18									-
19									-
20									-
21	.69	41.37	161.37	111.96	318.92	110.83	8.04	73.63	

**Factor Analysis**

Variables:

- Zscore(铅) [z铅]
- Zscore(镉) [z镉]

**Factor Analysis: Options**

Missing Values

- ☒ Exclude cases listwise
- ☐ Exclude cases pairwise
- ☐ Replace with mean

Coefficient Display Format

- ☒ Sorted by size
- ☐ Suppress absolute values less than: .10

Continue Cancel Help

Descriptives... Extraction... Rotation... Scores... Options...

# 聚类分析

**聚类分析数据 - SPSS Data Editor**

File Edit View Data Transform **Analyze** Graphs Utilities Window Help

11 : 锶 0.331

	铝	钙
1	4.51	3.79
2	4.76	4.39
3	5.49	4.83
4	4.55	4.39
5	5.11	4.16
6	4.97	4.26
7	5.02	4.23
8	4.41	4.07
9	4.69	5.48
10	5.20	5.01
11	4.97	4.61
12	4.85	4.13

Reports  
Descriptive Statistics  
Tables  
Compare Means  
General Linear Model  
Mixed Models  
Correlate  
Regression  
Loglinear  
**Classify**  
Data Reduction  
Scale  
Nonparametric Tests  
Time Series  
Survival  
Multiple Response  
Missing Value Analysis...

TwoStep Cluster...  
K-Means Cluster...  
**Hierarchical Cluster...**  
Discriminant...

	钠	磷
3	3.19	5.72
10	3.09	4.96
6	3.33	3.68
7	2.89	5.35
2	3.03	4.04
17	3.06	4.39
1	2.85	4.99
5	2.75	4.38

**聚类分析数据 - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Window Help

11 : 锶 0.331

**Hierarchical Cluster Analysis**

Variable(s):  
☒ 铝  
☒ 钙  
☒ 铁  
☒ 钾  
☒ 镁  
☒ 钠

Label Cases by:

Cluster  
☐ Cases ☒ Variables

Display  
☒ Statistics ☒ Plots

Statistics... Plots... Method... Save...

OK Paste Reset Cancel Help

	磷	钡	锰	锶	钛
5.72	.47	1.18	.36	2.21	
4.96	.53	.98	.37	2.26	
3.68	.56	.97	.38	2.62	
5.35	.46	1.13	.35	2.51	
4.81	.52	1.07	.35	2.31	
4.59	.50	1.09	.34	2.61	
3.68	.47	1.00	.33	2.38	
2.98	.44	.85	.30	2.02	
4.64	.48	1.17	.35	2.70	

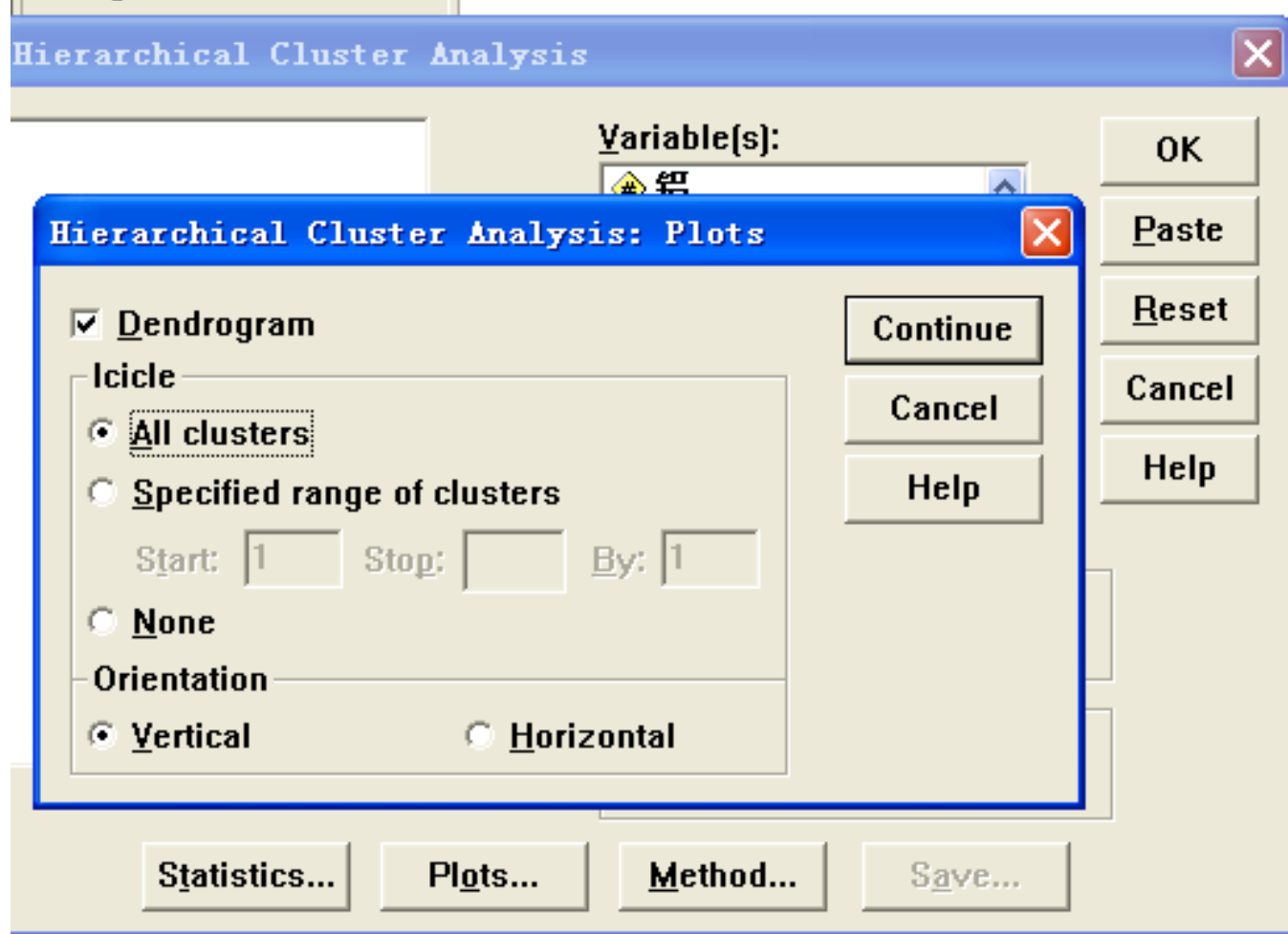
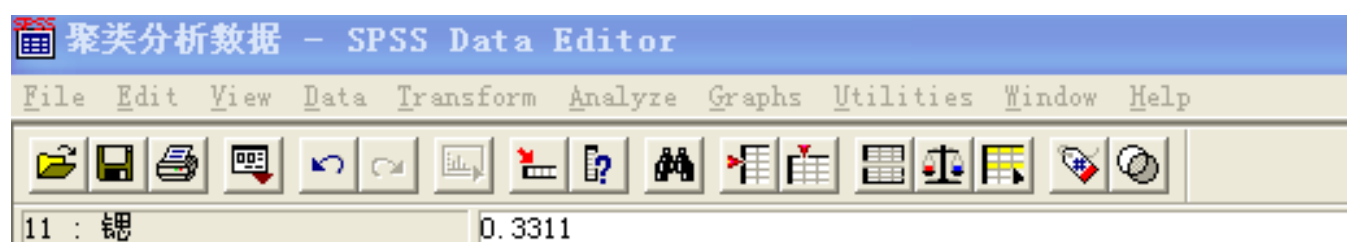
**Hierarchical Cluster Analysis: Statistics**

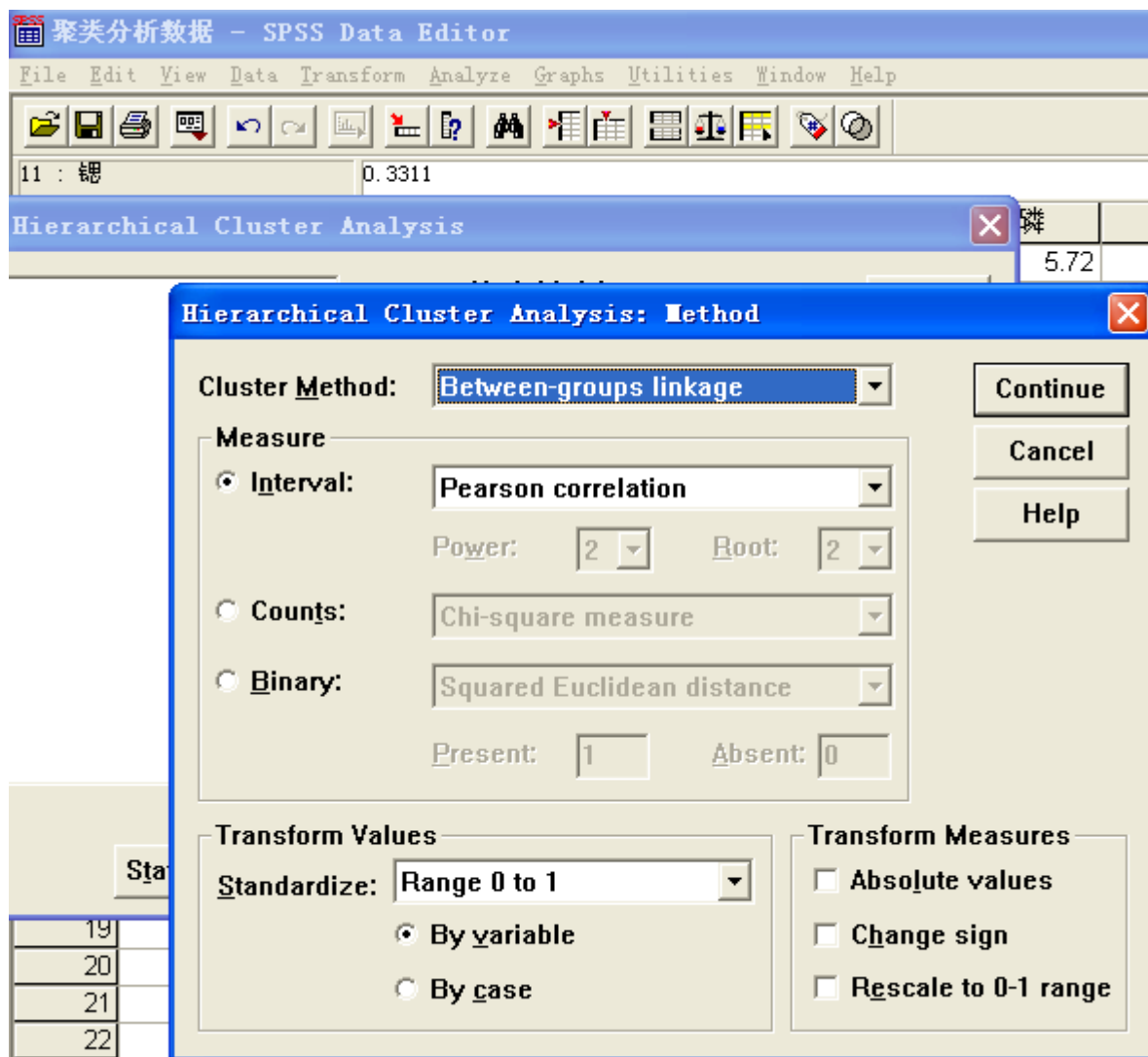
☒ Agglomeration schedule  
☒ Proximity matrix

Cluster Membership  
☐ None  
☒ Single solution: 4 clusters  
☐ Range of solutions: From through clusters

Continue Cancel Help

19	6.94	5.61	9.39	2.05	6.07
20	7.20	6.03	11.04	2.23	5.92
21	6.84	5.86	10.84	1.80	5.54





基于 SPSS 的聚类分析的实用方法（层次聚类法和迭代聚类法）

层次聚类法和迭代聚类法的主要区别在于:层次聚类法的聚类结果受奇异值的影响非常大,且聚类过程是单方向的,一旦某个样本进入某一类,就不可能从该类出来,再归入其他的类;迭代聚类法的聚类结果受奇异值和不合适的聚类变量的影响较小,对于不合适的初始聚类可以进行反复调整,但其缺点是聚类结果对初始聚类非常敏感,而且它也只能得到局部最优解。

### （一）层次聚类

Analyze--> Classify--> Hierarchical Cluster

在“Cluster”组中选择聚类类型：要进行变量聚类选择指定“Variables”；要进行观测聚类指定“Cases”。

指定参与分析的变量，将选定的变量通过按钮箭头转移到箭头按钮右侧的“Variable[s]：”矩形框中；将标识变量通过下面一个按钮箭头转移到按钮右侧的“Label Cases by:”下面的矩形框中。

如果不使用系统默认值，或由于参与分析的变量量纲不一致需要指定选择项，则应该根据需要有选择性地执行下述某些步骤。

#### 1. 确定聚类方法

在主对话框中，点击“Method”按钮，展开分层聚类分析的方法选择对话框，即“Hierarchical Cluster Analysis: Method”。在对话框中根据需要指定聚类方法、距离测度的方法、对数值进行转换方法，即标准化数值的方法和对测度的转换方法。

##### （1）聚类方法选择

“Cluster Method:”表中列出可以选择的聚类方法：

Between-groups linkage 组内连接



Within-groups linkage 组内连接

Nearest neighbor 最近邻法

Furthest neighbor 最远邻法

Centroid clustering 重心聚类法

Median clustering 中位数法

Ward's method Ward 最小方差法。

(后三种聚类方法应与欧氏距离平方法一起使用)

几种方法的具体情况见下面的英文文档

## (2) 对距离的测度方法选择

在 Method 中指定的是用哪两点间的距离的大小决定是否合并两类。距离的具体计算方法还根据参与距离的变量类型从以下三种对话框选择其一，展开选择菜单后再进行具体方法的选择。这三个对话框分别对应于等间隔测度的变量（一般为连续变量）、计数变量（一般为离散变量）和二值变量。这里只考虑连续变量的情况

“Interval”(系统默认)

Euclidean distance: Euclidean 距离，即两观察单位间的距离为其值差的平方和的平方根，该技术用于 Q 型聚类；

Squared Euclidean distance: Euclidean 距离平方，即两观察单位间的距离为其值差的平方和，该技术用于 Q 型聚类；

Cosine: 变量矢量的余弦，这是模型相似性的度量；

Pearson correlation: 相关系数距离，适用于 R 型聚类；

Chebychev: Chebychev 距离，即两观察单位间的距离为其任意变量的最大绝对差值，该技术用于 Q 型聚类；

Block: City-Block 或 Manhattan 距离，即两观察单位间的距离为其值差的绝对值和，适用于 Q 型聚类；

Minkowski: 距离是一个绝对幂的度量，即变量绝对值的第  $p$  次幂之和的平方根； $p$  由用户指定

Customized: 距离是一个绝对幂的度量，即变量绝对值的第  $p$  次幂之和的第  $r$  次根， $p$  与  $r$  由用户指定。

## (3) 确定标准化的方法: “Transform Value”

“Standardize” 下为标准化列表

对数据进行标准化的可选择的方法有：

① None 不进行标准化，是系统默认值。

② Z scores 把数值标准化到 Z 分数。

③ Range -1 to 1 把数值标准化到 -1 到 +1 范围内。选择该项，对每个值用变量或观测量的值的范围去除。如果值范围是 0，所有值保持不变。

④ Maximum magnitud of 1 把数值标准化到最大值为 1。该方法是把正在标准化的变量或观测量的值用最大值去除。如果最大值为 0，则改用最小值去除，其商加 1。

⑤ Range 0 to 1 把数值标准化到 0 到 1 的范围内，对正在被标准化的变量或观测量的值剪去最小值，然后除以范围。如果范围是 0，对变量或观测量的所有值都设置成 0.5。

⑥ Mean of 1 把数值标准化到一个均值的范围内，对正在被标准化的变量或观测量的值除以这些值的均值。如果均值是 0，对变量或观测量的所有值都加 1，使其均值为 1。

⑦ Standard deviation of 1 把数值标准化到单位标准差。该方法对正在被标准化的变量或观测量的值除以这些值的标准差，如果标准差为 0，则这些值保持不变。

## (4) 测度的转换方法选择

对距离测度数值进行转换，在距离计算完成后进行。可选择的转换方法有三种，在“Methd”对话框右下角的标有“Transform Mearure”的框中选择。

① Absolute Values 把距离值标准化。当数值符号表示相关方向，且只对负相关关系感兴趣时使用此方法进行变换。

② Change sign 把相似性值变为不相似性值，或相反。用求反的方法使距离顺序颠倒。

③ Rescale to 0-- 1 range 通过首先去掉最小值然后除以范围的方法使距离标准化。对于已经按某种换算方法标准化了的测度，一般不再使用此方法进行转换。

## 2. 选择要求输出的统计量: Statistics 对话框

Agglomeration schedule 输出聚合过程表

Proximity matrix: 输出的是每个案例之间的欧氏距离平方表 (Q 型聚类)。

Cluster membership 决定聚合的群数。试探性地做时就选 none，做完后根据判断的合适的群数在输入确定的群数，这时会得出一个更多的结果 cluster membership，即在此群数下，各案例所属的群。当然也可选择 Range of solutions 确定群数的范围。

### 3. 选择统计图表: Plot

Dendrogram 树形图;

Icicle 冰柱图:

对于生成什么样的冰柱图还可以进一步用以下选择项确定:

All clusters 聚类的每一步都表现在图中。可用此种图查看聚类的全过程。但如果参与聚类的个体很多会造成图过大，没有必要。可以使用下面一个选择项限定显示的范围。

Specified range of clusters 指定显示的聚类范围。当选择此项时，该项下面的选择框加亮，表示等待输入显示范围。在 Start 后的矩形框中输入要求显示聚类过程的起始步数，在 Stop 后的矩形框中输入显示中止于哪一步，把显示的两步之间的增量输入到 By 后面的矩形框中。输入到矩形框中的数字必须是正整数。

例如，输入的结果是：Start: 3 Stop: 10 By: 2

生成的冰柱图从第三步开始，显示第三、五、七、九步聚类的情况。

None:不生成冰柱图

对于显示方向可以用 Orientation 下面的选择项确定:

Vertical 纵向显示的冰柱图。(系统默认)

HoriZontal 显示水平的冰柱图。

### 4. 生成新变量的选择: save

聚类分析的结果可以用新变量保存在工作数据文件中。单击主对话框的“save”按钮，展开相应的对话框。可以看出只能生成一个表明参与聚类的个体最终被分配到哪一类的新变量。通过对话框可以选择是否建立新变量和建立的新变量含义。

None 不建立新变量。

Single solution: 单一结果。生成一个新变量表明每个个体聚类最后所属的类。在该项后面的矩形框中指定类数。如果指定 5 clusters，则新变量的值为 1-- 5。

Range of solutions: 指定范围内的结果。生成若干个新变量，表明聚为若干个类时，每个个体聚类后所属的类。在该项后面的矩形框中指定显示范围，即把表示从第几类显示到第几类的数字分别输入到 From 后面的矩形框和 through 后面的矩形框中。例如输入结果是“From 4 through 6”，在聚类结束后在数据窗中原变量后面增加了 3 个新变量分别表明分为 4 类时、分为 5 类时和分为 6 类时的聚类结果。即聚为 4、5、6 类时各观测量分别属于哪一类。

新变量选择完成后按“Continue”按钮，返回到主对话框。

#### (二) 迭代聚类

Analyze--> Classify--> K-Means Cluster

“Method”框，给出两个可选择的聚类方法:

1) Iterate and classify 选择初始类中心，在迭代过程中使用 k-Means 算法不断更换类中心，把观测量分派到与之最近的以类中心为标志的类中去;

2) Classify only 只使用初始类中心对观测量进行分类。

Number of clusters 输入通过层次聚类或其他方式得出的聚类的合适的层数。

“Cluster Centers”对话框有两项:

1) 选择 Read initial from 要求使用指定数据文件中的观测量作为初始类中心。选择此项单击鼠标键后，再按其后的“File”按钮，显示选择文件的对话框，指定文件所在位置（路径）和文件名。按“OK”按钮返回。在“Center”选择框中的“file”按钮后面显示文件全名（包括路径）。

2) 选择 Write final as 要求把聚类结果中的各类中心数据保存到指定的文件中。操作方法与上相似。

save 对话框中有两项:

1) 选择 Cluster Membership 建立一个新变量，系统默认变量名为那 qc1\_1。其值表示聚类结果，即各观测量被分配到哪一类。其值为 1、2、3...的序号。该变量存入输入数据文件（New Data 窗中）。

2) 选择 Distance from cluster center 建立一个新变量。系统默认变量名为那 qc1\_2。聚类结束后把各观测量距所属类中心

间的欧氏距离存入输入数据文件（数据窗中）。

**Iterate 对话框：**

设置迭代参数的对话框。如果选择了“**Iterate and classify**”方法进行聚类，还可以进一步选择迭代参数。

1) **Maximum Iterations:** 限定 K-Means 算法中的迭代次数。改变后面框中的数字，则改变迭代次数。当达到限定的迭代次数时迭代停止。系统默认值为 10。

2) **Convergence Criterion:** 指定聚类判据。其值必须大于 0，小于 1。系统默认值为 0.02。即当两次迭代计算的最小的类中心的变化距离小于初始类中心距离的百分之二时迭代停止。

3) **Use Running means** 选择此项，限定在每个观测量被分配到一类后即刻计算新的类中心。如果不选择此项，则在完成了所有观测量的一个分配后再计算各类的类中心。不选择此项会节省迭代时间。

**Option 对话框：**输出统计量的选择项与缺失值处理

在 **Statiststacs** 组中可以选择要求计算和输出的统计量有：

**Initial cluster centers** 初始类中心

**ANOVA table** 方差分析表

**Cluster information for each case** 每个观测量的分类信息。如分配到哪一类和该观测量距所属类中心的距离。

以上三项可以选择其中几项，也可以全选。

在 **Missing Values** 组中选择一种处理带有缺失值观测量的方法，共有两种选择：

1) **Exclude cases listwise** 将带有缺失值的观测量从分析中剔除。

2) **Exclude cases pairwise** 只有当一个观测量的全部聚类变量值缺失时才将其从分析中剔除。否则根据其它非缺失变量值把它分配到最近的一类中去。

其它有用的使用：

1) 筛减 cases:

因为聚类分析计算的速度相对比较慢，而且如果 case 太多，结果也难于解释。有时可以将庞大的样本中进行随机筛选，取出小样本，进行聚类分析。下面是一种方法：

**data-->select cases**, 选中 **random sample of cases**, 点击 **sample** 复选框，弹出一个新的选框。Approximately( )% of all cases. 在括号里填入从所有样本中选取的案例百分比。然后 ok。

另外，在 **unselected cases are** 选框下有两种选择，**filtered** 并不从原数据中将没有选中的 case 删除，只是“过滤”掉了；**deleted** 则是删掉了，然后形成一个新的数据。

2) 求“群重心”的步骤：

之前须先将聚类结果保存在了原数据中。这里设保存的结果列命名为“cluster-3”。

**Analyse-->compare means-->means**

**Dependant list** 选入准备参与比较的聚类变量和非聚类变量。**Independent list** 选入保存的聚类结果“cluster-3”。

**Option** 复选框，**cell statistics** 选入 **mean**, 然后 **continue,ok**.

3) 使用 k-means 确定初始群重心时，

**centers-->initial from format** 选定，未知的在于 file 的格式。简介如下：

第一列：表示分成的群组数，如 1、2、3...

以后各列为各聚类变量，注意与所在群组相对应。如果实在搞不定，可以先不选择 **initial from format**，而选下面的 **write final as**，呵呵，知道什么意思吗，先试一下，以确定正确的格式，再把数据替换一下就可以了。