**多元统计分析 作业4 聚类分析**

地信一班 罗皓文 15303096

**实验环境：**

OS: Windows 7 Pro SP1 x64

CPU: Intel Xeon E3-1241 v3 @ 3.50GHz 3.50GHz

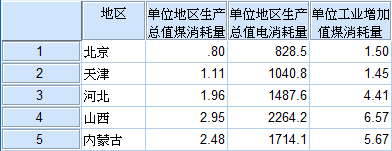
RAM: 8.00 Gb

Soft: SPSS Statistics 19

**一、系统聚类分析**

用“例5.2.sav”数据做系统聚类分析；

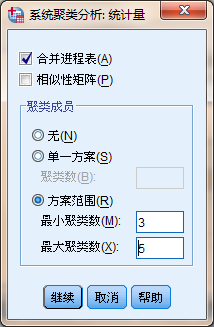
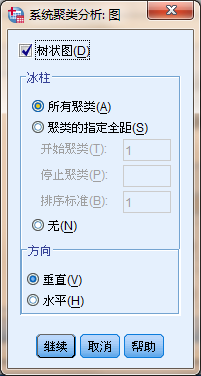
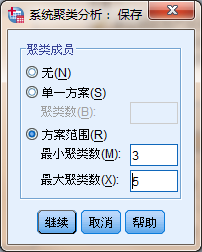
部分数据示例：



数据变量视图：

（1）使用SPSS系统聚类模块：分析(A) - 分类(F) - 系统聚类...(H)

（2）选择变量，指定输出和聚类参数，选择组间连接方法，使用欧氏距离



得到聚类结果：

1.凝聚过程表：

如表1，从该表“凝聚顺序表”可知具体的聚类过程。

| **表1 Agglomeration Schedule** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Stage | Cluster Combined | | Coefficients | Stage Cluster First Appears | | Next Stage |
| Cluster 1 | Cluster 2 | Cluster 1 | Cluster 2 |
| 1 | 16 | 23 | 2.233 | 0 | 0 | 12 |
| 2 | 8 | 9 | 3.372 | 0 | 0 | 13 |
| 3 | 10 | 19 | 8.775 | 0 | 0 | 6 |
| 4 | 15 | 18 | 12.107 | 0 | 0 | 7 |
| 5 | 2 | 7 | 18.742 | 0 | 0 | 7 |
| 6 | 10 | 30 | 40.632 | 3 | 0 | 11 |
| 7 | 2 | 15 | 82.564 | 5 | 4 | 13 |
| 8 | 11 | 17 | 153.252 | 0 | 0 | 11 |
| 9 | 6 | 26 | 338.923 | 0 | 0 | 19 |
| 10 | 13 | 22 | 382.170 | 0 | 0 | 16 |
| 11 | 10 | 11 | 502.346 | 6 | 8 | 16 |
| 12 | 16 | 20 | 641.040 | 1 | 0 | 18 |
| 13 | 2 | 8 | 958.731 | 7 | 2 | 15 |
| 14 | 14 | 21 | 2916.311 | 0 | 0 | 20 |
| 15 | 2 | 12 | 3212.477 | 13 | 0 | 20 |
| 16 | 10 | 13 | 3979.857 | 11 | 10 | 18 |
| 17 | 24 | 27 | 4957.292 | 0 | 0 | 24 |
| 18 | 10 | 16 | 7879.253 | 16 | 12 | 23 |
| 19 | 3 | 6 | 8514.451 | 0 | 9 | 23 |
| 20 | 2 | 14 | 10646.757 | 15 | 14 | 22 |
| 21 | 5 | 25 | 11995.307 | 0 | 0 | 26 |
| 22 | 1 | 2 | 36827.193 | 0 | 20 | 25 |
| 23 | 3 | 10 | 50584.899 | 19 | 18 | 25 |
| 24 | 4 | 24 | 54879.839 | 0 | 17 | 28 |
| 25 | 1 | 3 | 85388.787 | 22 | 23 | 26 |
| 26 | 1 | 5 | 292360.596 | 25 | 21 | 28 |
| 27 | 28 | 29 | 1430209.203 | 0 | 0 | 29 |
| 28 | 1 | 4 | 1573615.897 | 26 | 24 | 29 |
| 29 | 1 | 28 | 10036226.822 | 28 | 27 | 0 |

2.各聚集组员

如表2所示，该表显示了在不同聚类数条件下，各个样本的聚类标签。

3.垂直冰柱图

如图1所示，看图时应从下往上看，两个省份之间的黄柱所对应的聚类数，即为两个省份在划分为此聚类数时属于同一类，并且在此以后一直属于同一类。

| **表2 Cluster Membership** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Case | 5 Clusters | 4 Clusters | 3 Clusters | Case | 5 Clusters | 4 Clusters | 3 Clusters |
| 1:北京 | 1 | 1 | 1 | 16:河南 | 1 | 1 | 1 |
| 2:天津 | 1 | 1 | 1 | 17:湖北 | 1 | 1 | 1 |
| 3:河北 | 1 | 1 | 1 | 18:湖南 | 1 | 1 | 1 |
| 4:山西 | 2 | 2 | 2 | 19:广东 | 1 | 1 | 1 |
| 5:内蒙古 | 3 | 1 | 1 | 20:广西 | 1 | 1 | 1 |
| 6:辽宁 | 1 | 1 | 1 | 21:海南 | 1 | 1 | 1 |
| 7:吉林 | 1 | 1 | 1 | 22:重庆 | 1 | 1 | 1 |
| 8:黑龙江 | 1 | 1 | 1 | 23:四川 | 1 | 1 | 1 |
| 9:上海 | 1 | 1 | 1 | 24:贵州 | 2 | 2 | 2 |
| 10:江苏 | 1 | 1 | 1 | 25:云南 | 3 | 1 | 1 |
| 11:浙江 | 1 | 1 | 1 | 26:陕西 | 1 | 1 | 1 |
| 12:安徽 | 1 | 1 | 1 | 27:甘肃 | 2 | 2 | 2 |
| 13:福建 | 1 | 1 | 1 | 28:青海 | 4 | 3 | 3 |
| 14:江西 | 1 | 1 | 1 | 29:宁夏 | 5 | 4 | 3 |
| 15:山东 | 1 | 1 | 1 | 30:新疆 | 1 | 1 | 1 |

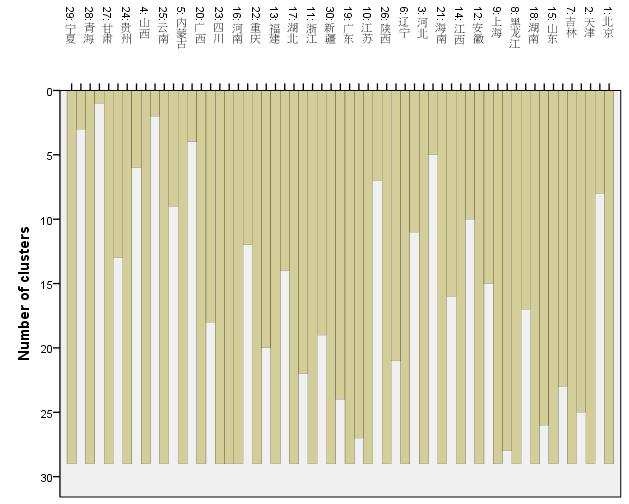


图1 垂直冰柱图

4.树状图/谱系图

与冰柱图类似，结果如图2所示。

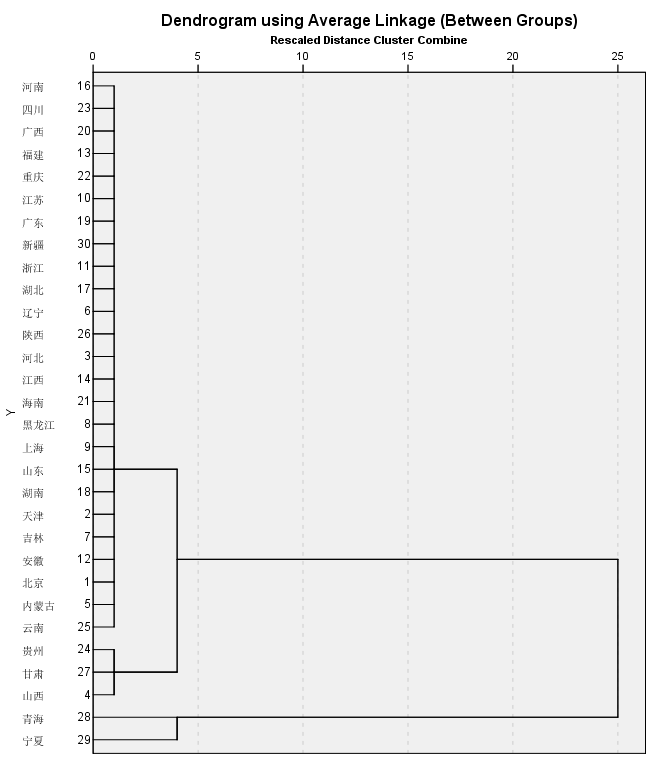


图2 垂直冰柱图

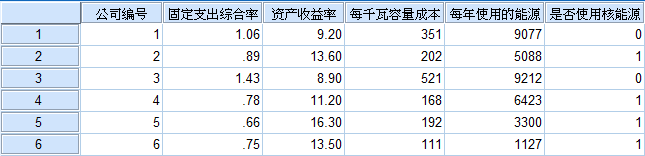
最终结果发现，依据煤电使用情况，可将各省市分为4类：

青海、宁夏各自成一类，贵州、甘肃、山西为一类，其余省市为一类。

**二、K均值聚类分析**

用“作业k均值聚类.sav”数据做K均值聚类分析。

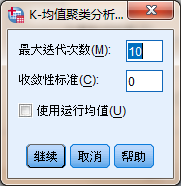
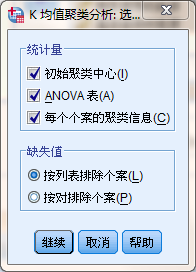
部分数据示例：

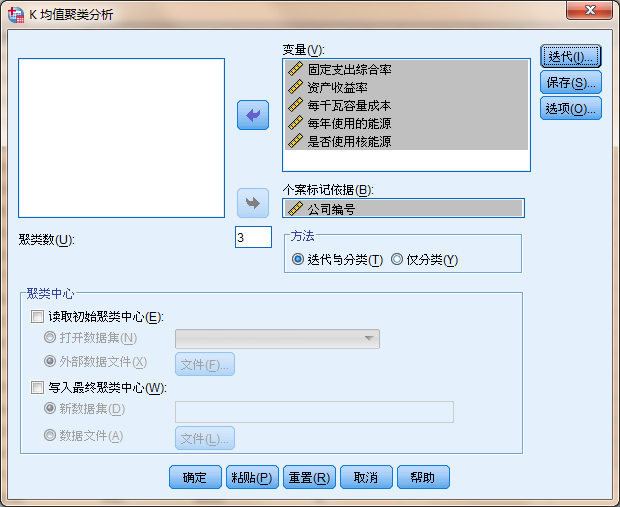


数据变量视图：

（1）使用SPSS的K-均值聚类模块：分析(A) - 分类(F) - K均值聚类...(K)

（2）选择变量，指定聚类参数，选择组间连接方法，使用欧氏距离



结果分析

1.初始聚类中心

从表3可知初始和最终的聚类中心，3类中心位置同初始位置相比，均发生了变化。

2.聚类成员分析

从表4知，每一个样品属于哪一类，还可以知道每一个样品到最终聚类中心的距离。

3.每个聚类中的样本数

从表5知，聚类1所包含样本数最多，聚类3包含样本数最少。

| **表3 Cluster Centers** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Initial Cluster Centers** | | | **Final Cluster Centers** | | |
| 1 | 2 | 3 | 1 | 2 | 3 |
| 固定支出综合率 | 1.09 | .36 | 1.25 | 1.13 | .67 | 1.17 |
| 资产收益率 | 6.10 | 16.30 | 7.50 | 7.72 | 14.51 | 7.88 |
| 每千瓦容量成本 | 296 | 184 | 376 | 527 | 170 | 319 |
| 每年使用的能源 | 9673 | 1093 | 17441 | 8153 | 2744 | 15005 |
| 是否使用核能源 | 0 | 1 | 0 | 0 | 1 | 0 |

| **表4 Cluster Membership** | | | |
| --- | --- | --- | --- |
| Case Number | 公司编号 | Cluster | Distance |
| 1 | 1 | 1 | 940.697 |
| 2 | 2 | 2 | 2343.881 |
| 3 | 3 | 1 | 1059.020 |
| 4 | 4 | 1 | 1766.950 |
| 5 | 5 | 2 | 556.092 |
| 6 | 6 | 2 | 1618.422 |
| 7 | 7 | 1 | 1283.657 |
| 8 | 8 | 3 | 1924.654 |
| 9 | 9 | 1 | 262.947 |
| 10 | 10 | 2 | 711.170 |
| 11 | 11 | 3 | 2436.429 |
| 12 | 12 | 1 | 2000.241 |
| 13 | 13 | 2 | 1569.596 |
| 14 | 14 | 1 | 1537.520 |
| 15 | 15 | 2 | 276.423 |
| 16 | 16 | 3 | 987.993 |
| 17 | 17 | 2 | 1969.967 |
| 18 | 18 | 1 | 1994.825 |
| 19 | 19 | 3 | 1500.523 |
| 20 | 20 | 2 | 465.877 |
| 21 | 21 | 1 | 1505.427 |
| 22 | 22 | 2 | 1651.391 |

| **表5 Final Cluster Centers** | | | |
| --- | --- | --- | --- |
|  | Cluster | | |
| 1 | 2 | 3 |
| 固定支出综合率 | 1.13 | .67 | 1.17 |
| 资产收益率 | 7.72 | 14.51 | 7.88 |
| 每千瓦容量成本 | 527 | 170 | 319 |
| 每年使用的能源 | 8153 | 2744 | 15005 |
| 是否使用核能源 | 0 | 1 | 0 |