利用平均得到更精确的聚类结果

1. 【3.xls】经过极差正规化后的数据为

年份 X1 X2 X3

1993 0 0 0.5

1994 0.25 0.667 0

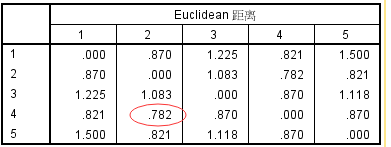
1995 0.5 1 1

1996 0.75 0.333 0.5

1997 1 1 0

（年份作为样本号，不参与计算。X1，X2，X3为变量）

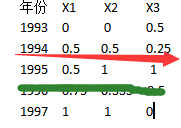
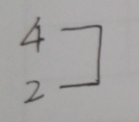
①按照Q型聚类，欧式距离得到结果



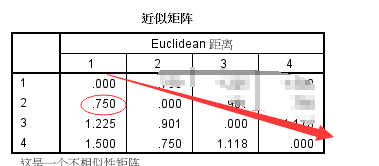
发现第4行第2列的值0.782最小，所以第一次合并把4和2合并起来，距离系数为0.782

②合并完之后，4和2是一群，把第4行(1996年)和第2行的**极差正规化数据**对应相加，求平均值，得到的结果为2的数据，再把第4行去掉。即：

（0.75+0.25）/2 = 0.5 (0.333+0.667)/2=0.5 (0.5+0)/2=0.25

③对上述的数据再进行欧式距离的计算，得到：



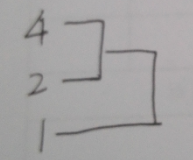
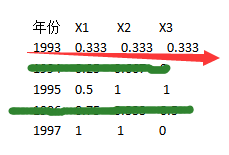
发现第2行第1列0.75最小，进行所以把2和1进行合并，距离系数为0.750

④合并完后对新的一群，**注意**，现在不是2和1是一群，而是4,2,1是一群，因为之前2和4合并了。所以要把4,2,1的极差正规化的数据相加，求平均值（除以3），再把第2行去掉。即：

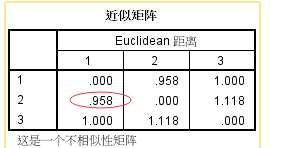
(0+0.25+0.75)/3=0.333

(0+0.667+0.333)/3=0.333

(0.5+0+0.5)/3=0.333



⑤对上述的数据进行求欧式距离：



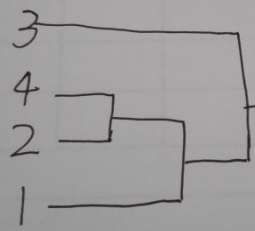
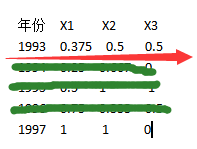
0.958，即原来的第三行第1列的数据最小，对3和1进行合并，距离系数为0.958。

⑥合并完3和1后，4,2,3,1是一个新的群，对这新的群进行数据的平均，得到的结果为1的数据，然后把3去掉。

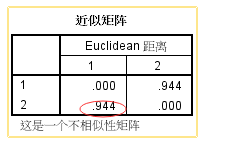
(0+0.25+0.75+0.5)/4=0.375

(0+0.667+0.333+1)/4=0.5

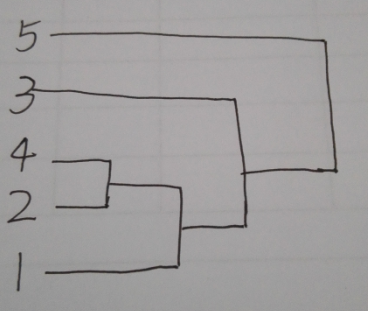
(0.5+0+0.5+1)/4=0.5



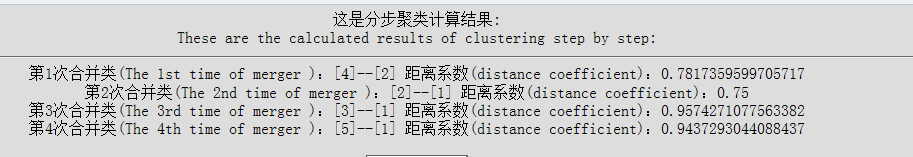
⑦对上述的数据再进行欧式距离的计算，得到：



即5和1合并，最后变成了一个群，结束。



老师的结果：

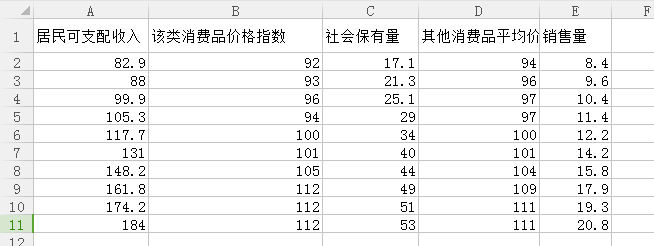




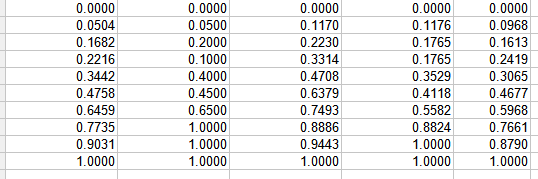
老师的图出bug了。

【聚类分析.xls的分析过程】

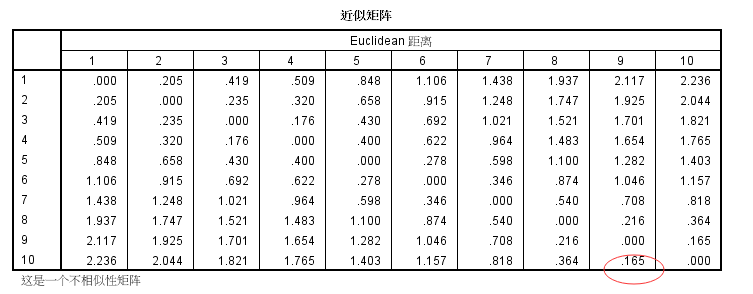
原始数据：



经过极差规格化之后：



①第一次欧式距离计算：

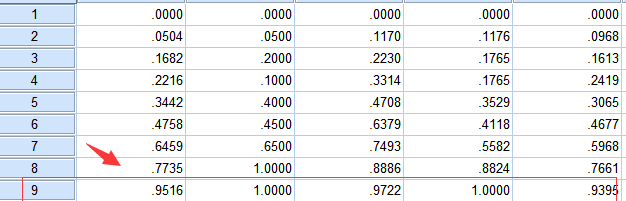


1. 第一次合并：10和9合并起来，距离系数为0.165。

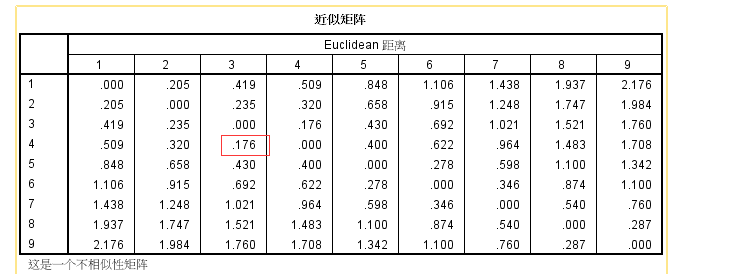


1. 把第10行的数据对应加到第9行中，并求平均数，得到的结果作为第9行的数据。然后再去掉第10行，得到，如下数据：

（1+0.9031）/2 = 0.9516……



②第二次欧式距离计算

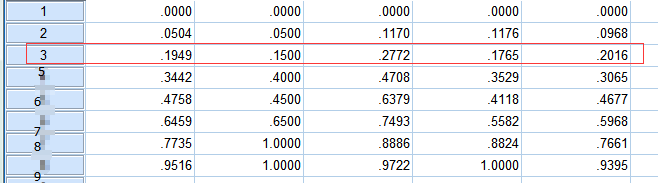


（1）第二次合并，把第4行和第3行合并，距离系数为0.176。

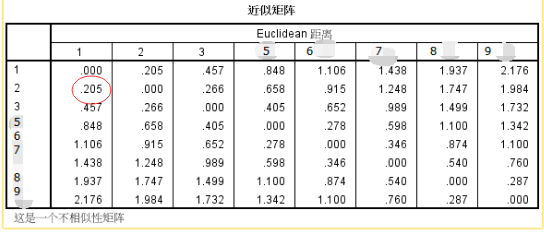


1. 把第4行的数据对应加到第3行中，并求平均数，得到的结果作为第3行的数据。然后再去掉第4行，得到，如下数据：

（0.2216+0.1682）/2 = 0.1949……



③第三次欧式距离：

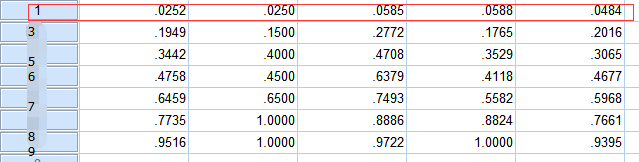


1. 第2行第1列的数据最小，2和1合并，距离系数为0.205

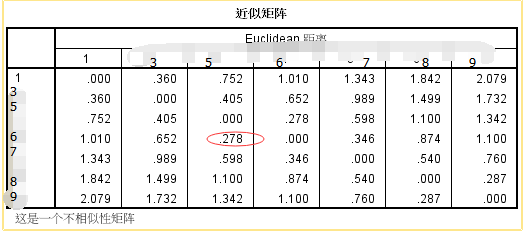


1. 把第2行的数据对应和第1行相加，并且求平均数，求完之后，作为第1行的数据。然后把第2行去掉。

（0+0.504）/2 = 0.0252



④第四次欧式距离：

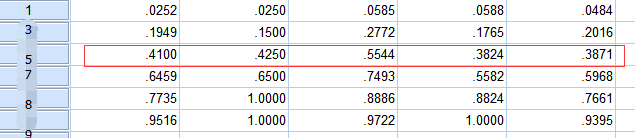


1. 发现第6行第5列的数据最小，距离系数为0.278，6和5合并：

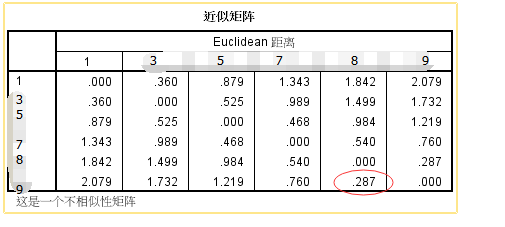


1. 把第6行的数据对应和第5行相加，并且求平均数，求完之后，作为第5行的数据。然后把第6行去掉。

（0.4758+0.3442）/2 = 0.4100 ……



⑤第5次欧式距离：

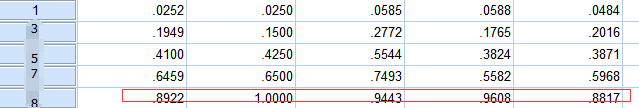


1. 第9行第8列的数据0.287最小，因此9和8合并，距离系数为0.287。

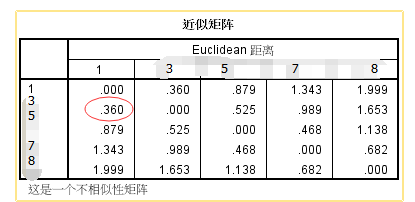


（2）10,9,8是一组，因此，要把10,9,8的数据相加，求平均数，作为第8行的数据，然后把第9行去掉。

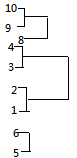
（1+0.9031+0.7735）/3 = 0.8922 ……



⑥第6次欧式距离

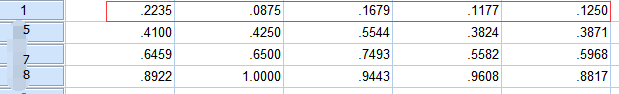


1. 发现第3行第1列的数据最小，所以3和1合并，距离系数为0.360

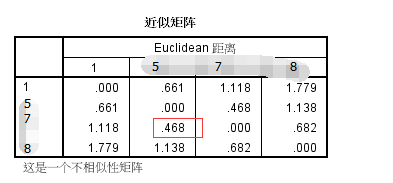


（2）3和1合并后，4,3,2,1成了新的一个群，所以把4,3,2,1的对应行数据相加，得到平均数，作为第1行的数据，然后把第3行去掉。

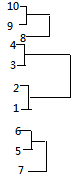
（0+0.504+0.1682+0.2216）/4 = 0.2235



⑦第7次欧式距离：

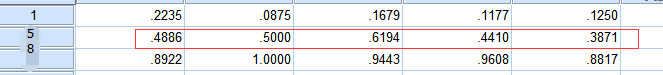


1. 发现地7行第5列的数据最小，距离系数为0.468，将7和5合并。

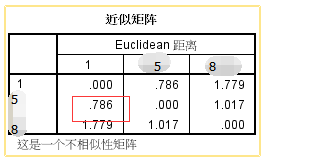


1. 合并后，5,6,7连成一个整体。对5,6,7对应数据相加，求平均数，作为第5行的数据，然后删掉第7行的数据。

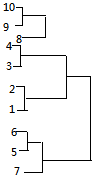
（0.3442+0.4758+0.6459）/3 = 0.4886 ……



⑧第8次欧式距离：



1. 发现第5行第1列的数据最小。将5和1合并，距离系数为0.786。

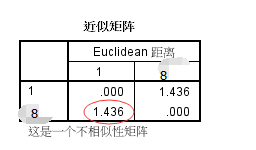


1. 合并后，1,2,3,4,5,6,7成为一个新的整体。对1,2,3,4,5，6,7的数据对应相加，求平均数，结果作为第1行的数据，第5行去掉。

（0+0.504+0.1682+0.2216+0.3442+0.4758+0.6459）/7=0.3371……

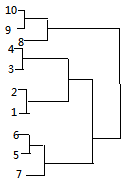


⑨第9次欧式距离：

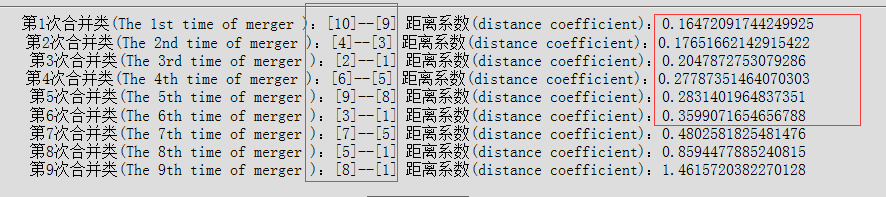


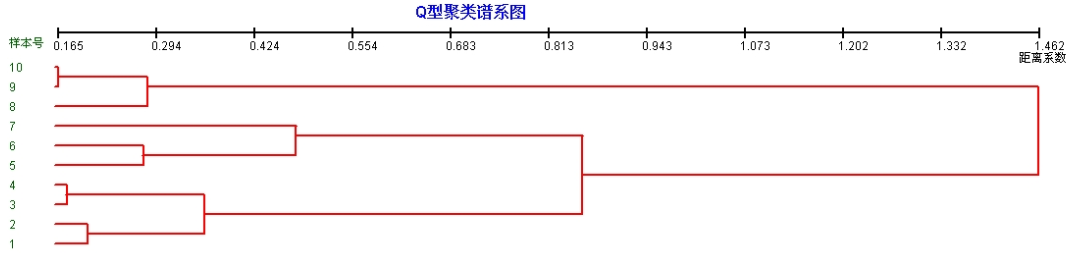
1和8合并，距离系数为1.436。

至此，全部数据合并成了1个群。



老师的结果：（合并一样，数据也差不多一样）



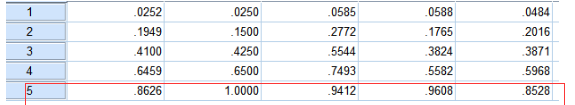


【注】

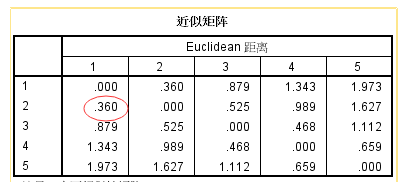
如果只是用两个两个合并后的数据的话，结果反而会更接近。

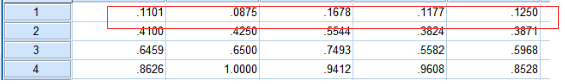
即在第5、6、7、8、9合并时候，就对他们要合并的两行（9,8），（3,1），（7,5），（5,1），（8,1）两行对应相加，求平均数。结果好像会更接近。为了方便，你用这种方法

第5次结果：

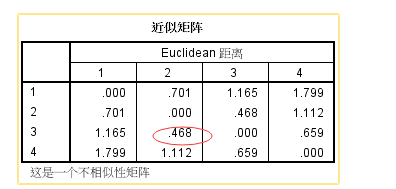


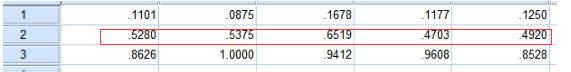
第6次：



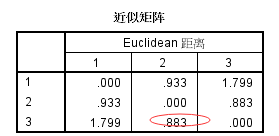


第7次：





第8次：





第9次：

