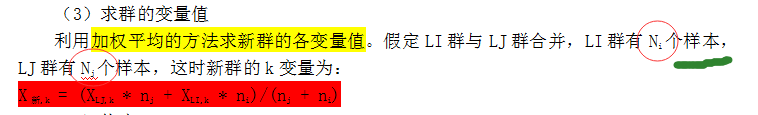
利用平均得到更精确的聚类结果



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

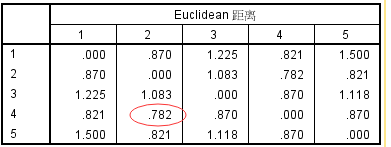
1995 0.5 1 1

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）

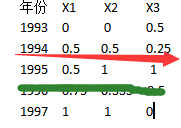
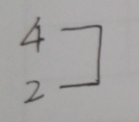
①按照Q型聚类，欧式距离得到结果



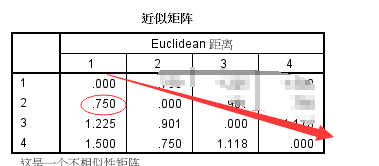
发现第4行第2列的值0.782最小，所以第一次合并把4和2合并起来，距离系数为0.782

②合并完之后，求新的群X4，2的变量值，4的样本数和2的样本数都是自己，即Ni=1,Nj=1,代入公式后把第4行去掉，新群X4,2的样本数为2（1+1），各个变量值为：

（0.75\*1+0.25\*1）/（1+1） = 0.5 (0.333\*1+0.667\*1)/（1+1）=0.5 (0.5\*1+0\*1)/（1+1）=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

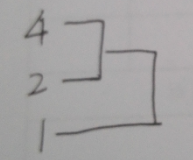
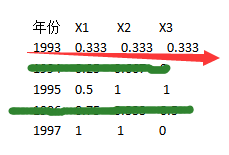
④求新群X42,1的变量值，第2行即X4,2群的变量值，而第1行现在只有它自己的样本，即只有1个。即Ni=2,Nj=1.代入公式后，把第2行去掉。新群X42,1的样本数为3（2+1）。

各个变量值为：

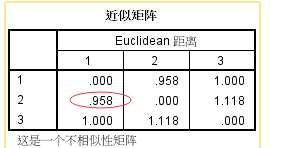
（0.5\*2+0\*1）/(2+1)=0.333

(0.5\*2+0\*1)/(2+1)=0.333

(0.25\*2+0.5\*1)/(2+1)=0.333



⑤对上述的数据进行求欧式距离：



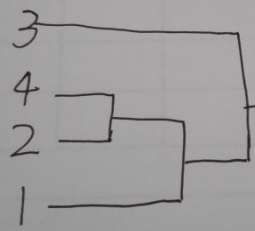
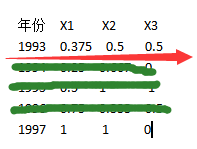
0.958，即原来的第三行第1列的数据最小，对3和1进行合并，距离系数为0.958。

⑥求新群X421,3的变量值,Ni=3,Nj=1，再把第3行去掉。

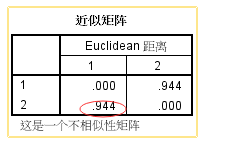
(0.333\*3+0.5\*1)/(3+1)=0.375

(0.333\*3+1\*1)/(3+1)=0.5

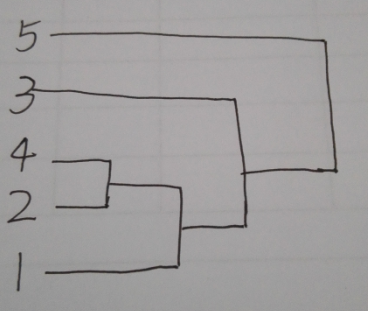
(0.333\*3+1\*1)/(3+1)=0.5



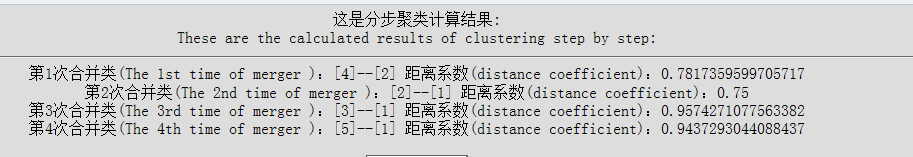
⑦对上述的数据再进行欧式距离的计算，得到：



即5和1合并，最后变成了一个群，结束。



与老师的结果一致：

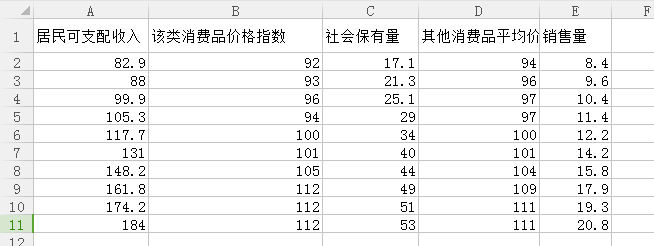




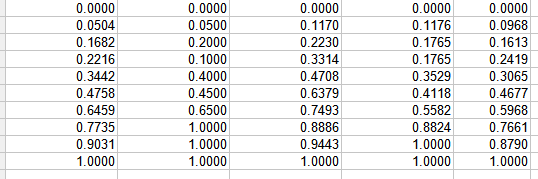
老师的图出bug了。

【聚类分析.xls的分析过程】

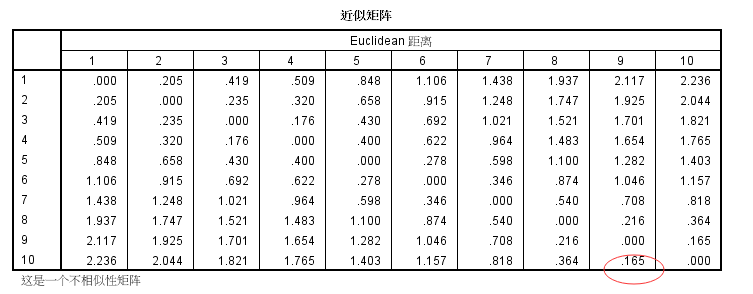
原始数据：



经过极差规格化之后：



①第一次欧式距离计算：



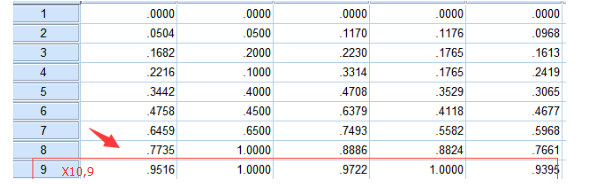
1. 第一次合并：10和9合并起来，距离系数为0.165。



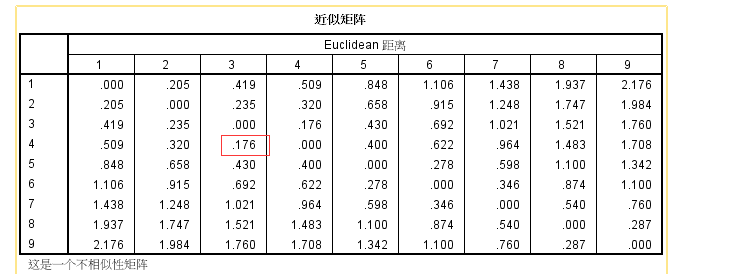
1. 求新群X10,9的变量值，=1，=1，算完新群X10,9的变量值后，把第10行去掉，新群的变量值如下：

（1\*1+0.9031\*1）/（1+1） = 0.9516

（1\*1+1\*1）/(1+1)=1……



②第二次欧式距离计算



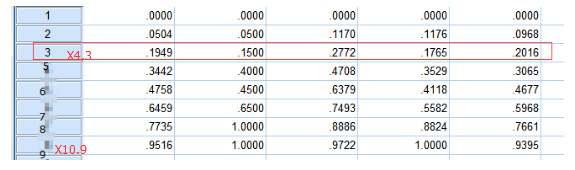
（1）第二次合并，把第4行和第3行合并，距离系数为0.176。



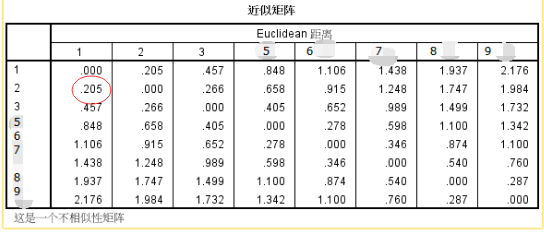
1. 求新群X4,3的变量值，N4=1，N3=1，求完之后把第4行去掉，新群X4,3的变量值如下：

（0.2216\*1+0.1682\*1）/（1+1） = 0.1949

（1\*1+2\*1）/（1+1）=1.5……



③第三次欧式距离：



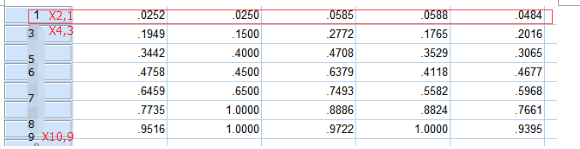
1. 第2行第1列的数据最小，2和1合并，距离系数为0.205



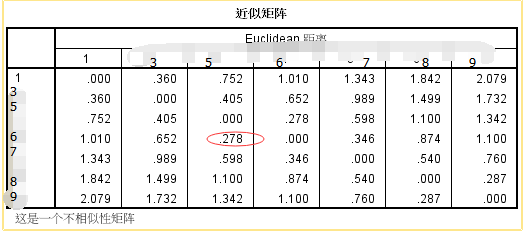
1. 求新群X2,1的变量值，N2=1，N1=1，求完之后把第2行去掉，新群X2,1的变量值如下：

（0\*1+0.504\*1）/（1+1） = 0.0252

（0.05\*1+0\*1）/（1+1）=0.25……



④第四次欧式距离：



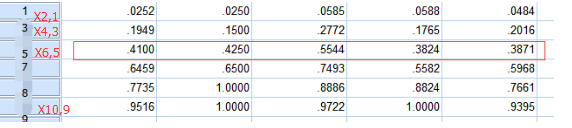
1. 发现第6行第5列的数据最小，距离系数为0.278，6和5合并：



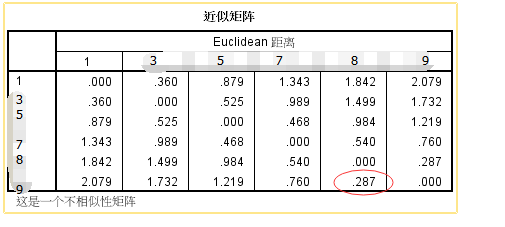
1. 求新群X6,5的变量值，N6=1，N5=1，求完之后把第6行去掉，新群X6,5的变量值如下：

（0.4758\*1+0.3442\*1）/(1+1)= 0.4100

（0.45\*1+0.4\*1）/(1+1)=0.425……



⑤第5次欧式距离：



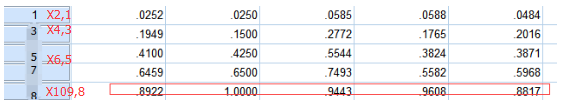
1. 第9行第8列的数据0.287最小，因此9和8合并，距离系数为0.287。



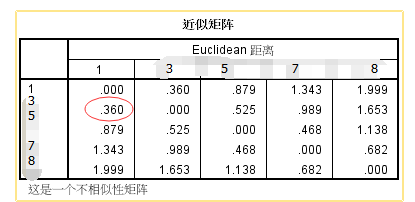
（2）第9行即之前求得群X10,9。它的样本数为N10,9=2，第8行的样本数N8=1，代入公式求得新群X109,8的各个变量值后，再把第9行去掉。新群X109,8的各个变量值为：

（0.9516\*2+0.7735\*1）/（2+1） = 0.8922

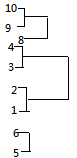
（1\*2+1\*1）/(2+1)=1……



⑥第6次欧式距离



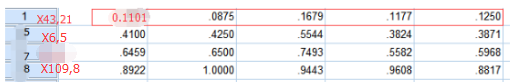
1. 发现第3行第1列的数据最小，所以3和1合并，距离系数为0.360



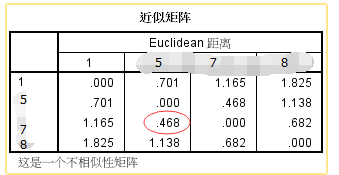
1. 第3行即X4,3群，第1行即X3,1群，N4,3=2, N3,1=2，代入公式，求新群X43,21的变量值，并把第3行去掉。

（0.1949\*2+0.0252\*2）/(2+2)=0.1101

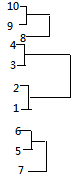
（0.1500\*2+0.025\*2）/(2+2)=0.0875……



⑦第7次欧式距离：



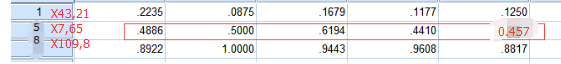
1. 发现地7行第5列的数据最小，距离系数为0.468，将7和5合并。



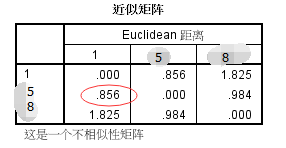
1. 第5行即X6,5群，第7行即原样本的单独还未合并的数据，N6,5=2, N7=1，代入公式，求新群X7,65的变量值，并把第7行去掉

（0.6459\*1+0.4100\*2）/(1+2) = 0.4886

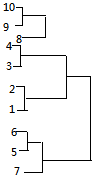
(0.6500\*1+0.4250\*2)/(1+2)=0.5 ……



⑧第8次欧式距离：



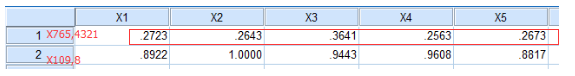
1. 发现第5行第1列的数据最小。将5和1合并，距离系数为0.856。



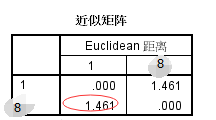
1. 第5行即X7,65群，第1行即X43,21群，N7,65=3，N43,21=4，代入公式，求新群X765,4321的变量值，并把第5行去掉。

（0.4886\*3+0.1101\*4）/(3+4)=0.2723

(0.5\*3+0.0875\*4)/(3+4)=0.2643……

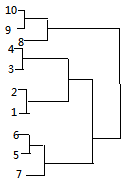


⑨第9次欧式距离：



1和8合并，距离系数为1.461。

至此，全部数据合并成了1个群X109,7654321,分析结束。



老师的结果基本一致（胜利）：

