# 题目一.聚类、主成分分析

1. 聚类分析

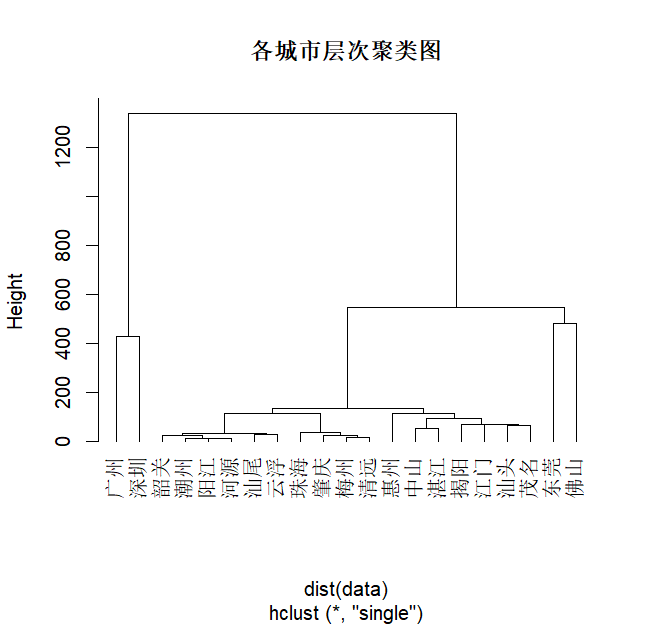
导入数据：

data <- read.csv("D://数据.csv", row=1)

1. 最短距离法

hclust\_result <- hclust(dist(data),method="single")

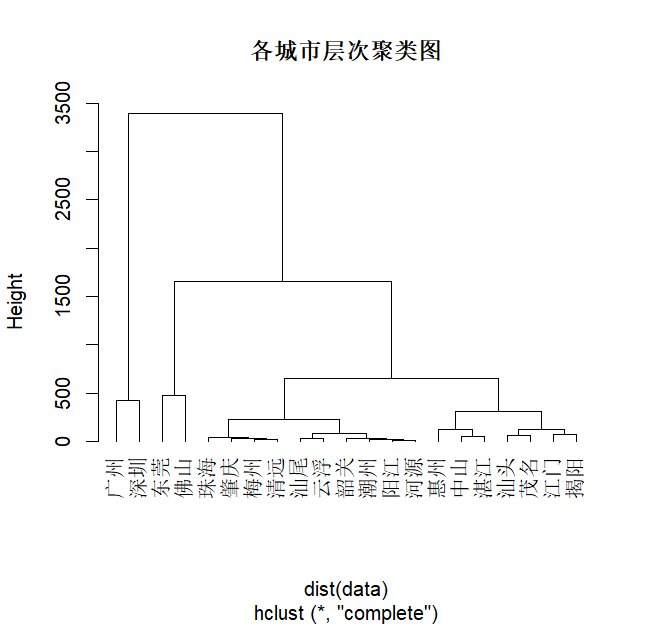
图像：



1. 最长距离法

hclust\_result <- hclust(dist(data),method="complete")

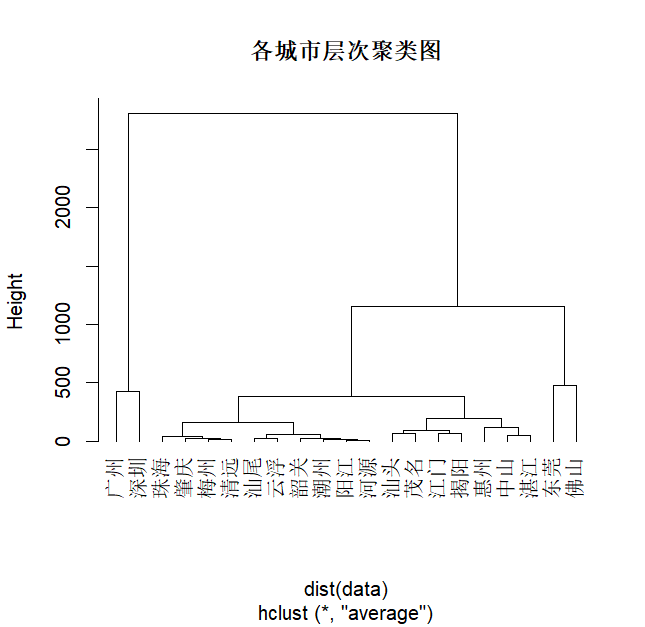
图像：



1. 平均距离法

hclust\_result <- hclust(dist(data),method="average")

图像：

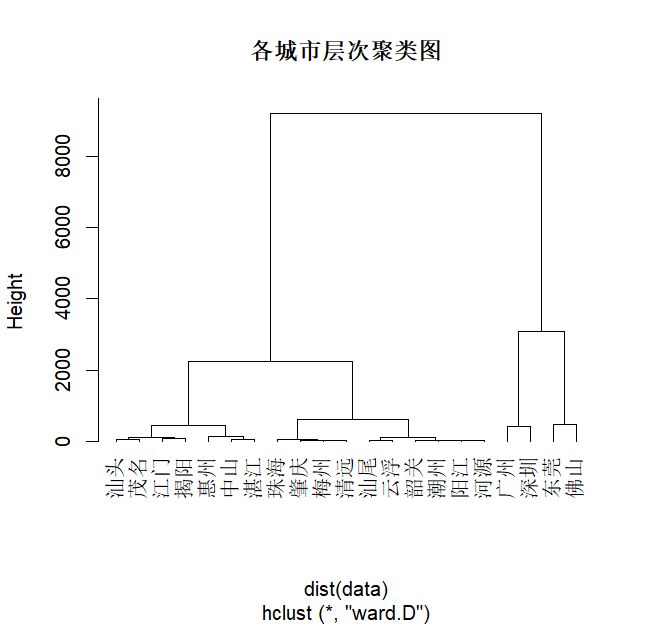


平均距离法

1. wald.D方法

hclust\_result <- hclust(dist(data),method="ward.D")

图像：



1. Kmeans聚类

从2：10选取最佳中心数的模型：

for (k in 1:10) {

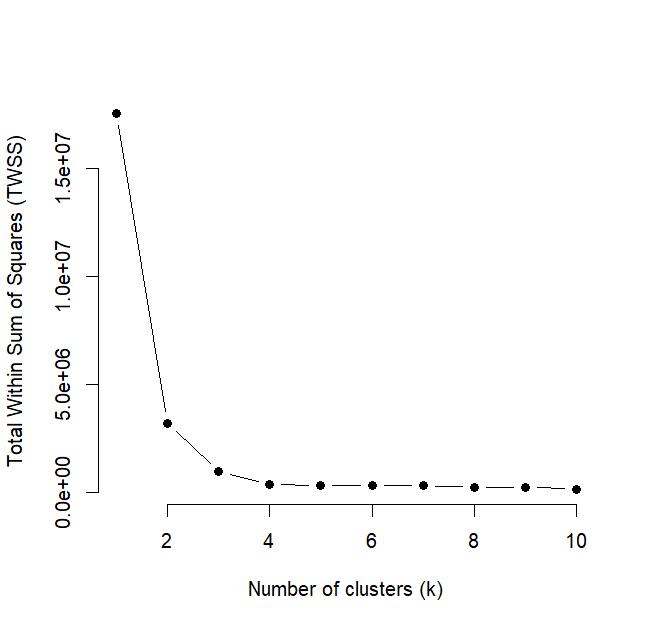
kmeans\_result <- kmeans(data, centers = k)

twss[k] <- kmeans\_result$tot.withinss

}

绘制肘部图：

plot(1:10, twss, type = "b", pch = 19, frame = FALSE, xlab = "Number of clusters (k)", ylab = "Total Within Sum of Squares (TWSS)")



肘部法图形

选取中心数为2的聚类结果：

kmeans\_result <- kmeans(data, centers = 2)

# 输出kmeans聚类结果

cluster\_labels <- kmeans\_result$cluster

cluster\_labels

> cluster\_labels

广州 深圳 珠海 惠州 东莞 中山 江门 佛山 肇庆 汕头 汕尾 潮州

2 2 1 1 2 1 1 1 1 1 1 1

揭阳 阳江 湛江 茂名 韶关 河源 梅州 清远 云浮

1 1 1 1 1 1 1 1 1

从上述聚类结果综合分析，可以看出分成2类较为合适。

|  |  |
| --- | --- |
| 第一类 | 第二类 |
| 广州、深圳 | 珠海、惠州、东莞、中山、江门、佛山、肇庆、汕头、汕尾、潮州、揭阳、阳江、湛江 茂名、韶关、河源、梅州、清远、云浮 |

二. 主成分分析

代码：

# 进行主成分分析

PCA=princomp(data,cor=T)

summary(PCA)

screeplot(PCA,type="lines")

#综合得分(m=1直接取分数)

PCA\_score\_abs <- abs(PCA$scores[,1])

sorted\_pca\_score <- sort(-PCA\_score\_abs)

rank(sorted\_pca\_score)

结果如下：

> summary(PCA)

Importance of components:

Comp.1 Comp.2 Comp.3

Standard deviation 2.6309808 0.212191743 0.163997917

Proportion of Variance 0.9888657 0.006432191 0.003842188

Cumulative Proportion 0.9888657 0.995297925 0.999140113

Comp.4 Comp.5 Comp.6

Standard deviation 0.063919786 0.0384430944 2.013499e-02

Proportion of Variance 0.000583677 0.0002111245 5.791681e-05

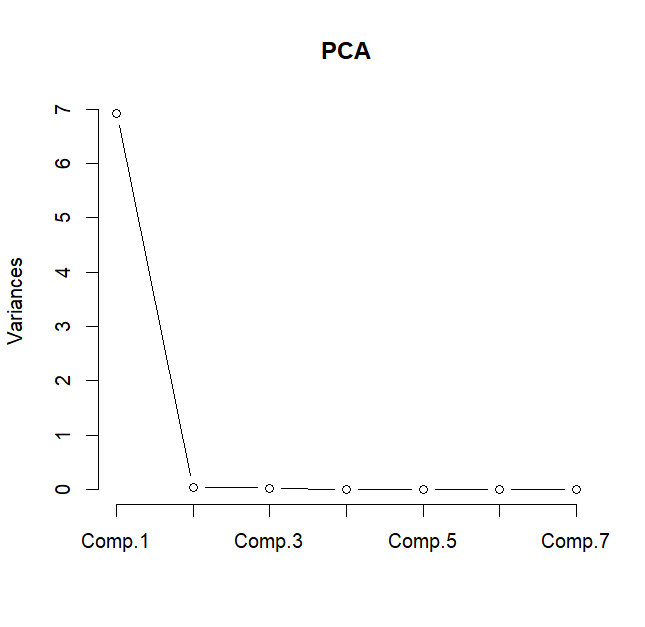
Cumulative Proportion 0.999723790 0.9999349148 9.999928e-01

Comp.7

Standard deviation 7.083678e-03

Proportion of Variance 7.168357e-06

Cumulative Proportion 1.000000e+00



> rank(sorted\_pca\_score)

广州 深圳 东莞 云浮 汕尾 佛山 河源 阳江 潮州 韶关 清远 梅州

1 2 3 4 5 6 7 8 9 10 11 12

肇庆 珠海 揭阳 茂名 汕头 江门 湛江 惠州 中山

13 14 15 16 17 18 19 20 21

从上述的分析可以看出Comp.1即电信业务总量(亿元)主成分占比最高，达到约了0.99，因此主成分只取一个。然后根据该主成分对每个城市进行排名，排名结果如上，广州为第一，中山为最后。

# 题目二.因子分析

1. 计算相关矩阵

#样本相关矩阵

R<-cor(data)

R

结果：

> R

X1 X2 X3 X4 X5

X1 1.0000000 0.76328027 0.84314861 0.96904443 0.1712224

X2 0.7632803 1.00000000 0.50350319 0.69238035 0.1398949

X3 0.8431486 0.50350319 1.00000000 0.87046843 0.1723038

X4 0.9690444 0.69238035 0.87046843 1.00000000 0.1795710

X5 0.1712224 0.13989493 0.17230376 0.17957098 1.0000000

X6 0.1268784 0.12880930 0.12298886 0.12746520 0.9846422

X7 0.2353095 0.07428481 0.26387630 0.26255241 0.4302291

X8 0.2341116 0.07247893 0.26337951 0.26294443 0.4349874

X9 0.1236413 0.14279039 0.07546233 0.06790907 -0.3029802

X10 0.3647862 0.24018246 0.33276188 0.26003630 0.1014948

X11 0.4552973 0.27500820 0.39702342 0.37094042 -0.1117388

X6 X7 X8 X9 X10

X1 0.1268784 0.23530947 0.23411156 0.12364128 0.3647862

X2 0.1288093 0.07428481 0.07247893 0.14279039 0.2401825

X3 0.1229889 0.26387630 0.26337951 0.07546233 0.3327619

X4 0.1274652 0.26255241 0.26294443 0.06790907 0.2600363

X5 0.9846422 0.43022906 0.43498740 -0.30298023 0.1014948

X6 1.0000000 0.44756724 0.45311704 -0.22674908 0.1074502

X7 0.4475672 1.00000000 0.99973452 -0.30353826 -0.1224743

X8 0.4531170 0.99973452 1.00000000 -0.30077878 -0.1258156

X9 -0.2267491 -0.30353826 -0.30077878 1.00000000 0.3618608

X10 0.1074502 -0.12247434 -0.12581556 0.36186081 1.0000000

X11 -0.1269058 -0.16108566 -0.15946369 0.49528784 0.8139465

X11

X1 0.4552973

X2 0.2750082

X3 0.3970234

X4 0.3709404

X5 -0.1117388

X6 -0.1269058

X7 -0.1610857

X8 -0.1594637

X9 0.4952878

X10 0.8139465

X11 1.0000000

每个变量间没有存在很强的相关性，故进行下一步因子分析。

二． 因子分析

使用"data"数据框进行分析。nfactors = 4参数指定了提取的因子数量为4个。rotate = "none"参数表示不进行因子旋转。

# 因子分析

pca\_result <- pca(data, nfactors = 4, rotate = "none")

# 旋转前的方差贡献率

pca\_result$Vaccounted

结果：

> pca\_result$Vaccounted

PC1 PC2 PC3 PC4

SS loadings 4.0886919 3.1024129 1.4357211 1.0182711

Proportion Var 0.3716993 0.2820375 0.1305201 0.0925701

Cumulative Var 0.3716993 0.6537368 0.7842569 0.8768270

Proportion Explained 0.4239140 0.3216570 0.1488550 0.1055740

Cumulative Proportion 0.4239140 0.7455710 0.8944260 1.0000000

可以看到因子数量选取4的时候，累计方差贡献率达到了0.87。

计算载荷矩阵的代码：

# 计算旋转前的因子载荷矩阵

factor\_loadings\_before\_rotation <- pca\_result$loadings

结果：

> print(factor\_loadings\_before\_rotation)

Loadings:

PC1 PC2 PC3 PC4

X1 0.927 -0.211 -0.229 -0.116

X2 0.710 -0.210 -0.190 -0.363

X3 0.852 -0.145 -0.216

X4 0.905 -0.146 -0.303 -0.128

X5 0.425 0.653 0.535 -0.302

X6 0.396 0.656 0.570 -0.253

X7 0.448 0.720 -0.138 0.497

X8 0.448 0.722 -0.134 0.495

X9 -0.645 0.237 0.260

X10 0.459 -0.485 0.602 0.210

X11 0.474 -0.650 0.377 0.312

PC1 PC2 PC3 PC4

SS loadings 4.089 3.102 1.436 1.018

Proportion Var 0.372 0.282 0.131 0.093

Cumulative Var 0.372 0.654 0.784 0.877

上述是旋转前的方差累计贡献率和载荷矩阵。

后计算选择后的方差累计贡献率和载荷矩阵：

# 因子旋转

rotated\_result <- fa(data, nfactors = 4, rotate = "varimax") # 使用varimax方法进行因子旋转

# 计算旋转后的方差贡献率

rotated\_result$Vaccounted

# 计算旋转后的因子载荷矩阵

factor\_loadings\_after\_rotation <- rotated\_result$loadings

输出结果：

Loadings:

MR1 MR3 MR4 MR2

X1 0.976 0.113 0.210

X2 0.698 0.114

X3 0.791 0.183 0.206

X4 0.983 0.136 0.104

X5 0.118 0.972 0.201

X6 0.952 0.227

X7 0.147 0.224 0.950 -0.146

X8 0.146 0.229 0.949 -0.146

X9 -0.211 -0.212 0.461

X10 0.197 0.147 0.807

X11 0.271 0.956

MR1 MR3 MR4 MR2

SS loadings 3.210 2.042 2.015 1.935

Proportion Var 0.292 0.186 0.183 0.176

Cumulative Var 0.292 0.477 0.661 0.836

从旋转前的因子方差贡献率来看，已经占到累计方差贡献率的87.7%，故选取前4个因子能较好地概况原始指标。

从旋转后的因子载荷矩阵可以看出，因子MR1在X1、X2、X3、X4上载荷量大，该因子得分高意味着公司能给予股东的回报一般高。因子MR3在X5、X6上载荷量大，该因子得分高意味着公司资产管理能力高；因子MR4在X7、X8上载荷量大，该因子得分高意味着公司短期偿债能力强；因子MR3在X9、X10、X11上载荷量大，该因子得分高意味着公司盈利能力较高。

# 计算各个公司的因子和综合因子得分

rotated\_result[["scores"]]

a<- rotated\_result[["Vaccounted"]][2,1:4]

a1<- t(matrix(rep(a, times=28),nrow=28, ncol=4,byrow=TRUE))

b<- rotated\_result[["scores"]]

scores<-(b %\*% a1/sum(a1[1:4,1]))[,1]

rank(-scores)

各个因子相对应的公司得分：

> rotated\_result[["scores"]]

MR1 MR3 MR4 MR2

上海医药 -0.39687230 0.122346698 -1.2651902 3.35905236

昆明制药 1.78446442 -0.433468371 0.8554822 0.13046820

片仔癀 0.44734574 4.500793584 0.6693499 -0.18612469

同仁堂 1.79662022 0.032991599 0.3437330 -0.17225172

天士力 0.65484271 -0.525502837 2.1601393 0.53363000

复星实业 2.09669051 -0.482602739 -1.2185541 -0.88966510

康美药业 2.06890068 -0.760613120 -0.9601697 0.08607700

江中药业 -0.27202113 -0.646436404 1.7692534 1.87515019

联环药业 -0.34536253 1.559002149 -0.1955916 0.40215474

交大昂立 -0.20648324 -0.118643781 0.5991981 0.42369976

双鹤药业 0.12891349 -0.489553724 -0.6395093 0.97204051

亚宝药业 -0.25212731 -0.418180114 0.5724360 0.48422908

东盛科技 -0.24700819 -0.804770806 2.3215933 -0.64697321

金宇集团 0.73347323 0.032019295 -0.1929223 -1.58173638

太极集团 -0.43948083 -0.542046587 -0.4341617 1.43178165

美罗药业 -0.72863824 0.116008314 -1.0159722 0.71459515

天药股份 0.57469813 -0.325571279 -1.0374286 -0.30838291

中新药业 -0.45388242 -0.521172207 0.3655530 0.03126206

星湖科技 -0.05232758 0.293619086 -0.4489819 -0.04373256

天坛生物 -0.18784791 0.279431089 0.5535012 -0.79738890

钱江生化 -0.17588641 0.026968318 -1.0703044 -0.50229853

迪康药业 -0.78899682 0.290179185 0.3657484 -0.94039494

金花股份 -0.39724958 0.232038273 -1.4296542 -1.03948870

鲁抗医药 -0.27832160 -0.338664186 -0.8324306 -0.46063885

通化东宝 -0.52661034 -0.001977229 0.5141827 -1.29904343

天目药业 -1.34658095 0.166884037 0.1261020 0.41127012

ST三普 -0.53354641 -0.533885152 -0.1646648 -0.89009862

ST金泰 -2.65670533 -0.709193092 -0.3107369 -1.09719228

计算综合因子得分，结果：

> rank(-scores)

上海医药 昆明制药 片仔癀 同仁堂 天士力 复星实业

7 2 1 4 3 9

康美药业 江中药业 联环药业 交大昂立 双鹤药业 亚宝药业

6 5 8 10 13 12

东盛科技 金宇集团 太极集团 美罗药业 天药股份 中新药业

11 17 16 20 18 19

星湖科技 天坛生物 钱江生化 迪康药业 金花股份 鲁抗医药

15 14 24 22 27 25

通化东宝 天目药业 ST三普 ST金泰

23 21 26 28

根据排名，可以看出，在偿债能力上，片仔癀排名第一，可能是与该公司独家生产和历史悠久有关。康美药业在股东回报上领先，但其他三个方面都不太好。天力士在盈利方面突出。同仁堂综合排名第4，在盈利能力因子上较低，但其它3个数据较好。

## 三、 小结

根据以上结果，可以认为上市公司业绩进行综合评定时可以主要考察该公司对股东的回报能力、盈利能力、偿债能力和资产管理能力等方面。而且对股东的回报能力主要考察每股收益、每股净产值、净产值收益率和扣除后每股收益。

从评价上市公司的四项主因子来看，对股东的回报放在首位，其次是资产管理、偿债、盈利能力。这一结论基本符合现代企业经营理论。