《多元统计分析》第七次上机作业
因子分析

3180103000 许乐乐

- 实验目的与要求：

  通过本试验项目，使学生理解并掌握如下内容：

  1. 熟悉潜在因子模型载荷矩阵的不同估计方法；
  2. 熟悉潜在因子个数的确定方法，因子得分的计算；
  3. 能够利用因子模型（或正交旋转）对所考虑问题做出合理的解释；

习题一

我国2010你那各地区城镇居民家庭平均每人全年消费数据如ex6.7所示，这些数据指标分别从食品（x1）,衣着（x2），居住(x3)，医疗(x4)，交通通信(x5)，教育(x6)，家政(x7)和耐用消费品(x8)来描述消费。试对该数据进行因子分析。

- 用`factanal()`进行因子分析。

  `factanal(x, factors, scores=(“none”, “regression”, “Bartlett”), rotation=”varimax”)`

  - `x` 为数值矩阵或数据表单
  - `factors` 为因子个数
  - `scores` 为因子得分的计算方法，包括`“regression”`,`“Barlett”`
  - `rotation` 为因子旋转方法，缺省为`"varimax"`，如果`rotation=”none”`则不做因子旋转

  注：该函数是基于极大似然方法来求解的。

```
d<-read_csv("ex6.7.csv")
factanal(d[,2:9],3,scores="Bartlett",rotation="varimax")
```

```
Call:
factanal(x = d[, 2:9], factors = 3, scores = "Bartlett", rotation = "varimax")

Uniquenesses:   # 是特殊方差
   x1    x2    x3    x4    x5    x6    x7    x8
0.108 0.426 0.005 0.200 0.041 0.253 0.108 0.292

Loadings:   # 是因子载荷矩阵
   Factor1 Factor2 Factor3
x1 0.923   0.153   0.127
x2 0.233   0.721
x3 0.538   0.337   0.769
x4         0.835   0.308
x5 0.916   0.246   0.246
x6 0.657   0.467   0.312
x7 0.849   0.272   0.312
x8 0.477   0.619   0.312

               Factor1 Factor2 Factor3
SS loadings      3.422   2.089   1.056  # 公共因子对变量的总方差贡献
```

```
Proportion Var   0.428   0.261   0.132  # 方差贡献率
Cumulative Var   0.428   0.689   0.821  # 累积方差贡献率

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 9.15 on 7 degrees of freedom.
The p-value is 0.242
```

取公共因子个数m=3，3个公共因子的累计贡献率在0.8以上（0.821），即说明前三个公共因子反映原始变量的信息已占总信息的80%以上。

- 方差最大正交旋转后的第一公共因子中，正载荷主要是食品(x1),交通通信(x5),家政(x7)，载荷值均在0.8以上，这些都属于日常支出，不妨称之为日常消费因子。
- 第二公共因子中，各变量的因子载荷的最大值为医疗(x4)，其载荷为0.835；衣着(x2),耐用消费品(x8)处在次要地位，这两者都属于季度支出，可以称为季度消费支出。
- 第三公共因子中，各变量的因子载荷的最大值为居住(x3)，其载荷为0.769，住房属于一次性大额支出，所以可以称为大额消费支出。

因此，影响消费的主要因素是食品(x1),居住(x3),医疗(x4)交通通信(x5),家政(x7)。

习题二

采用“体检数据”。这是一组4000多个样本的体检资料，分别有常规体检的一系列指标，

Sbp	Dbp	Sphygmus	Weight	Height	TC	TG	ALT	AST	T-BIL	IB	ALP	TP	Alb	GLB
收缩压	舒张压	脉搏	体重	身高	总胆固醇 (TC)	甘油三酯 (TG)	谷丙转氨酶 (ALT)	谷草转氨酶 (AST)	总胆红素 (T-BIL)	间接胆红素 (IB)	碱性磷酸酶 (ALP)	总蛋白 (TP)	白蛋白 (Alb)	球蛋白 (GLB)

其中，体检数据，请考虑下面的问题：

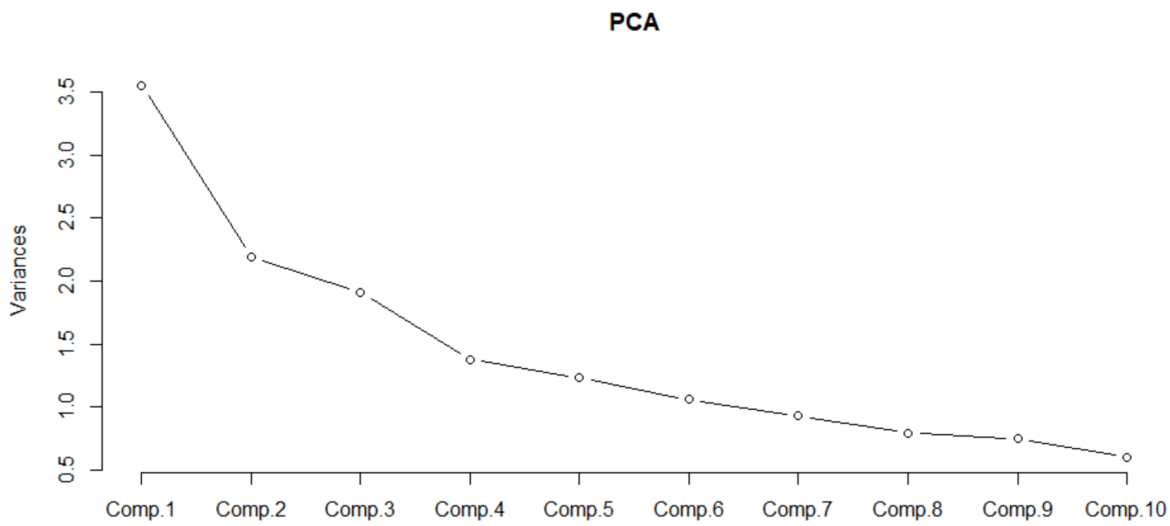
一、 利用主成分方法对变量进行降维，然后进行相应的主成分方法聚类分析；

```
d<-read_csv("hw8_2.csv")
d.a<-na.omit(d)
# 主成分分析
d.a<-d.na[,3:18]
# 计算主成分
PCA<-princomp(d.a, cor=T)
# 主成分的方差和累计占比，给出特征向量
summary(PCA,loadings=TRUE)
# 画碎石图
screeplot(PCA, type="lines")
# 主成分得分
PCAS<-PCA$scores
```

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4
Standard deviation     1.8852608 1.4820144 1.3810009 1.17354888
Proportion of Variance 0.2369472 0.1464244 0.1271442 0.09181447
Cumulative Proportion  0.2369472 0.3833717 0.5105159 0.60233036
                           Comp.5    Comp.6     Comp.7     Comp.8
Standard deviation     1.11047946 1.0288761 0.96309333 0.89071769
Proportion of Variance 0.08221098 0.0705724 0.06183658 0.05289187
Cumulative Proportion  0.68454133 0.7551137 0.81695032 0.86984219
                           Comp.9    Comp.10    Comp.11    Comp.12
Standard deviation     0.86399831 0.77506394 0.51872944 0.46045152
```

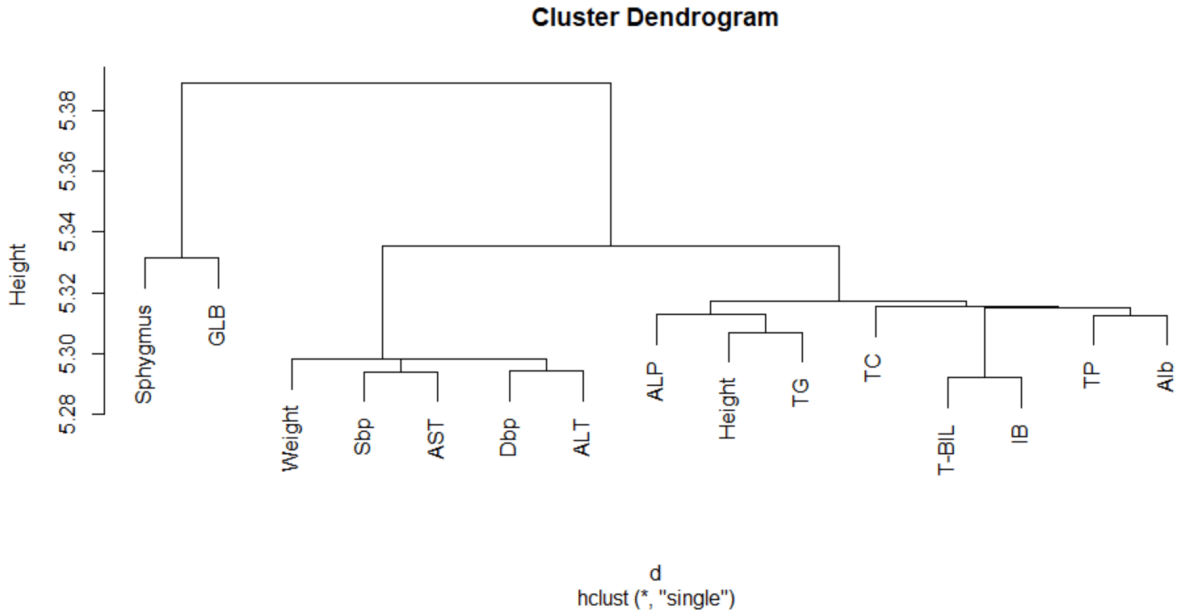
```
Proportion of Variance 0.04976621 0.04004827 0.01793868 0.01413437
Cumulative Proportion  0.91960839 0.95965667 0.97759535 0.99172972
                           Comp.13    Comp.14 Comp.15
Standard deviation     0.320359305 0.14636954       0
Proportion of Variance 0.006842006 0.00142827       0
Cumulative Proportion  0.998571730 1.00000000       1

Loadings:
          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
Sbp        0.335          0.109  0.428  0.345         0.189
Dbp        0.361          0.120  0.420  0.279                0.112
Sphygmus         -0.179 -0.103  0.180  0.474 -0.317 -0.426 -0.104
Weight     0.365  0.189  0.201  0.110 -0.313                0.235
Height     0.268  0.259         0.128 -0.479 -0.276         0.152
TC         0.206 -0.159                      0.539 -0.470 -0.259
TG         0.259 -0.197                -0.272  0.345 -0.259 -0.106
ALT        0.358          0.195 -0.500  0.154 -0.157
AST        0.330          0.171 -0.537  0.229
T-BIL      0.176  0.394 -0.490 -0.114  0.124  0.189
IB         0.184  0.383 -0.489 -0.105  0.112  0.224
ALP        0.225                             0.102  0.617 -0.626
TP         0.192 -0.465 -0.408        -0.157 -0.136  0.101  0.107
Alb        0.214        -0.351        -0.202 -0.473 -0.158 -0.434
GLB              -0.508 -0.251                0.179  0.242  0.453
          Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
Sbp        0.221                   0.687
Dbp        0.190  0.115  -0.198  -0.700
Sphygmus -0.625 -0.116
Weight   -0.185 -0.136   0.721          -0.202
Height   -0.247 -0.240  -0.608   0.129
TC         0.197 -0.547
TG       -0.260  0.732
ALT                                    0.711
AST        0.112         -0.204          -0.660
T-BIL                                            0.708
IB                                              -0.704
ALP      -0.344 -0.146
TP                                                      0.709
Alb        0.373  0.107                                -0.418
GLB      -0.161 -0.148                                -0.568
```



主成分个为7时，累计贡献率达到0.81，因此选取主成分个数为7。下面基于主成分进行聚类分析。

```
# 基于主成分进行聚类分析
loadingM<-PCA$loadings
d<-dist(scale(loadingM))
hc1<-hclust(d,method="single")
plot(hc1)
```



二、 构建因子分析模型，进行因子旋转，分析每个因子的意义及这些潜在的因子与年龄的关系。

```
d.scale<-scale(d.a)
factanal(d.scale,3,scores="Bartlett", rotation="varimax")
```

```
Call:
factanal(x = d.scale, factors = 3, scores = "Bartlett", rotation = "varimax")

Uniquenesses:
     Sbp      Dbp Sphygmus   Weight   Height       TC
   0.994    0.994    0.934    0.986    0.916    0.931
      TG      ALT      AST    T-BIL       IB      ALP
   0.909    0.910    0.953    0.027    0.016    0.937
      TP      Alb      GLB
   0.005    0.005    0.005

Loadings:
         Factor1 Factor2 Factor3
Sbp                      0.104
Dbp                      0.143
Sphygmus  0.153
Weight            0.112  0.155
Height            0.122  0.281
TC        0.161
TG        0.244
ALT                      0.122
AST       0.120
T-BIL             0.947  0.272
IB                0.957  0.261
ALP       0.129
```

```
TP        0.982           0.183
Alb       0.443           0.894
GLB       0.900          -0.429

                Factor1 Factor2 Factor3
SS loadings       2.137   1.871   1.314
Proportion Var    0.142   0.125   0.088
Cumulative Var    0.142   0.267   0.355

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 124644.8 on 63 degrees of freedom.
The p-value is 0
```

4以上的因子会报错。

```
Error in factanal(d.scale, 4, scores = "Bartlett", rotation = "varimax") :
  无法用这些初始值进行最佳化运算
```