# 多元统计分析期末作业

统计 1701 尹恒

2020/5/18

## 第一题

```
data1=read.table("T1.DAT")   # 读入数据
new_data1<-data1[,c(1:5)]    # 提取前五列作主成分分析
apply(new_data1,2,mean)    # 计算样本均值向量
```

```
##       V1       V2       V3       V4       V5
## 15.66923 17.07692 18.78462 15.50000 11.73077
```

**(a) 确定能有效的综合样本变异性的恰当的成分个数。构造崖底碎石图帮助你的求解。**

```
print(cov(new_data1),digits = 4) # 输出协方差矩阵
```

通过协方差阵 **S** 作主成分分析

```
##          V1      V2       V3      V4      V5
## V1   34.750 -4.2767 -18.0718 -15.973   5.716
## V2   -4.277 17.5134   0.4198  -7.868  -8.723
## V3  -18.072  0.4198  29.8447   9.349 -13.942
## V4  -15.973 -7.8682   9.3488  33.043  -9.942
## V5    5.716 -8.7233 -13.9422  -9.942  26.958
```

```
print(eigen(cov(new_data1))) # 输出特征值-特征向量
```

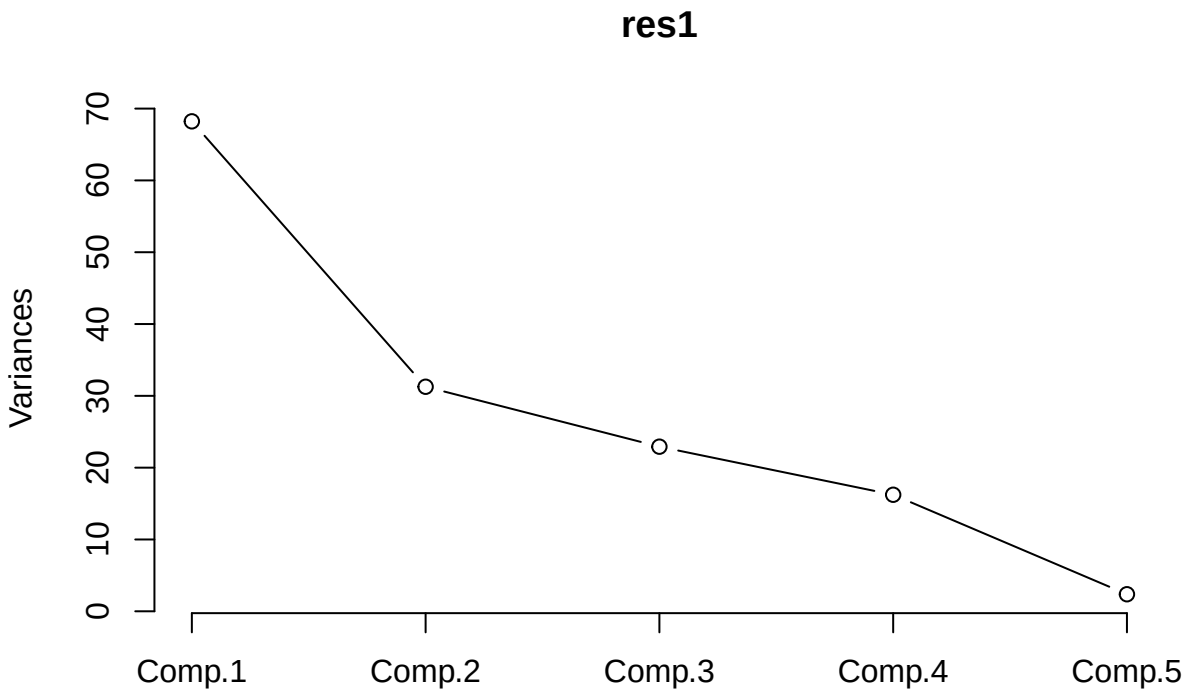
```
## eigen() decomposition
## $values
## [1] 68.752385 31.508994 23.100973 16.354182  2.392411
##
## $vectors
##             [,1]        [,2]       [,3]        [,4]       [,5]
## [1,]  0.57943538  0.07917988 -0.6428795 -0.30939267 -0.3859629
## [2,] -0.04165689  0.61192825  0.1399143  0.51462195 -0.5825777
```

```
## [3,] -0.52428496  0.21883511  0.1192554 -0.73403767 -0.3524249
## [4,] -0.49309245 -0.57215650 -0.4221873  0.30427403 -0.3983365
## [5,]  0.38013742 -0.49398633  0.6120997 -0.08970196 -0.4782893
```

```
res1=princomp(new_data1,cor=FALSE) # 当 cor=FALSE 表示用样本的协方差阵 S 做主成分分析
summary(res1,loadings=TRUE)
```

```
## Importance of components:
##                            Comp.1    Comp.2    Comp.3   Comp.4     Comp.5
## Standard deviation     8.2597530 5.5916560 4.7878255 4.028446 1.54078160
## Proportion of Variance 0.4838005 0.2217242 0.1625582 0.115082 0.01683505
## Cumulative Proportion  0.4838005 0.7055248 0.8680829 0.983165 1.00000000
##
## Loadings:
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## V1  0.579         0.643  0.309  0.386
## V2         0.612 -0.140 -0.515  0.583
## V3 -0.524  0.219 -0.119  0.734  0.352
## V4 -0.493 -0.572  0.422 -0.304  0.398
## V5  0.380 -0.494 -0.612         0.478
```

```
screeplot(res1,type="lines")# 输出崖底碎石图
```




由上述程序整理可得主成分的系数为:

| 变量 | $\hat{e}_1$ | $\hat{e}_2$ | $\hat{e}_3$ | $\hat{e}_4$ | $\hat{e}_5$ |
|---|---|---|---|---|---|
| 独立性 | 0.579 | - | 0.643 | 0.309 | 0.386 |
| 支持力 | - | 0.612 | -0.140 | -0.515 | 0.583 |
| 仁爱心 | -0.524 | 0.219 | -0.119 | 0.734 | 0.352 |
| 顺从性 | -0.493 | -0.572 | 0.422 | -0.304 | 0.398 |
| 领导能力 | 0.380 | -0.494 | -0.612 | - | 0.478 |
| 方差（$\hat{\lambda}$） | 68.752 | 31.509 | 23.101 | 16.354 | 2.392 |
| 占总方差的累计百分比 | 0.484 | 0.706 | 0.868 | 0.983 | 1.000 |

由碎石图可以看出选择前三个主成分较为合理，由表格得前三个主成分占总方差的累计百分比为 86.8%，与碎石图得到的结果基本一致。

```
print(cor(new_data1),digits = 4) # 输出相关矩阵
```

**通过相关阵 R 作主成分分析**

```
##          V1       V2       V3       V4       V5
```

```
## V1  1.0000 -0.17336 -0.56116 -0.4714  0.1868
## V2 -0.1734  1.00000  0.01836 -0.3271 -0.4015
## V3 -0.5612  0.01836  1.00000  0.2977 -0.4915
## V4 -0.4714 -0.32708  0.29771  1.0000 -0.3331
## V5  0.1868 -0.40147 -0.49153 -0.3331  1.0000
```

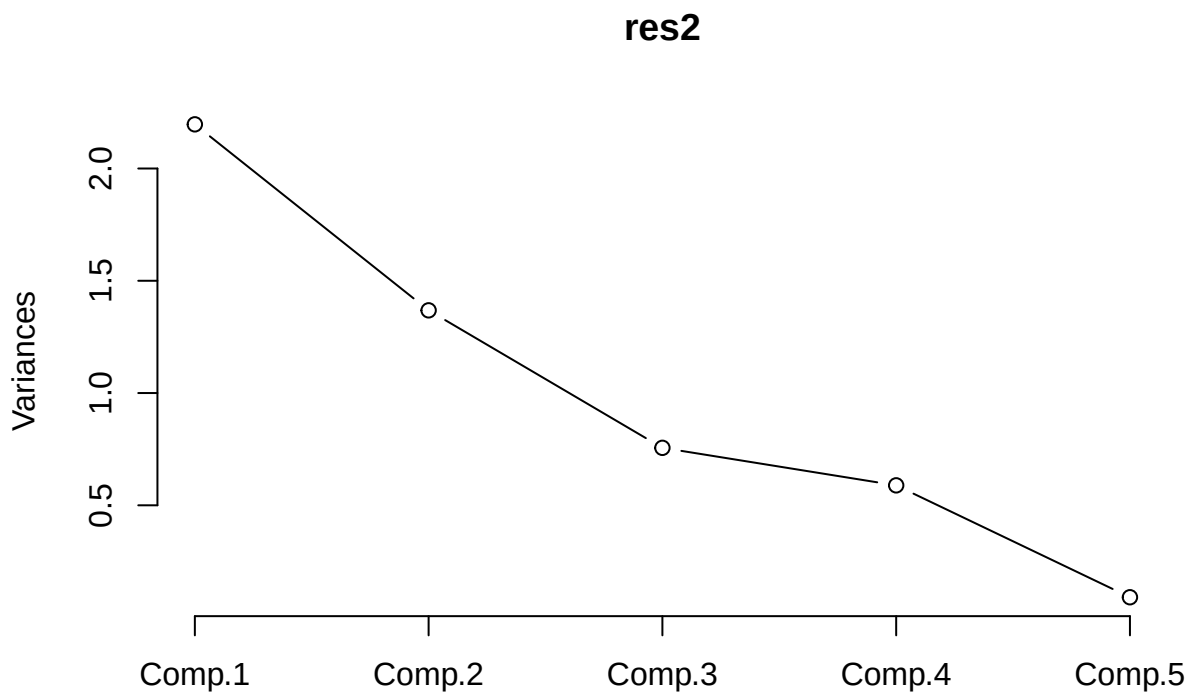
```
print(eigen(cor(new_data1))) # 输出特征值-特征向量
```

```
## eigen() decomposition
## $values
## [1] 2.19662443 1.36824960 0.75586304 0.58878599 0.09047694
##
## $vectors
##            [,1]         [,2]       [,3]         [,4]      [,5]
## [1,]  0.5209626  0.086521361  0.6674512  0.253099293 0.4599582
## [2,] -0.1213677  0.788216689 -0.1870605 -0.350892684 0.4537257
## [3,] -0.5482732 -0.007941356 -0.1150943  0.732694760 0.3863226
## [4,] -0.4391410 -0.490952547  0.2949415 -0.525281896 0.4507873
## [5,]  0.4694885 -0.360736798 -0.6475184 -0.007238184 0.4797052
```

```
res2=princomp(new_data1,cor=TRUE) # 当 cor=TRUE 表示用样本的相关矩阵 R 做主成分分析
summary(res2,loadings=TRUE)
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3    Comp.4     Comp.5
## Standard deviation     1.4821014 1.1697220 0.8694038 0.7673239 0.30079386
## Proportion of Variance 0.4393249 0.2736499 0.1511726 0.1177572 0.01809539
## Cumulative Proportion  0.4393249 0.7129748 0.8641474 0.9819046 1.00000000
##
## Loadings:
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## V1  0.521         0.667  0.253  0.460
## V2 -0.121  0.788 -0.187 -0.351  0.454
## V3 -0.548        -0.115  0.733  0.386
## V4 -0.439 -0.491  0.295 -0.525  0.451
## V5  0.469 -0.361 -0.648         0.480
```

```
screeplot(res2,type="lines") # 输出崖底碎石图
```




由上述程序整理可得主成分的系数为:

| 变量 | $\hat{e}_1$ | $\hat{e}_2$ | $\hat{e}_3$ | $\hat{e}_4$ | $\hat{e}_5$ |
|---|---|---|---|---|---|
| 独立性 | 0.521 | - | 0.667 | 0.253 | 0.460 |
| 支持力 | -0.121 | 0.788 | -0.187 | -0.351 | 0.454 |
| 仁爱心 | -0.548 | - | -0.115 | 0.733 | 0.386 |
| 顺从性 | -0.439 | -0.491 | 0.295 | -0.525 | 0.451 |
| 领导能力 | 0.469 | -0.361 | -0.648 | - | 0.480 |
| 方差（$\hat{\lambda}$） | 2.197 | 1.368 | 0.756 | 0.589 | 0.090 |
| 占总方差的累计百分比 | 0.439 | 0.713 | 0.864 | 0.981 | 1.000 |

由碎石图可以看出选择前三个主成分较为合理，由表格得前三个主成分占总方差的累计百分比为 86.4%，与碎石图得到的结果基本一致。

**(b) 解释样本主成分.**

选取的三个样本主成分为：

$$y_1 = 0.579 \times x_1 - 0.524 \times x_3 - 0.493 \times x_4 + 0.380 \times x_5$$

$$y_2 = 0.612 \times x_2 + 0.219 \times x_3 - 0.572 \times x_4 - 0.494 \times x_5$$

$$y_3 = 0.643 \times x_1 - 0.140 \times x_2 - 0.119 \times x_3 + 0.422 \times x_4 - 0.612 \times x_5$$

第一主成分占总方差的 48.4%，其中和独立性与领导能力正相关，和仁爱心与顺从性负相关，表现出秘鲁青年心理大部分呈现独立自主。

第二主成分占总方差的 22.2%，其中和支持力与仁爱心正相关，和领导能力与顺从性负相关，表现出秘鲁青年心理少部分呈现顺从仁爱。

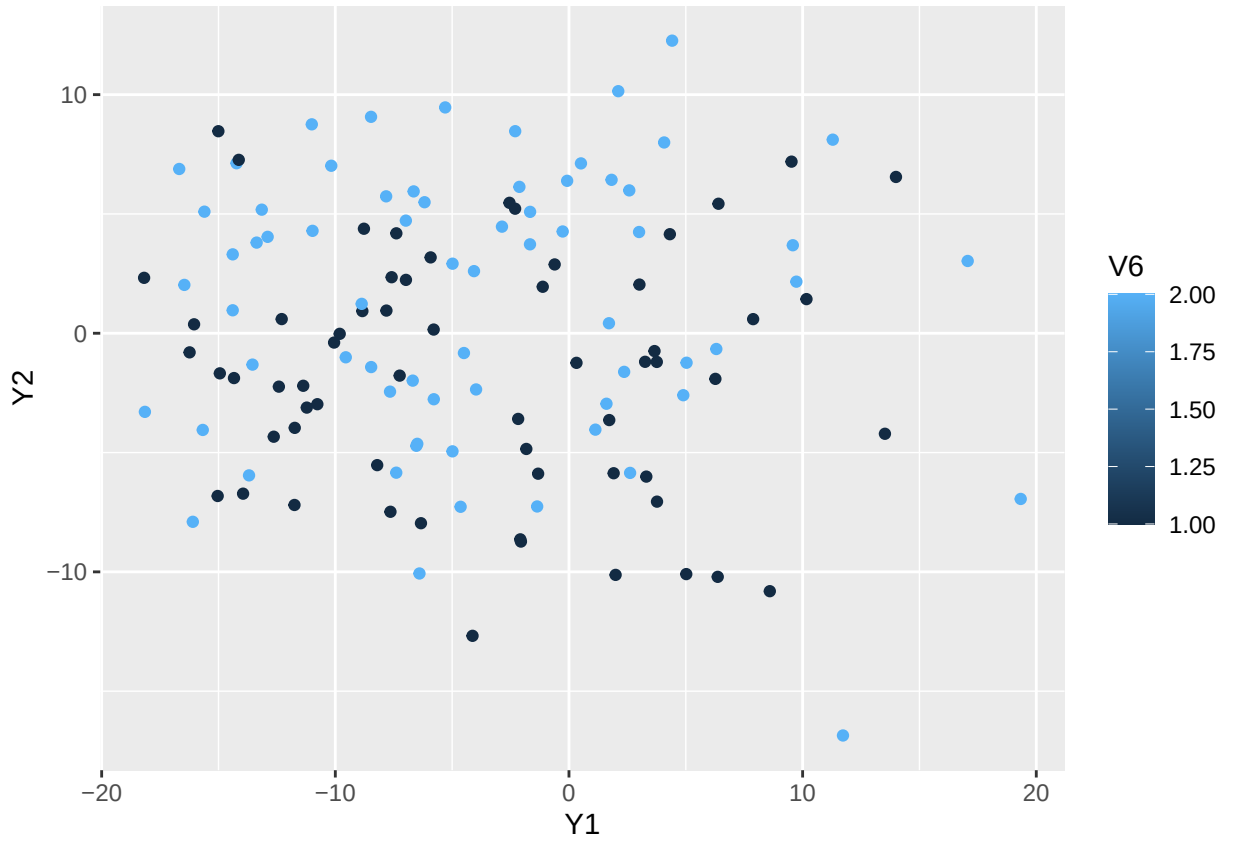
第三主成分占总方差的 16.3%，是五个因素的线性组合，表现更一般的情况，各种因素都会影响秘鲁青年心理状况。

**(c) 用前两个主成分的值，将 $(\hat{y_1}, \hat{y_2})$ 的值画在图中**

```
library(ggplot2)
attach(data1)# 把第一二主成分的值加入原表格
data1$Y1<-0.579*V1-0.524*V3-0.493*V4+0.380*V5 # 第一主成分
data1$Y2<-0.612*V2+0.219*V3-0.572*V4-0.494*V5 # 第二主成分
detach(data1)
```

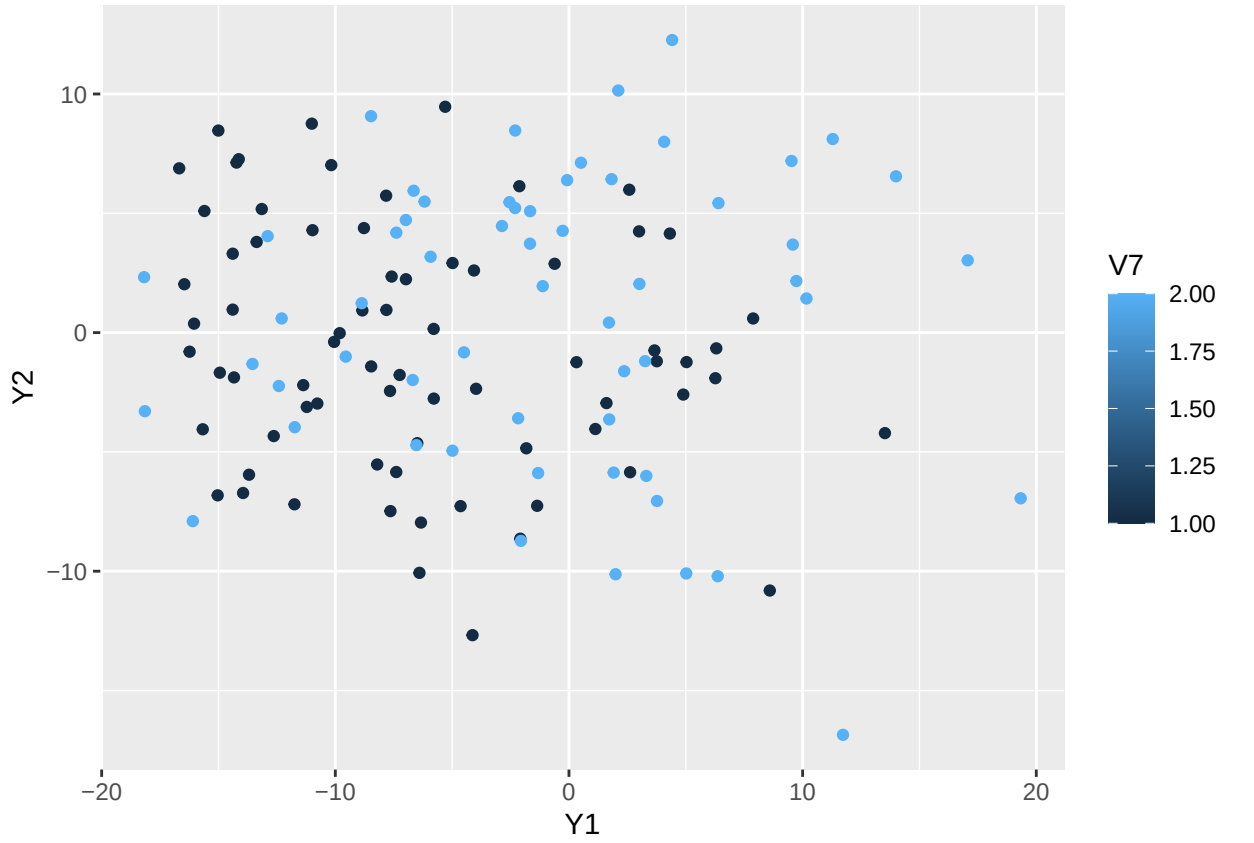
以性别作为分类指标作图

```
ggplot(data1,aes(x=Y1,y=Y2,colour=V6))+geom_point()
```



以地位作为分类指标作图

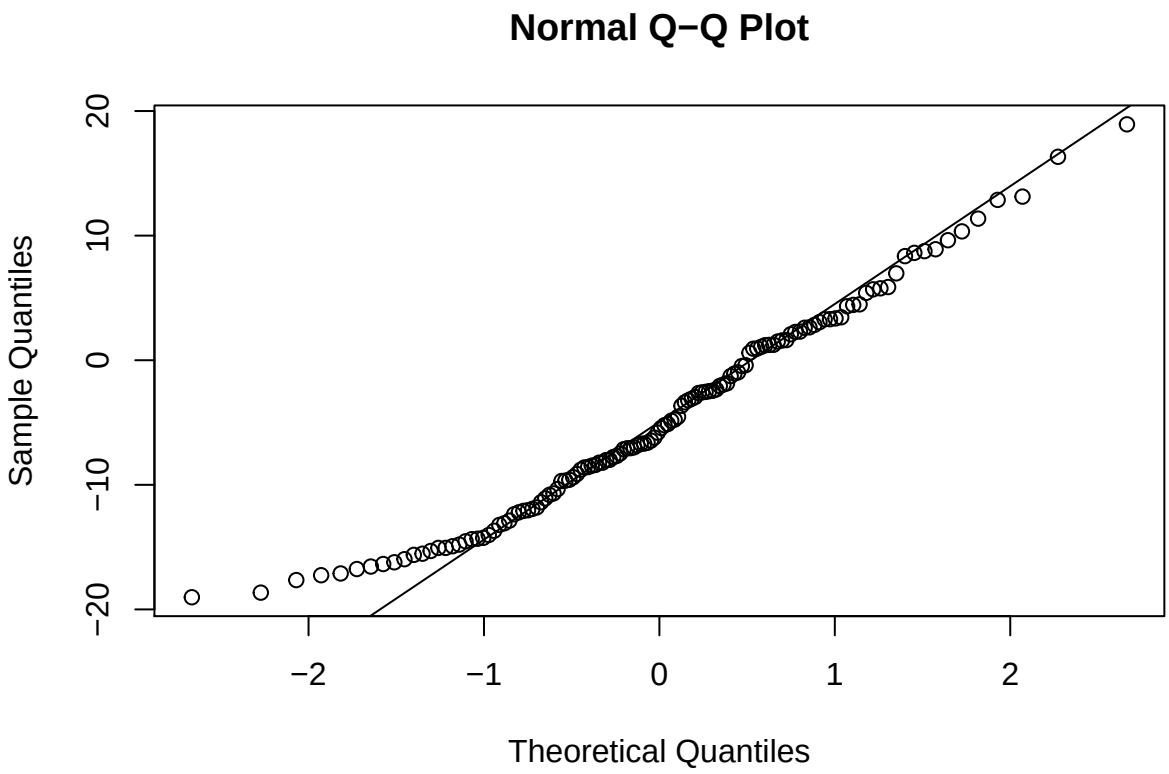
```
ggplot(data1,aes(x=Y1,y=Y2,colour=V7))+geom_point()
```




从图看出在右下角存在几个离群值。

**(d) 用第一主成分 Q-Q 图，解释该图。**

```
lamda<-eigen(cov(new_data1))
spc_mat<-lamda$vectors[,1]
prin_y<-t(t(spc_mat)%*%t(new_data1))
qqnorm(prin_y)
qqline(prin_y)
```




此图说明第一主成分在最左端拟合程度不好，并且在最右上角存在可疑点。

## 第二题

### (a) 求主成分解或极大似然解

因子分析，利用 psych 包中的 fa(r=cor2,nfactors=2,fm="pa",rotate="none") 函数，该函数为多元统计分析的一个包；nfactors 为因子个数,fm 为估计解的方法:pa 为主成分法,ml 为极大似然估计法;rotate 为是否进行旋转

```
library(psych)
```

```
data2=read.table("T2.DAT")
cor2<-cor(scale(data2))
# 主成分法
m2<-fa(r=cor2,nfactors=2,fm="pa",rotate="none")
```

```
## maximum iteration exceeded
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

```
m2$loadings
```

```
##
## Loadings:
##    PA1    PA2
## V1  0.984 -0.168
## V2  0.933 -0.117
## V3  0.934
## V4  0.717  0.870
## V5  0.722
## V6  0.579 -0.291
## V7  0.906 -0.273
##
##                  PA1   PA2
## SS loadings    4.901 0.965
## Proportion Var 0.700 0.138
## Cumulative Var 0.700 0.838
```

```
m3<-fa(r=cor2,nfactors=3,fm="pa",rotate="none")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## An ultra-Heywood case was detected. Examine the results carefully
```

```
m3$loadings
```

```
##
## Loadings:
##    PA1    PA2    PA3
## V1  0.976
## V2  0.951        -0.312
## V3  0.936         0.146
## V4  0.652  0.641  0.279
## V5  0.723  0.179
## V6  0.641 -0.573  0.390
## V7  0.917 -0.175 -0.274
##
```

```
##                  PA1   PA2   PA3
## SS loadings    4.934 0.813 0.424
## Proportion Var 0.705 0.116 0.061
## Cumulative Var 0.705 0.821 0.882
```

```
# 极大似然估计法
m22<-fa(r=cor2,nfactors=2,fm="ml",rotate="none")
m22$loadings
```

```
##
## Loadings:
##    ML1    ML2
## V1  0.695  0.669
## V2  0.669  0.695
## V3  0.795  0.494
## V4  0.983 -0.167
## V5  0.655  0.312
## V6  0.250  0.569
## V7  0.558  0.812
##
##                  ML1   ML2
## SS loadings    3.333 2.283
## Proportion Var 0.476 0.326
## Cumulative Var 0.476 0.802
```

```
m33<-fa(r=cor2,nfactors=3,fm="ml",rotate="none")
m33$loadings
```

```
##
## Loadings:
##    ML1    ML3    ML2
## V1  0.901  0.381
## V2  0.775  0.600
## V3  0.931  0.202
## V4  0.733 -0.118  0.666
## V5  0.689  0.225  0.169
## V6  0.757 -0.132 -0.636
## V7  0.762  0.608 -0.110
##
##                  ML1   ML3   ML2
## SS loadings    4.445 0.998 0.901
```

```
## Proportion Var 0.635 0.143 0.129
## Cumulative Var 0.635 0.778 0.906
```

**(b) 求旋转载荷，比较这两组旋转载荷，解释因子解**

```
m20<-fa(r=cor2,nfactors=2,fm="ml",rotate="varimax")
m20$loadings
```

```
##
## Loadings:
##    ML2   ML1
## V1 0.852 0.452
## V2 0.868 0.419
## V3 0.717 0.602
## V4 0.148 0.987
## V5 0.501 0.525
## V6 0.619
## V7 0.946 0.277
##
##                  ML2   ML1
## SS loadings    3.545 2.071
## Proportion Var 0.506 0.296
## Cumulative Var 0.506 0.802
```

第一因子可以把数学能力与销售利润联系起来，表现销售人员的销售能力第二因子可以把创造力和新客户销售额和联系起来，表现销售人员推销能力

```
m30<-fa(r=cor2,nfactors=3,fm="ml",rotate="varimax")
m30$loadings
```

```
##
## Loadings:
##    ML3   ML1   ML2
## V1 0.793 0.374 0.438
## V2 0.911 0.317 0.185
## V3 0.651 0.544 0.438
## V4 0.255 0.964
## V5 0.542 0.465 0.207
## V6 0.299       0.950
## V7 0.917 0.180 0.298
##
##                  ML3   ML1   ML2
```

```
## SS loadings    3.175 1.718 1.453
## Proportion Var 0.454 0.245 0.208
## Cumulative Var 0.454 0.699 0.906
```

第一，二因子同上。第三因子把抽象推理能力和新客户销售额和销售增长联系起来，表现销售人员的判断能力。比较这两组旋转载荷，三个因子的累计方差达到了 90%，比两个因子的累计方差高。

### (c) 列出共性方差，特殊方差，比较结果并解释

列出分析的完整数据如下:（communalities 为共性方差，特殊方差为 1-共性方差)

```
m20$communalities
```

```
##        V1        V2        V3        V4        V5        V6        V7
## 0.9308084 0.9296171 0.8766888 0.9950000 0.5264156 0.3863585 0.9711829
```

```
m30$communalities
```

```
##        V1        V2        V3        V4        V5        V6        V7
## 0.9614288 0.9655182 0.9118758 0.9950000 0.5533880 0.9950000 0.9624919
```

两个因子的共性方差和特殊方差为：

| | 共性方差 | 特殊方差 |
|---|---|---|
| 销售增长 | 0.93 | 0.069 |
| 销售利润 | 0.93 | 0.070 |
| 新客户销售额 | 0.88 | 0.123 |
| 创造力 | 1.00 | 0.005 |
| 机械推理 | 0.53 | 0.474 |
| 抽象推理 | 0.39 | 0.614 |
| 数学能力 | 0.97 | 0.029 |

三个因子的共性方差和特殊方差为：

| | 共性方差 | 特殊方差 |
|---|---|---|
| 销售增长 | 0.96 | 0.039 |
| 销售利润 | 0.97 | 0.034 |
| 新客户销售额 | 0.91 | 0.088 |
| 创造力 | 1.00 | 0.005 |
| 机械推理 | 0.55 | 0.447 |
| 抽象推理 | 1.00 | 0.005 |
| 数学能力 | 0.96 | 0.038 |

比较两个表格，三个因子的共性方差基本都接近 1，并且三个因子的累计方差高，所以选三个因子

**(d) 对 $m=2$ 和 $m=3$ 做假设检验。**

由公式（9-39）如下，把 n=50，p=7，m=2,3 代入得：

$$(n-1-(2p+4m+5)/6)\ln\frac{\left|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\Psi}\right|}{|\mathbf{S}_n|}>\chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

$$43.833\times ln(\frac{0.000075933}{0.000018427})=62.1>\chi^2(0.01)=11.3$$

所以我们拒绝原假设 $H_0$，综合以上分析选择 $m=3$。

## 第三题

**(a) 使用二次判别方法将 $X_0^{'}=[3.5,1.75]$ 分类到总体 $\pi_1,\pi_2\ \pi_3$**

```
library(MASS)
data3=read.table("T3.DAT")
qd<-qda(V5~V2+V4,data3,prior=c(1/3,1/3,1/3)) # 二次判别
```

```
predict(qd,newdata = data.frame(V2=3.5,V4=1.75))
```

```
## $class
## [1] 2
## Levels: 1 2 3
##
## $posterior
##              1         2         3
## 1 6.391308e-46 0.7807453 0.2192547
```

根据后验概率第二类最大，所以分类在第二类

**(b) 使用线性判别方法将 $X_0^{'}=[3.5,1.75]$ 分类到总体 $\pi_1,\pi_2\ \pi_3$**

```
ld<-lda(V5~V2+V4,data3,prior=c(1/3,1/3,1/3)) # 线性判别
```

```
predict(ld,newdata = data.frame(V2=3.5,V4=1.75))
```

```
## $class
## [1] 2
## Levels: 1 2 3
##
## $posterior
```

```
##               1         2         3
## 1 3.209389e-14 0.7187594 0.2812406
##
## $x
##        LD1      LD2
## 1 2.136514 1.636255
```

所以线性判别得分为 2.136514 和 1.636255

根据后验概率第二类最大，所以分类在第二类

**(c) 用 (b) 中的线性判别函数将样本观测值分类。计算 $APER$ 和 $\hat{E}(AER)$**

```
pred<-predict(ld) # 用模型对学习样本分类
tab1<-table(data3$V5,pred$class)
tab1
```

```
##
##      1  2  3
##   1 50  0  0
##   2  0 49  1
##   3  0  4 46
```

观察表格其中出错了 5 个值, 根据公式 $APER = \frac{5}{150} = 0.033$

```
ld1<-lda(V5~V2+V4,data3,prior=c(1/3,1/3,1/3),CV=T) #CV=T 运用提留方法
tab2<-table(data3$V5,ld1$class)
tab2
```

```
##
##      1  2  3
##   1 50  0  0
##   2  0 48  2
##   3  0  4 46
```

观察表格其中出错了 6 个值根据公式 $\hat{E}(AER) = \frac{6}{150} = 0.04$

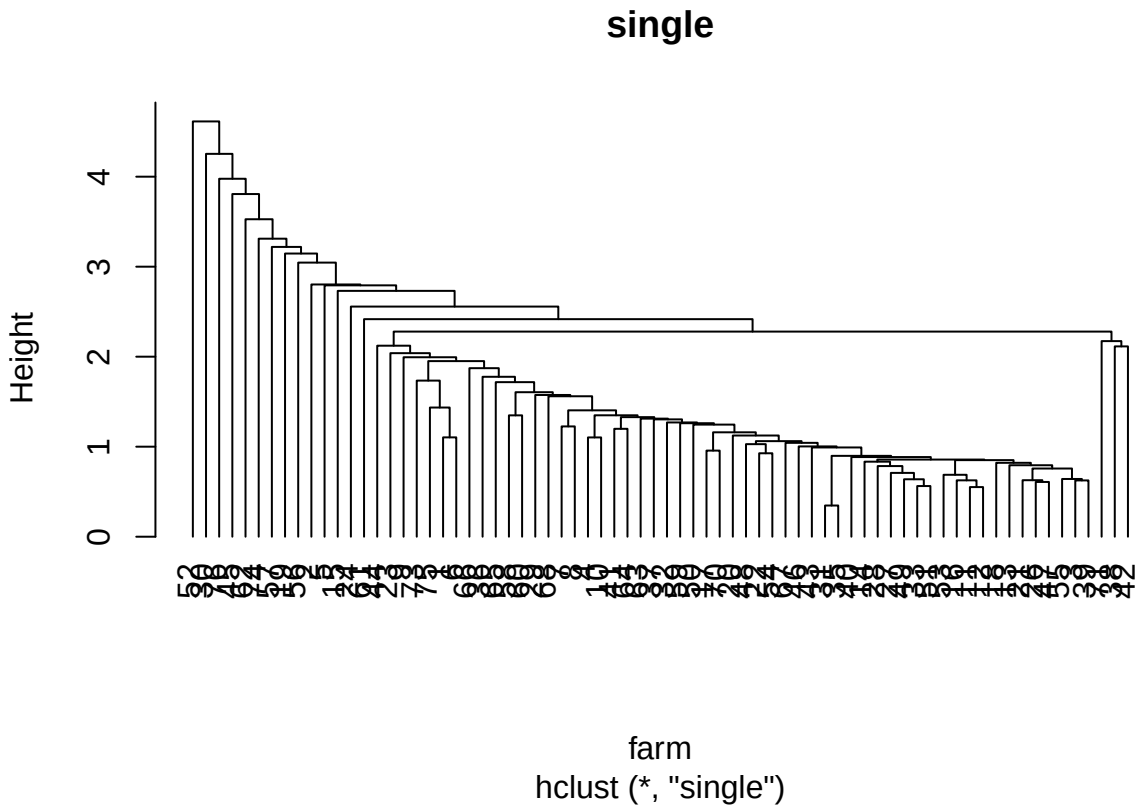
## 第四题

**(a) 用单连接法，完全连接法对农场做聚类。构造连接树图并比较结果。**

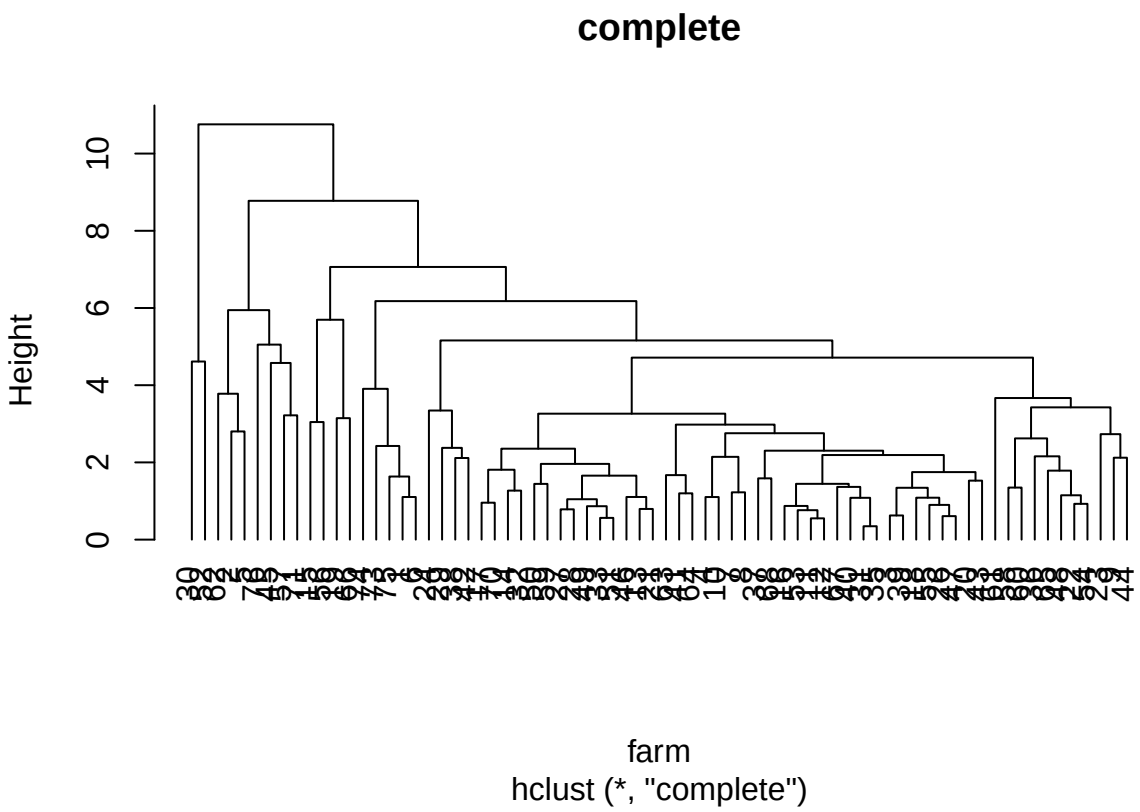
```
data4=read.table("T4.DAT")   # 读入数据
data4<-data4[-c(25,34,69,72),]   # 去掉离群值 25,34,69,72
dim(data4)   # 输出数据的维度
```

```
## [1] 72  9
```

```
dist4=dist(scale(data4), method = "euclidean", p = 2)
# 对标准化后的数据计算欧氏距离，"euclidean" 表示欧氏距离，维度为 2
D4_single<-hclust(dist4,method="single")
D4_com<-hclust(dist4,method="complete")
# 进行聚类分析，"single" 为最短距离法，"complete" 为最长距离法。
plot(D4_single,hang=-1,main="single",sub=NULL,xlab="farm")
```




```
plot(D4_com,hang=-1,main="complete",sub=NULL,xlab="farm")
```




**(b) 用三个不同的 K 值对农场作聚类。**

取 K=8,16,32 进行 K 均值聚类分析如下

```
km1<-kmeans(scale(data4),8)
km1$cluster  # 输出分组结果
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  3  4  8  8  4  3  3  3  1  3  8  8  8  5  7  8  5  8  6  8  8  5  1  5  8  8
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
##  5  5  2  8  5  5  8  1  8  7  8  8  3  1  8  1  4  8  8  5  8  5  8  2  8  5
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
##  8  7  7  1  5  1  1  4  8  3  1  6  8  3  5  7  3  6  3  4
```

```
print(cutree(D4_single,k=8))  # 输出最短距离法分 K 组的结果
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
##  1  1  2  1  1  1  1  1  1  1  1  1  1  1  1  3  1  1  1  1  1  1  4  1  1
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
```

```
##  1  1  5  1  1  1  1  6  1  1  1  1  1  1  1  1  1  7  1  8
```

```
km2<-kmeans(scale(data4),16)
km2$cluster
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  2  6 11 16  6  2  9  9 10 16 12 12 13 13  3 12  8 11 14 16 11  8  4  5 11 13
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
## 13  5 15 12  8 13 12  4  8  3 11 12  9  5 11 10  1 13 11  4 13 13 13 15 12  8
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
## 11  3  1  4  8  4 10  6 13  9  4 14 12  9  8  3  2 14  2  7
```

```
print(cutree(D4_single,k=16))
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  1  2  1  1  3  1  1  1  1  1  1  1  1  1  4  1  1  1  5  1  1  1  1  6  1  1
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
##  1  1  7  1  1  1  1  1  1  8  1  1  1  8  1  1  9  1  1  1  1  1  1 10  1  1
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
##  1 11 12  1  1  1 13 14  1  1  1  1  1  1  1  8  1 15  1 16
```

```
km3<-kmeans(scale(data4),32)
km3$cluster
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  2 25 27 22 25  2 20 20  4 22 11 11 23 29 18 11 24 21 17 21 23  3  1 32 23 26
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
## 26 19 10  9 29 26  9  1  9 28 27 11 13 28  6  4 15 21 23  3 26 16 26 12 11  3
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
## 27 18  8  3 16  3  5 25 13 13  1 14 11 30 24  8  2  7  2 31
```

```
print(cutree(D4_single,k=32))
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27
##  1  2  3  4  5  1  6  6  7  4  3  3  3  3  8  3  3  3  9  3  3  3 10 11  3  3
## 28 29 30 31 32 33 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
##  3 12 13  3  3  3  3 14  3 15  3  3  3 16  3 17 18  3  3  3  3  3  3 19  3  3
## 55 56 57 58 59 60 61 62 63 64 65 66 67 68 70 71 73 74 75 76
##  3 20 21 22  3 22 23 24  3  3 25 26  3 27  3 28 29 30 31 32
```

将分类结果与（a）比较得当 K=32 时结果更好