

# 实验四：判别分析

沈雨萱 3180104691

## 目录

1 实验概况	1
2 实验结果	1

## 1 实验概况

一. 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）处理判别分析的基本步骤；（2）熟悉各类判别方法；

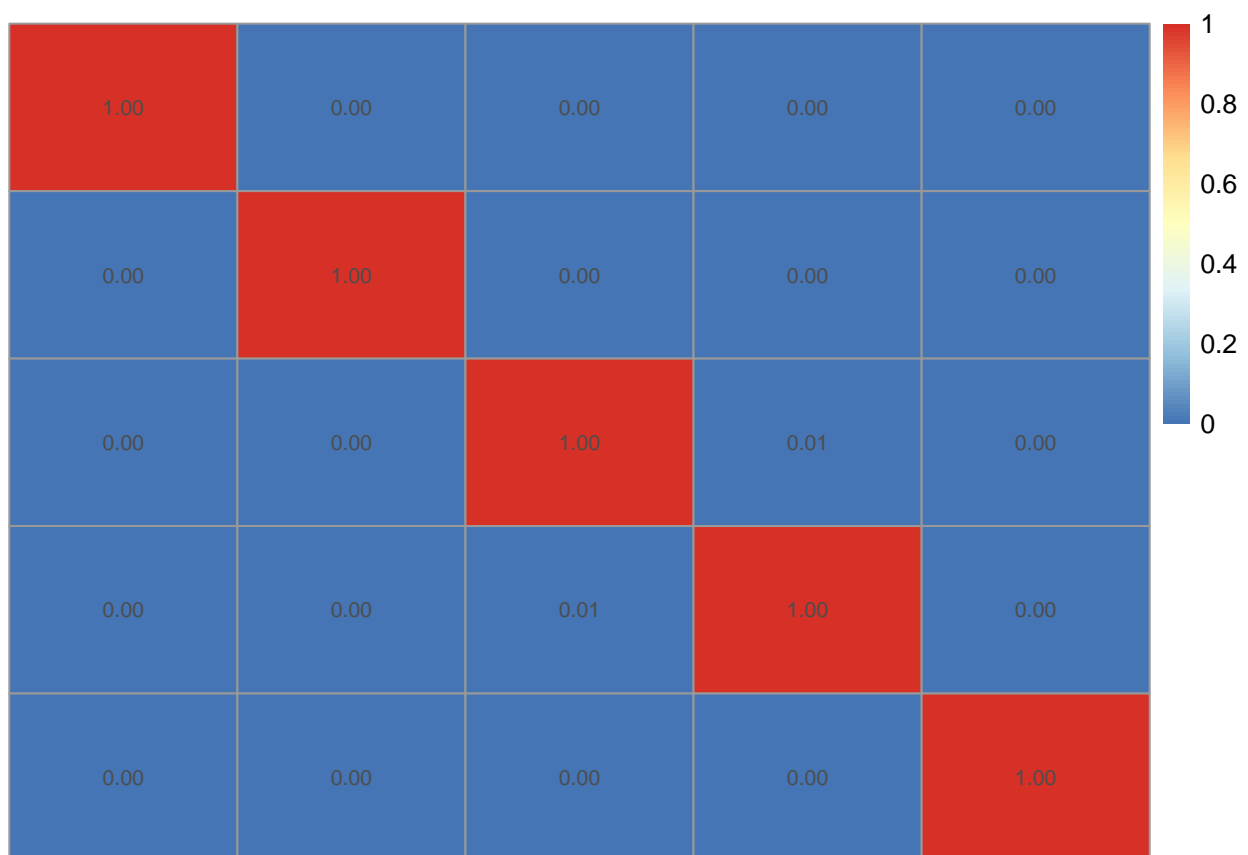
二. 实验内容 1. 利用第五章的数据和上机指导五，熟悉 R 在判别分析中的应用（请动手操作）；2. 采用“肝胆病患者检查数据”（见附件）。这是一组医院病人的资料，基本包括了四变量分别为：总胆红素 (umol/L)，白蛋白 (g/L)，碱性磷酸酶，谷丙转氨酶和医生诊断结果，希望通过这组数据（学习样本）建立判别肝、胆疾病的判别函数，并应用于“体检数据”中，根据体检资料（见附件）分析是否有得肝胆疾病的可能性；

## 2 实验结果

a. 均值的显著性检验

```
data <- read.csv(" 肝胆病患者检查数据.csv",encoding = "UTF-8",na.strings=c("", " ", "NA"))
d1=data %>% filter(group==1)
d1m=d1[1:nrow(d1),2:5]
d2=data %>% filter(group==2)
d2m=d2[1:nrow(d2),2:5]
d3=data %>% filter(group==3)
d3m=d3[1:nrow(d3),2:5]
d4=data %>% filter(group==4)
d4m=d4[1:nrow(d4),2:5]
```

```
d5=data %>% filter(group==5)
d5m=d5[1:nrow(d5),2:5]
dlist <- list(d1m,d2m,d3m,d4m,d5m)
a <- matrix(0,5,5)
for (i in 1:5){
  for (j in 1:5){
    a[i,j] = HotellingsT2(dlist[[i]],dlist[[j]])$p.value
  }
}
pheatmap(a,cluster_rows = FALSE, cluster_cols = FALSE, display_numbers = TRUE)
```



可以看到 5 中疾病类型组间均值存在显著性差异

b. 建立判别函数

```
datar <- data[,2:6]
attach(datar)
lda(group~BIL+Alb+ALP+ALT, datar)
```

## Call:

```
## lda(group ~ BIL + Alb + ALP + ALT, data = datar)
##
## Prior probabilities of groups:
##          1          2          3          4          5
## 0.31395349 0.04360465 0.17151163 0.09883721 0.37209302
##
## Group means:
##          BIL          Alb          ALP          ALT
## 1  88.96111 33.05463 139.00926 310.74074
## 2 155.84667 34.84000 192.13333 1303.60000
## 3  19.15593 42.70339  77.64407  31.77966
## 4  23.54412 39.04706 109.41176  47.67647
## 5  14.05234 47.69375  60.10938  23.28906
##
## Coefficients of linear discriminants:
##          LD1          LD2          LD3          LD4
## BIL -0.002624856 -0.0014151382  0.0169912889 -0.006453130
## Alb  0.158422305 -0.1181964481  0.0033508644 -0.099499017
## ALP -0.005435544  0.0001076113 -0.0121008443 -0.017931771
## ALT -0.002135831 -0.0028019445 -0.0007891008  0.001607685
##
## Proportion of trace:
##      LD1      LD2      LD3      LD4
## 0.8617 0.1300 0.0071 0.0011
```

c. 利用判别函数做判别

```
lda.sol <- lda(group~BIL+Alb+ALP+ALT, datar)
results <- predict(lda.sol, datar)
table(datar$group, results$class)
```

```
##
##      1  2  3  4  5
## 1  88  7  8  1  4
## 2   7  8  0  0  0
## 3   5  0 12  0 42
## 4  10  0 13  3  8
## 5   0  0  0  0 128
```

可以看到对疾病类型 3,4（慢胆，急胆）的预测效果不佳, 对疾病类型 1, 2, 5 的效果尚可

c. 根据判别函数去分析体检数据

```
datatj <- read.csv(" 体检资料.csv",encoding = "UTF-8",na.strings=c("", " ", "NA"))
tj_l <- datatj[,c(16,18,20,22)]
colnames(tj_l) <- c('ALT','BIL','ALP','Alb')
results <- predict(lda.sol, tj_l)
table(results$class)
```

```
##
```

```
##  1   2   3   4   5
```

```
##  0   0   3   0 219
```

根据所建立的判别函数确定体检对象大多为正常，有少数可能有慢胆。