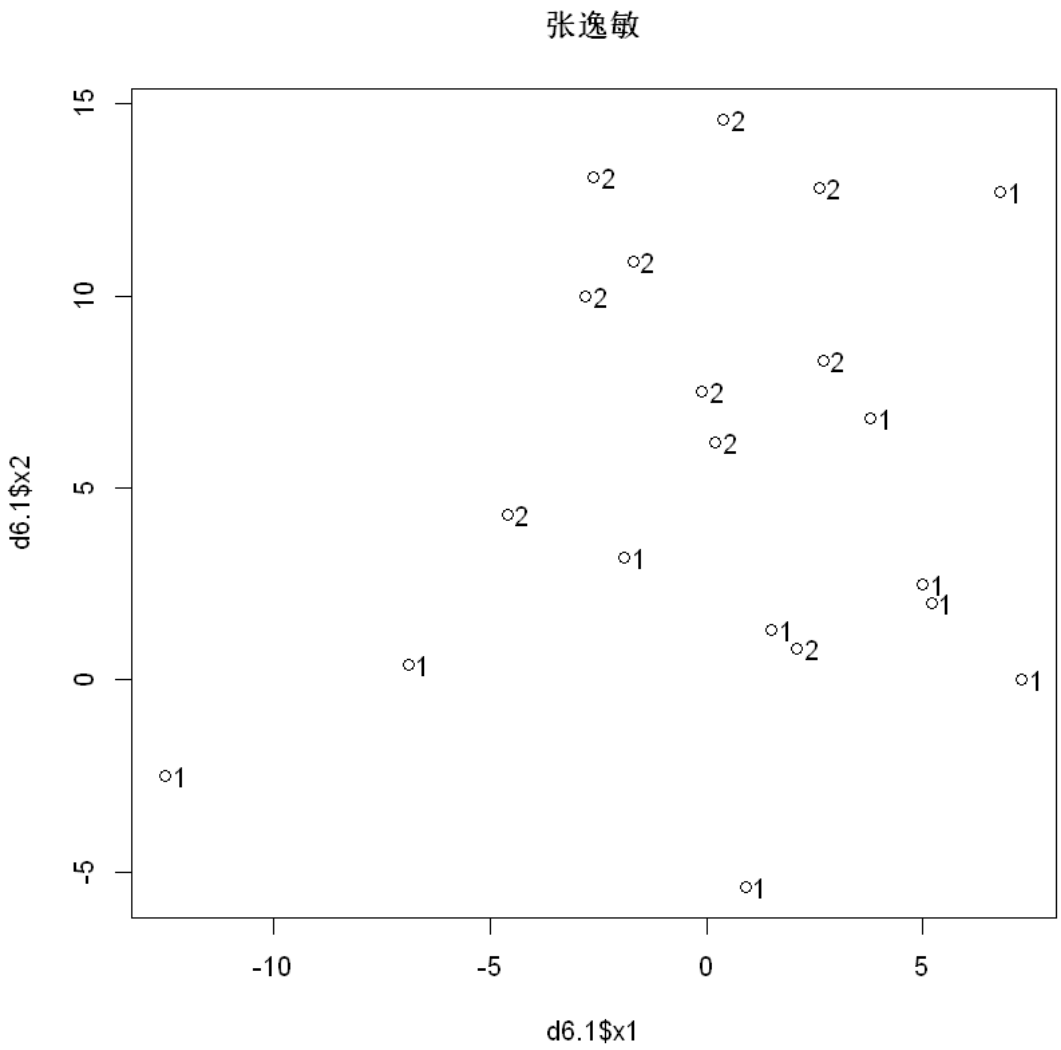
实验二-利用 R 语言实现判别分析

张逸敏

2023 年 4 月 6 日

0.1 6.2 线性判别分析

[1]:
```
setwd("C:\\Users\\zym\\Desktop\\multivariate\\exp2\\")
d6.1 = read.table('./data/6_1.txt', header=T)
plot(d6.1$x1, d6.1$x2, main="张逸敏")
text(d6.1$x1, d6.1$x2, d6.1$G, adj=-0.5)
```



[2]:
```
d6.1
```

A data.frame: 20 × 4

G <int>	x1 <dbl>	x2 <dbl>	G1 <chr>
1	-1.9	3.2	雨
1	-6.9	0.4	雨
1	5.2	2.0	雨
1	5.0	2.5	雨
1	7.3	0.0	雨
1	6.8	12.7	雨
1	0.9	-5.4	雨
1	-12.5	-2.5	雨
1	1.5	1.3	雨
1	3.8	6.8	雨
2	0.2	6.2	晴
2	-0.1	7.5	晴
2	0.4	14.6	晴
2	2.7	8.3	晴
2	2.1	0.8	晴
2	-4.6	4.3	晴
2	-1.7	10.9	晴
2	-2.6	13.1	晴
2	2.6	12.8	晴
2	-2.8	10.0	晴

[3]:
```
options(warn = -1)
library(MASS)
attach(d6.1)
ld = lda(G~x1+x2)
ld
```

```
Call:
lda(G ~ x1 + x2)

Prior probabilities of groups:
  1   2
0.5 0.5

Group means:
```

```
     x1   x2
1  0.92 2.10
2 -0.38 8.85

Coefficients of linear discriminants:
          LD1
x1 -0.1035305
x2  0.2247957
```

[4]:
```
Z = predict(ld)
newG = Z$class
cbind(G, Z$x, newG)
```

A matrix: 20 × 3 of type dbl

	G	LD1	newG
1	1	-0.28674901	1
2	1	-0.39852439	1
3	1	-1.29157053	1
4	1	-1.15846657	1
5	1	-1.95857603	1
6	1	0.94809469	2
7	1	-2.50987753	1
8	1	-0.47066104	1
9	1	-1.06586461	1
10	1	-0.06760842	1
11	2	0.17022402	2
12	2	0.49351760	2
13	2	2.03780185	2
14	2	0.38346871	2
15	2	-1.24038077	1
16	2	0.24005867	2
17	2	1.42347182	2
18	2	2.01119984	2
19	2	1.40540244	2
20	2	1.33503926	2

[5]:
```
# 真实为 1 的样本中有一个错判为 2, 真实为 2 的样本中有一个错判为 1
(tab=table(G, newG))
```

```
   newG
G   1 2
  1 9 1
  2 1 9
```

[6]:
```
sum(diag(prop.table(tab))) # 符合率
```

0.9

[7]:
```
# 预测新样本
predict(ld, data.frame(x1=8.1,x2=2.0))
```

$class 1 Levels: 1.’ 1’ 2.’ 2’

$posterior A matrix: 1 × 2 of type dbl

	1	2
1	0.9327428	0.06725717

$x A matrix: 1 × 1 of type dbl

	LD1
1	-1.591809

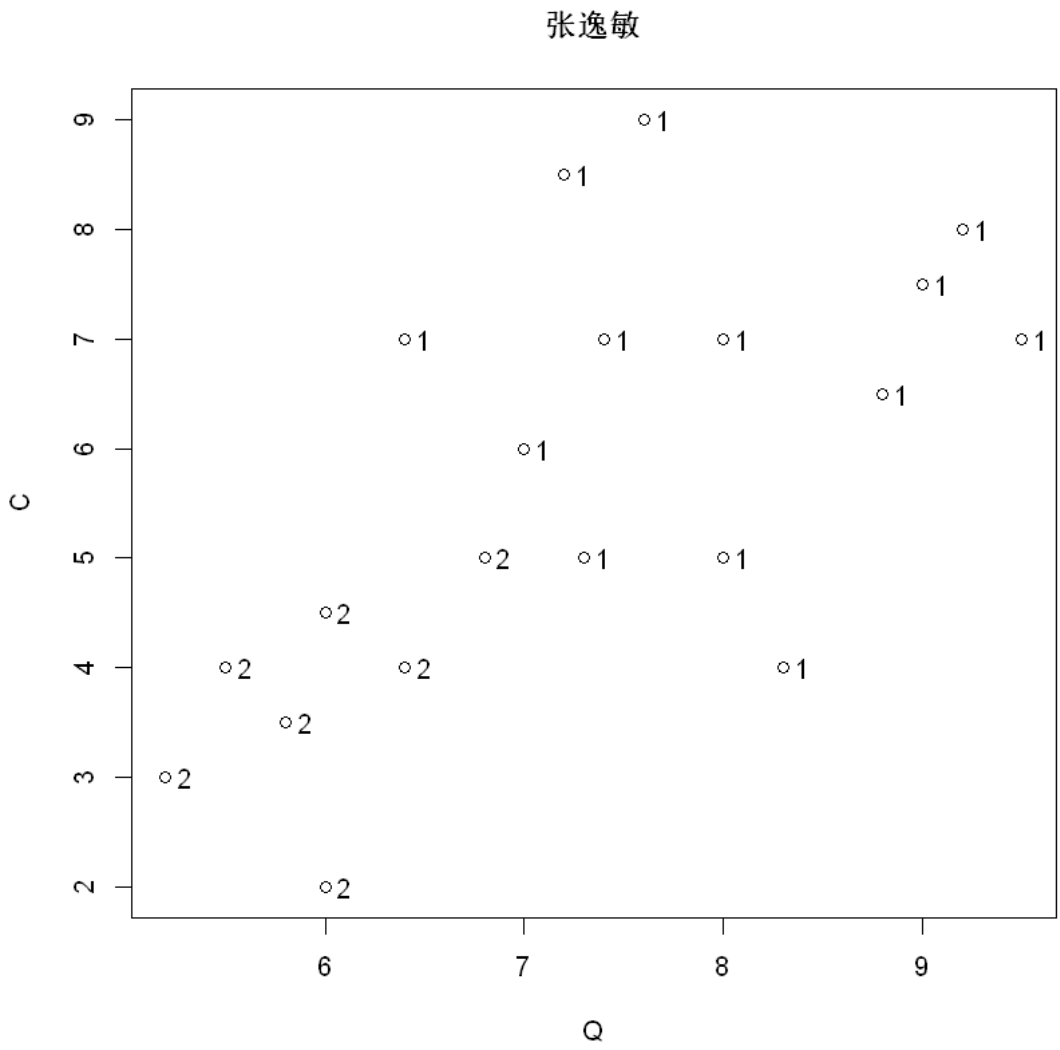
0.2 6.3 距离判别法

[1]:
```
options(warn = -1)
d6.2 = read.table("./data/6_2.txt", header = T)
attach(d6.2)
d6.2
# G=1 表示畅销，G=2 表示滞销；Q：质量评分；C：功能评分；P：销售价格
```

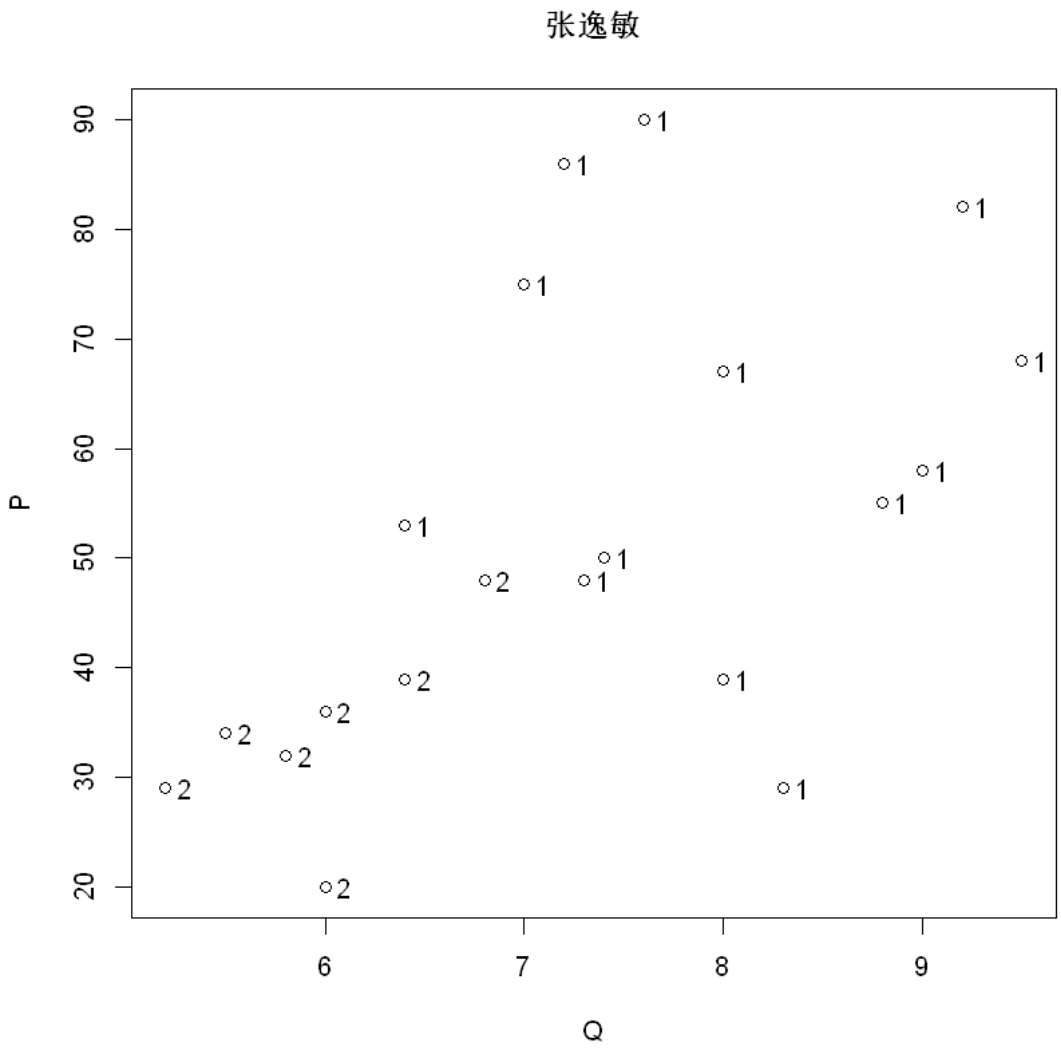
A data.frame: 20 × 4

G <int>	Q <dbl>	C <dbl>	P <int>
1	8.3	4.0	29
1	9.5	7.0	68
1	8.0	5.0	39
1	7.4	7.0	50
1	8.8	6.5	55
1	9.0	7.5	58
1	7.0	6.0	75
1	9.2	8.0	82
1	8.0	7.0	67
1	7.6	9.0	90
1	7.2	8.5	86
1	6.4	7.0	53
1	7.3	5.0	48
2	6.0	2.0	20
2	6.4	4.0	39
2	6.8	5.0	48
2	5.2	3.0	29
2	5.8	3.5	32
2	5.5	4.0	34
2	6.0	4.5	36

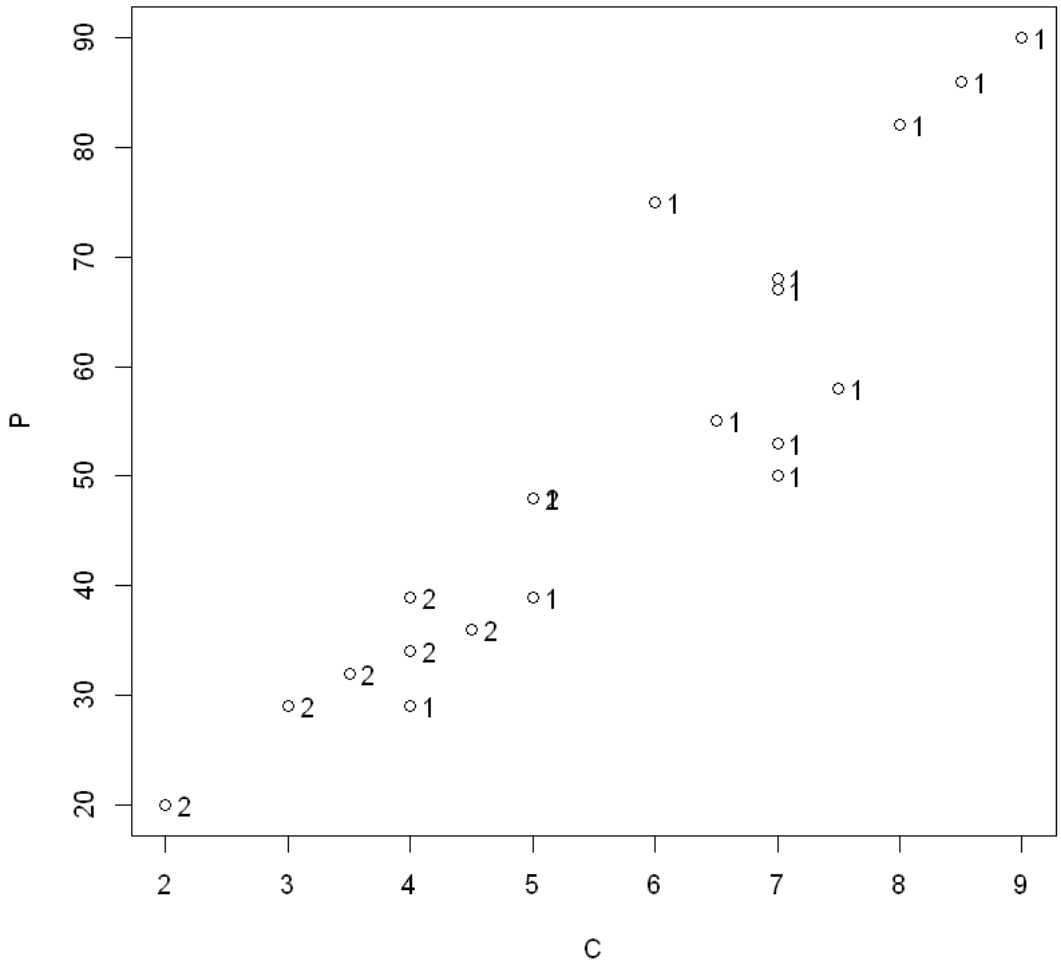
[2]:
```
# 质量评分和功能评分的散点图
plot(Q,C,main="张逸敏")
text(Q,C,G,adj=-0.8)
```



[3]:
```
# 质量评分和销售价格的散点图
plot(Q,P,main="张逸敏")
text(Q,P,G,adj=-0.8)
```



[4]:
```
# 功能评分和销售价格的散点图
plot(C,P,main="张逸敏")
text(C,P,G,adj=-0.8)
```



0.2.1 二次判别函数 qda() 的用法

[5]:
```
library(MASS)
qd=qda(G~Q+C+P)
qd
```

```
Call:
qda(G ~ Q + C + P)

Prior probabilities of groups:
```

```
   1    2
0.65 0.35


Group means:
         Q        C        P
1 7.976923 6.730769 61.53846
2 5.957143 3.714286 34.00000
```

[6]:
```
predict(qd)
cbind(G,newG=predict(qd)$class)
# 发现预测值和真实值全都一样
```

$class 1. 1 2. 1 3. 1 4. 1 5. 1 6. 1 7. 1 8. 1 9. 1 10. 1 11. 1 12. 1 13. 1 14. 2 15. 2 16. 2 17. 2 18. 2 19. 2 20. 2

Levels: 1. ' 1' 2. ' 2'

$posterior A matrix: 20 × 2 of type dbl

	1	2
1	1.0000000000	1.038272e-13
2	0.9999999943	5.696656e-09
3	0.9999999669	3.313397e-08
4	0.9999999957	4.297984e-09
5	0.9999999969	3.117297e-09
6	1.0000000000	7.625938e-15
7	1.0000000000	4.793512e-18
8	0.9999999952	4.810094e-09
9	0.9990859902	9.140098e-04
10	1.0000000000	3.454373e-11
11	1.0000000000	3.770380e-11
12	0.9976747923	2.325208e-03
13	0.5401116448	4.598884e-01
14	0.0006321752	9.993678e-01
15	0.0094288115	9.905712e-01
16	0.1453044024	8.546956e-01
17	0.0004122557	9.995877e-01
18	0.0009545983	9.990454e-01
19	0.0019820983	9.980179e-01
20	0.0346609285	9.653391e-01

A matrix: 20 × 2 of type int

G	newG
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
2	2
2	2
2	2
2	2
2	2
2	2
2	2

[7]:
```
# 新样本判定
predict(qd, data.frame(Q=8,C=7.5,P=65))
```

$class 1 Levels: 1.′ 1′ 2.′ 2′

$posterior A matrix: 1 × 2 of type dbl

	1	2
1	0.9998462	0.0001537705

0.2.2 线性判别函数的应用

[8]:
```
library(MASS)
ld = lda(G~Q+C+P)
ld
```

```
Call:
lda(G ~ Q + C + P)
```

```
Prior probabilities of groups:
   1    2
0.65 0.35

Group means:
         Q        C        P
1 7.976923 6.730769 61.53846
2 5.957143 3.714286 34.00000

Coefficients of linear discriminants:
          LD1
Q -0.82211427
C -0.64614217
P  0.01495461
```

[9]:
```
W = predict(ld)
cbind(G, Wx = W$x, newG=W$class)
```

A matrix: 20 × 3 of type dbl

	G	LD1	newG
1	1	-0.1069501	1
2	1	-2.4486840	1
3	1	-0.3569119	1
4	1	-0.9914270	1
5	1	-1.7445428	1
6	1	-2.5102440	1
7	1	0.3574261	1
8	1	-2.6388274	1
9	1	-1.2304672	1
10	1	-1.8499498	1
11	1	-1.2578515	1
12	1	-0.1244489	1
13	1	0.3531596	1
14	2	2.9416056	2
15	2	1.6046131	2
16	2	0.7642167	2
17	2	3.0877463	2
18	2	2.3162705	2
19	2	2.2697429	2
20	2	1.5655239	2

[10]:
```
# 线性判别函数判断结果，和二次判别函数结果相同
predict(ld,data.frame(Q=8,C=7.5,P=65))
```

$class 1 Levels: 1.′ 1′ 2.′ 2′

$posterior A matrix: 1 × 2 of type dbl

	1	2
1	0.999266	0.0007339578

$x A matrix: 1 × 1 of type dbl

	LD1
1	-1.583447

0.3 多总体距离判别

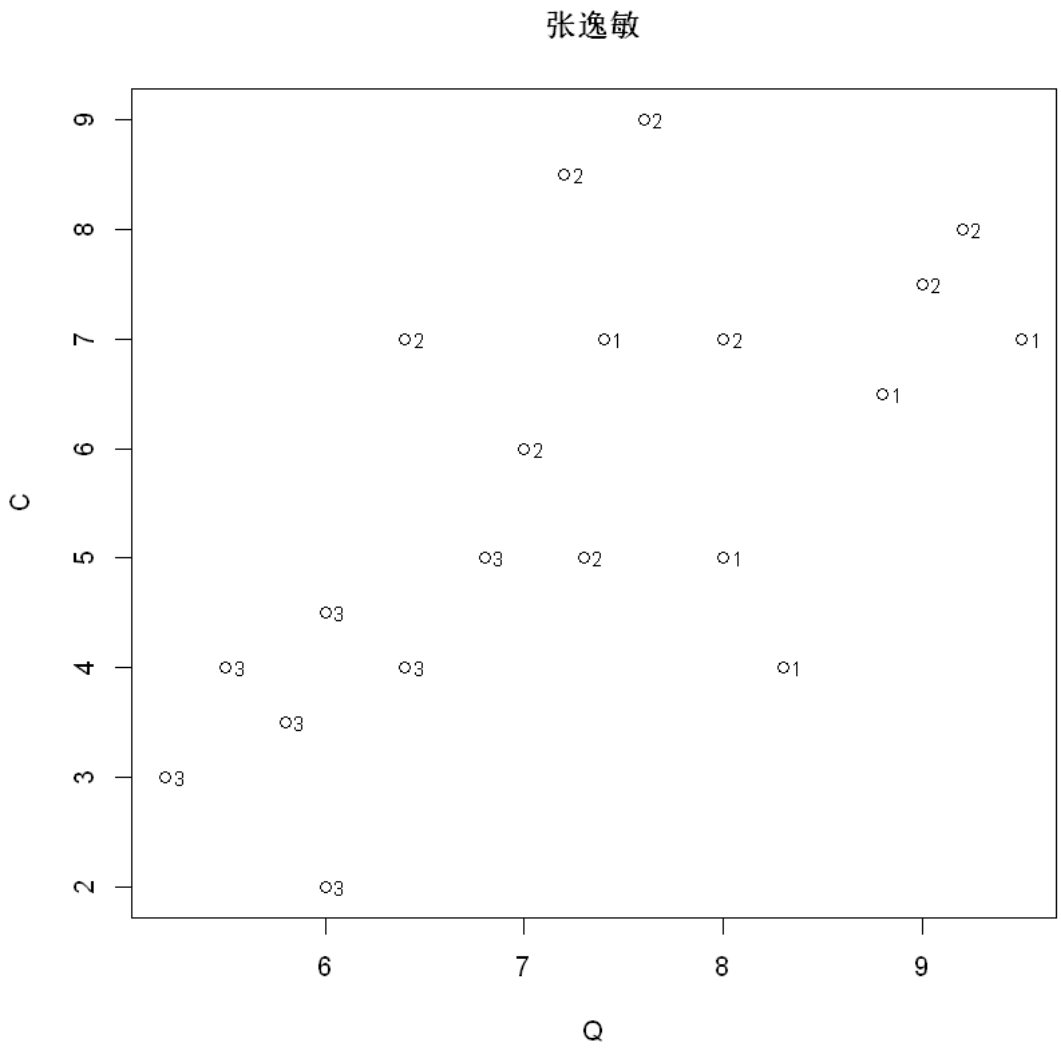
[11]:
```
d6.3 = read.table("./data/6_3.txt", header = T)
d6.3 # G=1 是畅销，G=2 是平销，G=3 是滞销
```

A data.frame: 20 × 4

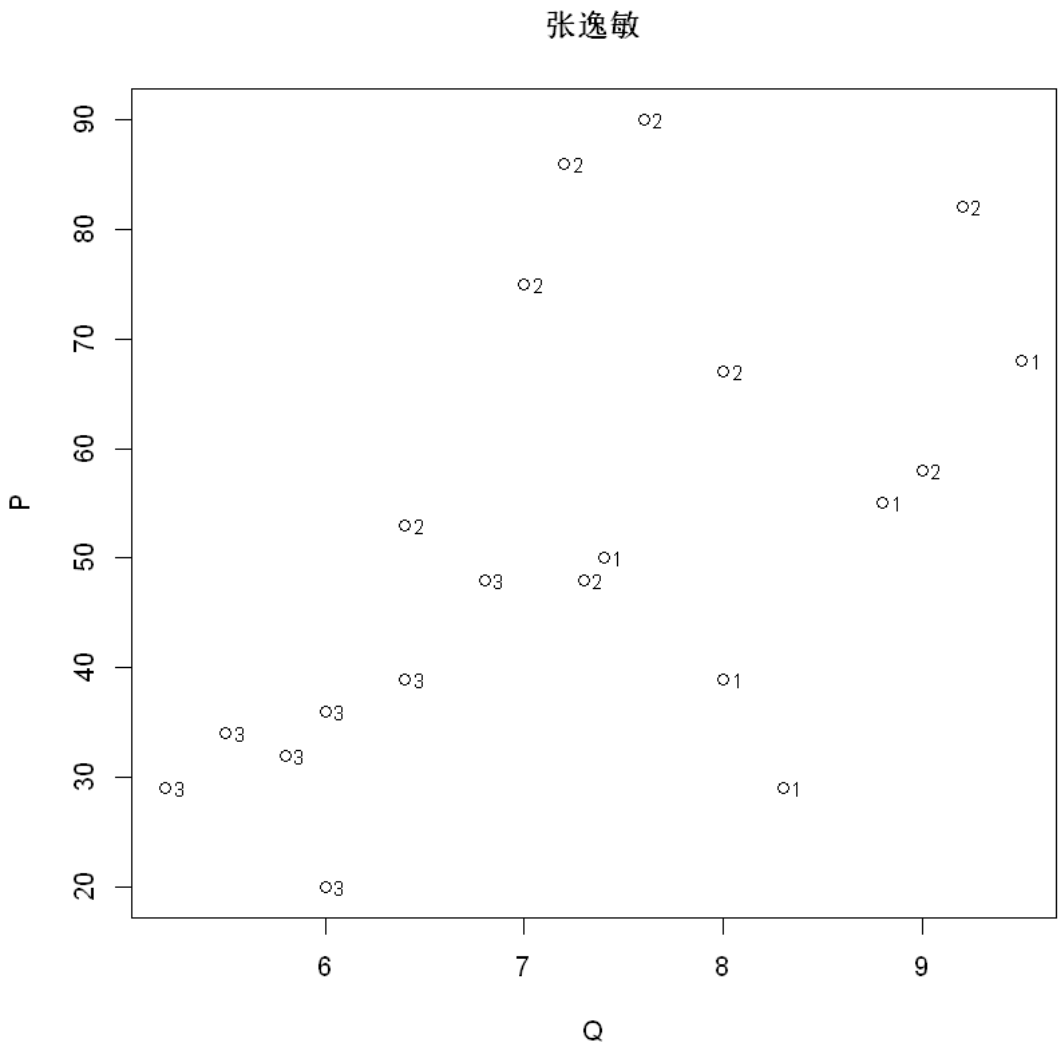
G <int>	Q <dbl>	C <dbl>	P <int>
1	8.3	4.0	29
1	9.5	7.0	68
1	8.0	5.0	39
1	7.4	7.0	50
1	8.8	6.5	55
2	9.0	7.5	58
2	7.0	6.0	75
2	9.2	8.0	82
2	8.0	7.0	67
2	7.6	9.0	90
2	7.2	8.5	86
2	6.4	7.0	53
2	7.3	5.0	48
3	6.0	2.0	20
3	6.4	4.0	39
3	6.8	5.0	48
3	5.2	3.0	29
3	5.8	3.5	32
3	5.5	4.0	34
3	6.0	4.5	36

[12]:
```
options(warn = -1)
attach(d6.3)
plot(Q,C, main="张逸敏")
text(Q,C,G,adj=-0.8,cex=0.75)
```

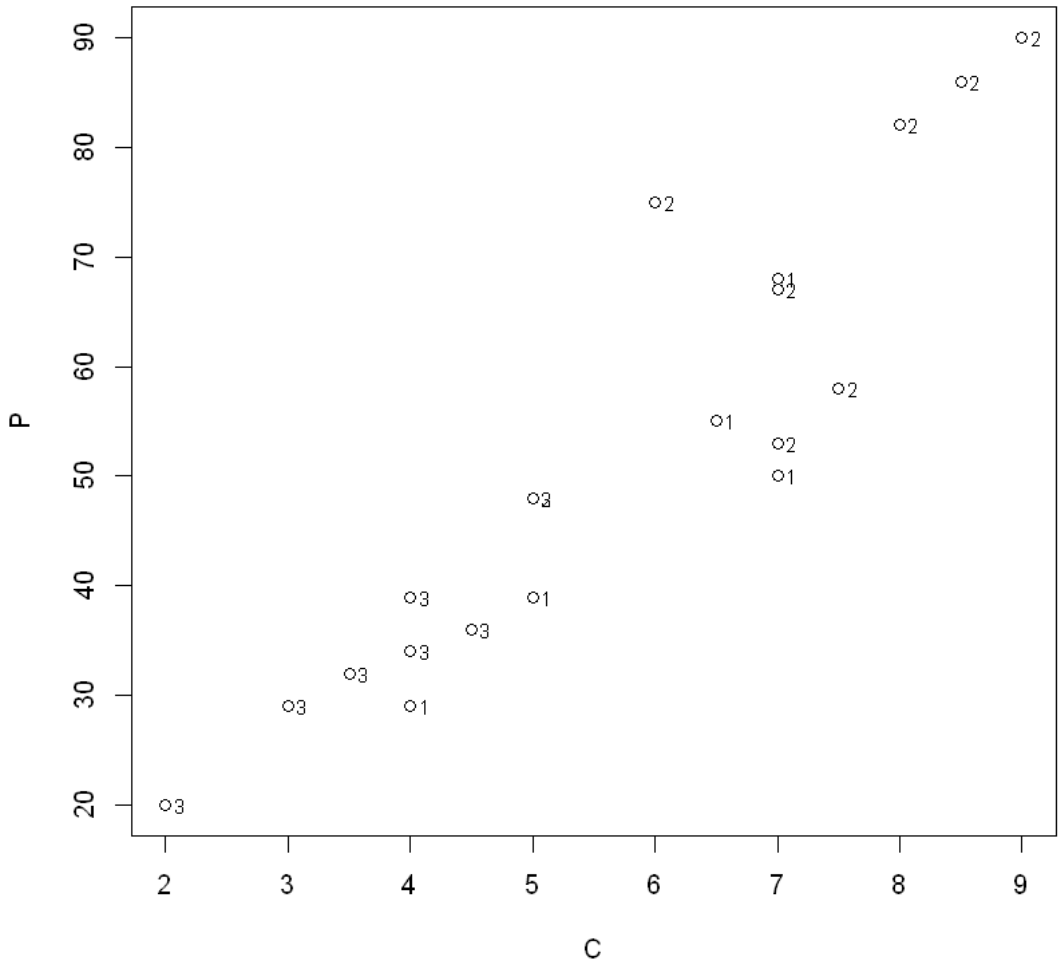
```
The following objects are masked from d6.2:

    C, G, P, Q
```



[13]:
```
plot(Q,P, main="张逸敏")
text(Q,P,G,adj=-0.8,cex=0.75)
```



[14]:
```
plot(C,P, main="张逸敏")
text(C,P,G,adj=-0.8,cex=0.75)
```



0.3.1 线性判别

[15]:
```
ld = lda(G~Q+C+P)
ld
```

```
Call:
lda(G ~ Q + C + P)

Prior probabilities of groups:
   1    2    3
```

```
0.25 0.40 0.35

Group means:
         Q        C      P
1 8.400000 5.900000 48.200
2 7.712500 7.250000 69.875
3 5.957143 3.714286 34.000

Coefficients of linear discriminants:
          LD1         LD2
Q -0.81173396  0.88406311
C -0.63090549  0.20134565
P  0.01579385 -0.08775636

Proportion of trace:
   LD1    LD2
0.7403 0.2597
```

[16]:
```
Z = predict(ld)
newG = Z$class
cbind(G,Z$x,newG)
```

A matrix: 20 × 4 of type dbl

	G	LD1	LD2	newG
1	1	-0.1409984	2.582951755	1
2	1	-2.3918356	0.825366275	1
3	1	-0.3704452	1.641514840	1
4	1	-0.9714835	0.548448277	1
5	1	-1.7134891	1.246681993	1
6	2	-2.4593598	1.361571174	1
7	2	0.3789617	-2.200431689	2
8	2	-2.5581070	-0.467096091	2
9	2	-1.1900285	-0.412972027	2
10	2	-1.7638874	-2.382302324	2
11	2	-1.1869165	-2.485574940	2
12	2	-0.1123680	-0.598883922	2
13	2	0.3399132	0.232863397	3
14	3	2.8456561	0.936722573	3
15	3	1.5592346	0.025668216	3
16	3	0.7457802	-0.209168159	3
17	3	3.0062824	-0.358989534	3
18	3	2.2511708	0.008852067	3
19	3	2.2108260	-0.331206768	3
20	3	1.5210939	0.035984885	3

[17]:
```
(tab=table(G,newG))
```

```
   newG
G   1 2 3
  1 5 0 0
  2 1 6 1
  3 0 0 7
```

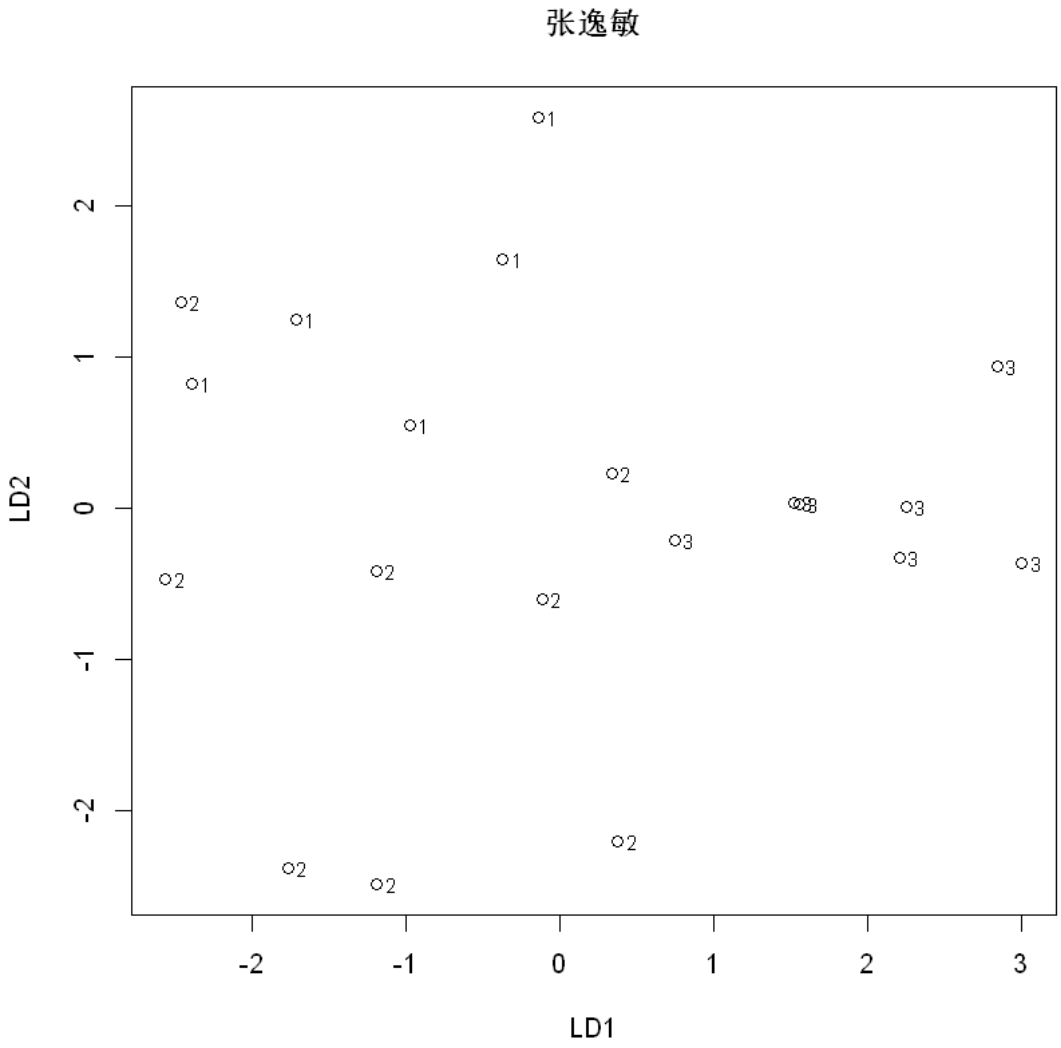
[18]:
```
diag(prop.table(tab,1))
```

1 1 **2** 0.75 **3** 1

[19]:
```
sum(diag(prop.table(tab)))
```

0.9

[20]:
```
plot(Z$x, main="张逸敏")
text(Z$x[,1],Z$x[,2],G,adj=-0.8,cex=0.75)
```



[21]:
```
predict(ld, data.frame(Q=8,C=7.5,P=65))
```

$class 2 Levels: 1. ′ 1′ 2. ′ 2′ 3. ′ 3′

$posterior A matrix: 1 × 3 of type dbl

	1	2	3
1	0.2114514	0.786773	0.001775594

$x A matrix: 1 × 2 of type dbl

	LD1	LD2
1	-1.537069	-0.1367865

0.3.2 二次判别

[22]:
```
(qd=qda(G~Q+C+P))
```

```
Call:
qda(G ~ Q + C + P)

Prior probabilities of groups:
   1    2    3
0.25 0.40 0.35

Group means:
         Q        C      P
1 8.400000 5.900000 48.200
2 7.712500 7.250000 69.875
3 5.957143 3.714286 34.000
```

[23]:
```
Z=predict(qd)
newG = Z$class
cbind(G, newG)
```

A matrix: 20 × 2 of type int

G	newG
1	1
1	1
1	1
1	1
1	1
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	3
3	3
3	3
3	3
3	3
3	3
3	3
3	3

[24]:
```
(tab=table(G,newG))
```

```
   newG
G   1 2 3
  1 5 0 0
  2 0 7 1
  3 0 0 7
```

[25]:
```
sum(diag(prop.table(tab)))
```

0.95

[26]:
```
predict(ld, data.frame(Q=8,C=7.5,P=65))
```

$class 2 Levels: 1. ' 1' 2. ' 2' 3. ' 3'

$posterior A matrix: 1 × 3 of type dbl

	1	2	3
1	0.2114514	0.786773	0.001775594

$x A matrix: 1 × 2 of type dbl

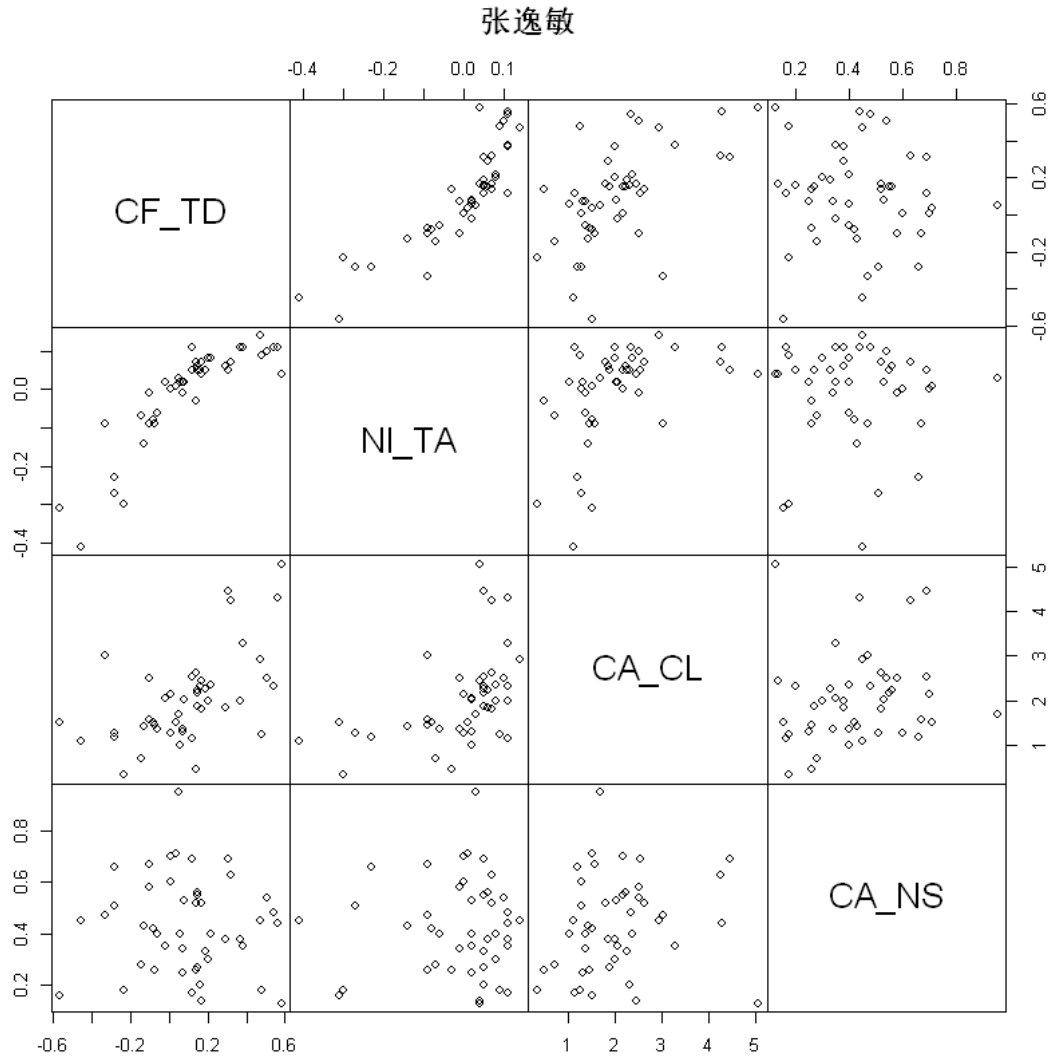
	LD1	LD2
1	-1.537069	-0.1367865

0.4 案例分析：企业财务状况的判别分析

[27]:
```
Case5=read.table("./data/6_case.txt", header = T)
Case5
plot(Case5[,2:5],gap=0, main="张逸敏") # 四个自变量，两两之间的散点图
```

A data.frame: 46 × 5

G <int>	CF_TD <dbl>	NI_TA <dbl>	CA_CL <dbl>	CA_NS <dbl>
1	0.51	0.10	2.49	0.54
1	0.08	0.02	2.01	0.53
1	0.38	0.11	3.27	0.35
1	0.19	0.05	2.25	0.33
1	0.32	0.07	4.24	0.63
1	0.31	0.05	4.45	0.69
1	0.12	0.05	2.52	0.69
1	-0.02	0.02	2.05	0.35
1	0.22	0.08	2.35	0.40
1	0.17	0.07	1.80	0.52
1	0.15	0.05	2.17	0.55
1	-0.10	-0.01	2.50	0.58
1	0.14	-0.03	0.46	0.26
1	0.14	0.07	2.61	0.52
1	0.15	0.06	2.23	0.56
1	0.16	0.05	2.31	0.20
1	0.29	0.06	1.84	0.38
1	0.54	0.11	2.33	0.48
1	-0.33	-0.09	3.01	0.47
1	0.48	0.09	1.24	0.18
1	0.56	0.11	4.29	0.44
1	0.20	0.08	1.99	0.30
1	0.47	0.14	2.92	0.45
1	0.17	0.04	2.45	0.14
1	0.58	0.04	5.06	0.13
2	-0.45	-0.41	1.09	0.45
2	-0.56	-0.31	1.51	0.16
2	0.06	0.02	1.01	0.40
2	-0.07	-0.09	1.45	0.26
2	-0.10	-0.09	1.56	0.67
2	-0.14	-0.07	0.71	0.28
2	0.04	0.01	1.50	0.71
2	-0.06	-0.06	1.37	0.40
2	0.07	-0.01	1.37	0.34
2	-0.13	-0.14	1.42	0.43
2	-0.23	-0.30	0.33	0.18
2	0.07	0.02	1.31	0.25
2	0.01	0.00	2.15	0.70
2	-0.28	-0.23	1.19	0.66



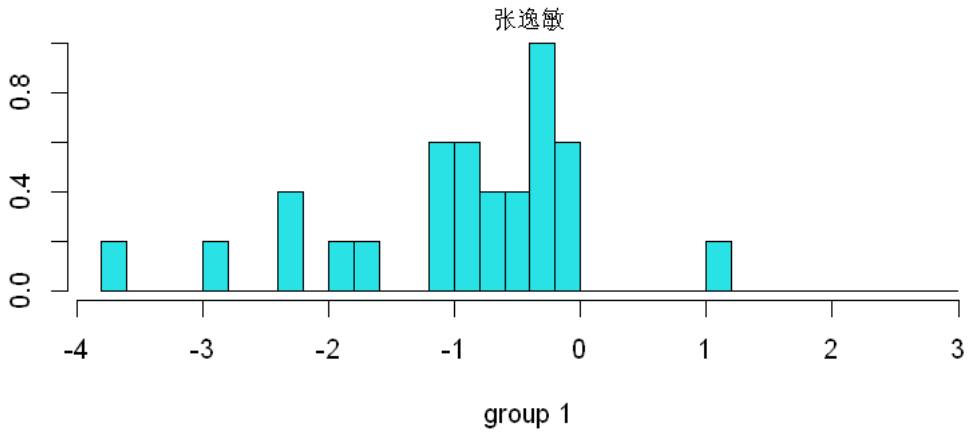
0.4.1 线性判别

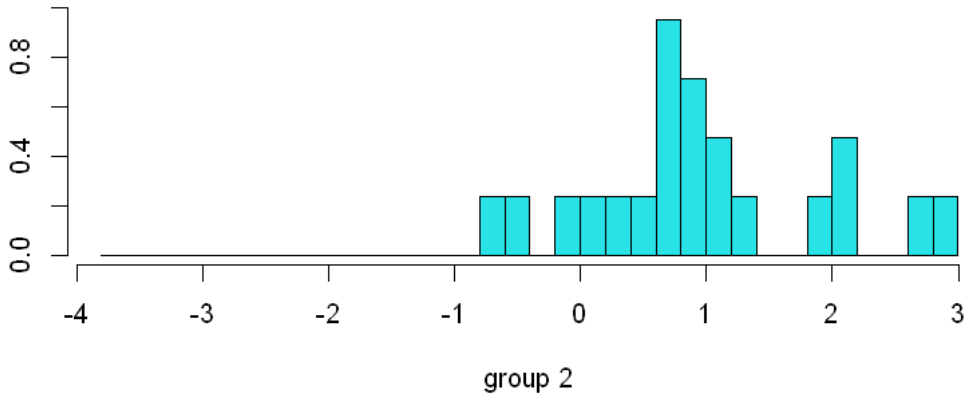
[28]:
```
library(MASS)
ld=lda(G~.,data=Case5);ld  # 线性判别
```

```
Call:
lda(G ~ ., data = Case5)

Prior probabilities of groups:
        1         2
```

```
0.5434783 0.4565217

Group means:
        CF_TD       NI_TA    CA_CL    CA_NS
1  0.23520000  0.05560000 2.593600 0.426800
2 -0.06809524 -0.08142857 1.366667 0.437619

Coefficients of linear discriminants:
             LD1
CF_TD -0.6291667
NI_TA -4.4458516
CA_CL -0.8892843
CA_NS  1.1844801
```

[29]:
```
plot(ld)
mtext("张逸敏")
```





[30]:
```
Zld=predict(ld)
data.frame(Case5$G,Zld$class,round(Zld$x,3))
tab1=table(Case5$G,Zld$class);tab1
sum(diag(tab1))/sum(tab1)
addmargins(tab1)
```

A data.frame: 46 × 3

	Case5.G <int>	Zld.class <fct>	LD1 <dbl>
1	1	1	-1.013
2	1	1	0.028
3	1	1	-1.895
4	1	1	-0.625
5	1	1	-2.210
6	1	1	-2.230
7	1	1	-0.395
8	1	1	-0.158
9	1	1	-0.783
10	1	1	-0.076
11	1	1	-0.268
12	1	1	-0.102
13	1	2	1.271
14	1	1	-0.778
15	1	1	-0.354
16	1	1	-0.813
17	1	1	-0.308
18	1	1	-1.005
19	1	1	-0.185
20	1	1	-0.265
21	1	1	-2.808
22	1	1	-0.569
23	1	1	-1.655
24	1	1	-0.971
25	1	1	-3.562
26	2	2	2.997
27	2	2	1.904
28	2	2	0.776
29	2	2	0.790
30	2	2	1.196
31	2	2	1.426
32	2	2	0.764
33	2	2	0.887
34	2	2	0.512
35	2	2	1.278
36	2	2	2.725
37	2	2	0.325
38	2	2	0.238
39	2	2	2.249

```
   1  2
1 24  1
2  3 18
```

0.91304347826087

A table: 3 × 3 of type dbl

	1	2	Sum
1	24	1	25
2	3	18	21
Sum	27	19	46

线性判别符合率：91.3%

0.4.2 二次判别

[31]:
```
qd=qda(G~.,data=Case5);qd  # 二次判别
Zqd=predict(qd)
#data.frame(Case5$G,Zqd$class,round(Zqd$post,3)*100)
tab2=table(Case5$G,Zqd$class);tab2
sum(diag(tab2))/sum(tab2)
addmargins(tab2)
```

```
Call:
qda(G ~ ., data = Case5)

Prior probabilities of groups:
        1         2
0.5434783 0.4565217

Group means:
        CF_TD       NI_TA    CA_CL    CA_NS
1  0.23520000  0.05560000 2.593600 0.426800
2 -0.06809524 -0.08142857 1.366667 0.437619
```

```
   1  2
1 24  1
2  2 19
```

0.934782608695652

A table: 3 × 3 of type dbl

	1	2	Sum
1	24	1	25
2	2	19	21
Sum	26	20	46

二次判别符合率：93.5%，高于线性判别