第五章 判别分析

一、盐泉含钾性判别

1、盐泉的特征数值

第一类盐泉

13.85 2.79 7.80 49.60 A

22.31 4.67 12.31 47.80 A

28.82 4.63 16.18 62.15 A

15.29 3.54 7.50 43.20 A

28.79 4.90 16.12 58.10 A

第二类盐泉

2.18 1.06 1.22 20.60 B

3.85 0.80 4.06 47.10 B

11.40 0.00 3.50 0.00 B

3.66 2.42 2.14 15.10 B

12.10 0.00 5.68 0.00 B

待判盐泉

8.55 3.38 5.17 26.10

28.60 2.40 1.20 127.00

20.70 6.70 7.60 30.20

7.90 2.40 4.30 33.20

3.19 3.20 1.43 9.90

12.40 5.10 4.43 24.60

16.80 3.40 2.31 31.30

15.00 2.70 5.02 64.00

2、sas程序

data chapter5;

input x1 X2 X3 x4 group $;

cards;

13.85 2.79 7.80 49.60 A

22.31 4.67 12.31 47.80 A

28.82 4.63 16.18 62.15 A

15.29 3.54 7.50 43.20 A

28.79 4.90 16.12 58.10 A

2.18 1.06 1.22 20.60 B

3.85 0.80 4.06 47.10 B

11.40 0.00 3.50 0.00 B

3.66 2.42 2.14 15.10 B

12.10 0.00 5.68 0.00 B

;

data chapter5test;

input x1 x2 x3 x4 group $;

cards;

8.55 3.38 5.17 26.10

28.60 2.40 1.20 127.00

20.70 6.70 7.60 30.20

7.90 2.40 4.30 33.20

3.19 3.20 1.43 9.90

12.40 5.10 4.43 24.60

16.80 3.40 2.31 31.30

15.00 2.70 5.02 64.00

；

proc discrim data=chapter5 testdata=chapter5 pool=yes list testlist;

class group;

var x1 x2 x3 x4;

testclass group;

run;

3、SAS结果分析

| **到 group 的广义平方距离** | | |
| --- | --- | --- |
| **从 group** | **A** | **B** |
| **A** | 0 | 37.02876 |
| **B** | 37.02876 | 0 |

从表中可以看出，A与B的广义平方距离为37.0826，

| **以下对象的线性判别函数: group** | | |
| --- | --- | --- |
| **变量** | **A** | **B** |
| **常数** | -42.24731 | -5.16272 |
| **x1** | 7.67412 | 2.93107 |
| **X2** | 5.54881 | 1.35698 |
| **X3** | -13.96307 | -5.37383 |
| **x4** | 1.18131 | 0.455 |

线性判别函数分别为

A=-42.24731+7.67421x1+5.54881x2-13.96307x3+1.18131x4

B=-5.16272+2.93107X1+1.35681X2-5.37383X3+0.445X4

最后，第2,3,6,7,8五个盐泉为含钾的。

二、胃癌的鉴别

1、胃癌检验的生化指标值

类别 x1 x2 x3 x4

1 228 134 20 11

1 245 134 10 40

1 200 167 12 27

1 170 150 7 8

1 100 167 20 14

2 225 125 7 14

2 130 100 6 12

2 150 117 7 6

2 120 133 10 26

2 160 100 5 10

3 185 115 5 19

3 170 125 6 4

3 165 142 5 3

3 135 108 2 12

3 100 117 7 2

2、SAS程序

Date chapter5;

Input group $ x1 x2 x3 x4;

Cards;

1 228 134 20 11

1 245 134 10 40

1 200 167 12 27

1 170 150 7 8

1 100 167 20 14

2 225 125 7 14

2 130 100 6 12

2 150 117 7 6

2 120 133 10 26

2 160 100 5 10

3 185 115 5 19

3 170 125 6 4

3 165 142 5 3

3 135 108 2 12

3 100 117 7 2

;

proc print;

run;

proc discrim pool=no distance list;

class group;

prios '1'=0.3333 '2'=0.3333 '3'=0.3333;

var x1 x2 x3;

run;

3、结果分析

| **到 group 的平方距离** | | | |
| --- | --- | --- | --- |
| **从 group** | **1** | **2** | **3** |
| **1** | 0 | 154.74171 | 25.32336 |
| **2** | 20.79790 | 0 | 2.63493 |
| **3** | 18.89697 | 126.06690 | 0 |
| **到 group 的广义平方距离** | | | |
| **从 group** | **1** | **2** | **3** |
| **1** | 18.46741 | 167.03226 | 40.60493 |
| **2** | 39.26532 | 12.29054 | 17.91650 |
| **3** | 37.36439 | 138.35745 | 15.281 |

故由上表可知当为平方距离时

d(2|1)=20.79790 d(1|2)=151.74171

当为广义平方距离为

d(2|1)=39.26532 d(1|2)=167.03226

| **“group”的出错数估计** | | | | |
| --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **合计** |
| **比率** | 0.2000 | 0.0000 | 0.0000 | 0.0667 |
| **先验** | 0.3333 | 0.3333 | 0.3333 |  |

出错概率约等于0.