**多元统计分析判别分析**

**1202110887 徐鸿飞**

**题目**（教材P98第2题）：某商学院在招收研究生时，以学生在大学期间的平均学分x1与管理能力考试成绩x2帮助录取研究生，对申请者规划为3类G1：录取；G2：为录取；G3:待定。下表记录了近期报考者的录取情况。

1 2.96 596

1 3.14 473

1 3.22 482

1 3.29 527

1 3.69 505

1 3.46 693

1 3.03 626

1 3.19 663

1 3.63 447

1 3.59 588

1 3.30 563

1 3.40 553

1 3.50 572

1 3.78 591

1 3.44 692

1 3.48 528

1 3.47 552

1 3.35 520

1 3.39 543

1 3.28 523

1 3.21 530

1 3.58 564

1 3.33 565

1 3.40 431

1 3.38 605

1 3.26 664

1 3.60 609

1 3.37 559

1 3.80 521

2 3.76 546

2 3.24 467

2 2.54 446

2 2.43 425

2 2.20 474

2 2.36 531

2 2.57 542

2 2.35 406

2 2.51 412

2 2.51 458

2 2.36 399

2 2.36 482

2 2.66 420

2 2.68 414

2 2.48 533

2 2.46 509

2 2.63 504

2 2.44 366

2 2.13 408

2 2.41 469

2 2.55 538

2 2.31 505

2 2.41 469

2 2.19 411

2 2.35 321

2 2.60 394

2 2.55 528

2 2.72 399

2 2.85 381

3 2.90 384

3 2.86 494

3 2.85 498

3 3.14 419

3 3.28 371

3 2.89 447

3 3.15 313

3 3.50 404

3 2.89 485

3 2.80 444

3 3.13 416

3 3.01 471

3 2.79 490

3 2.89 431

3 2.91 466

3 2.75 546

3 2.73 467

3 3.12 463

3 3.08 440

3 3.03 419

3 3.00 509

3 3.03 438

3 3.05 399

3 2.85 483

3 3.01 453

3 3.03 414

3 3.04 446

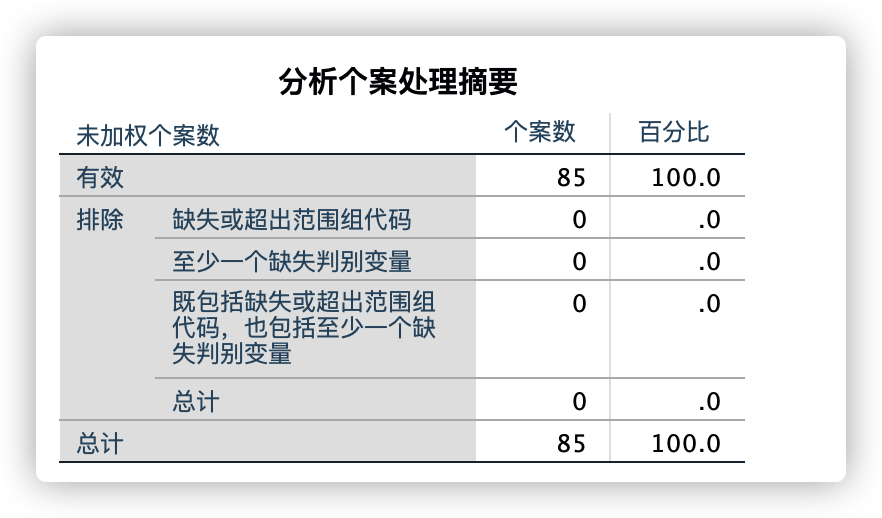
（1）在先验概率相等的的假定下，进行Bayes判别，并确定回代的交叉确认误判率；

（2）在先验概率由样本比例计算的假定下，进行Bayes判别，并确定回代的交叉确认误判率；

（3）设有两名新申请者的（x1，x2）分别为（3.61，513）和（2.91，497），利用所建立判别准则判别他们应该归为那一类？

**解答：**

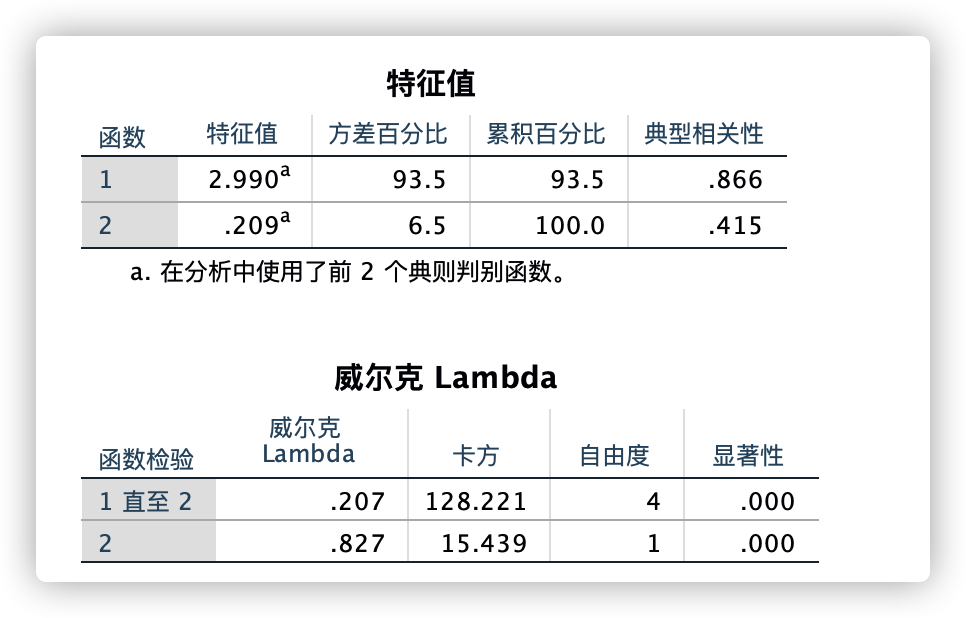
（1）因为自变量只有两个所以不使用步进法。



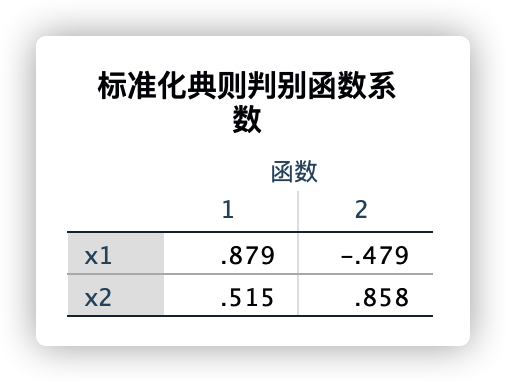
共有85组数据



第一组有29条数据，第二组有29条数据，第三组有27条数据



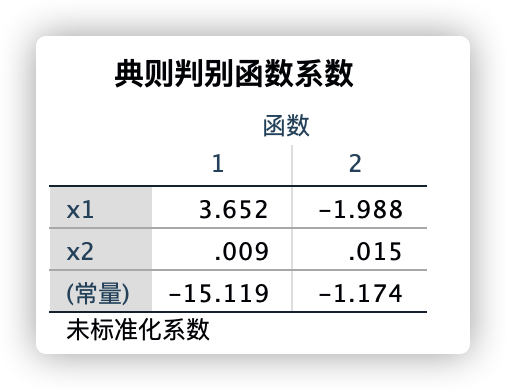
探寻自变量和因变量之间的关系，自变量和因变量的关系函数有2个，第一判别函数是更加准确的，此处1和2的显著性都是很低的，说明正确率都比较高。



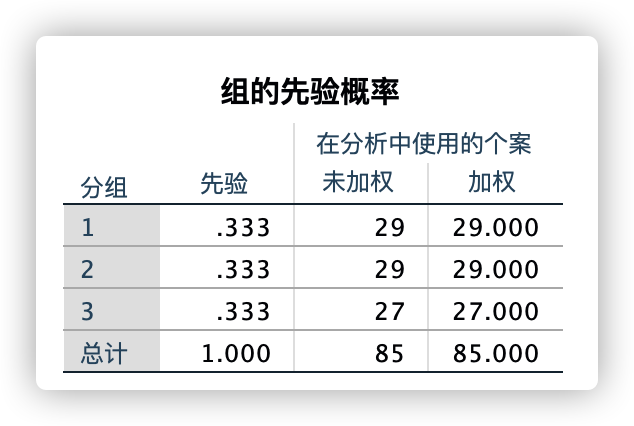
标准化函数的系数，需要先将数据进行标准化之后在进行计算判别，一般不使用。



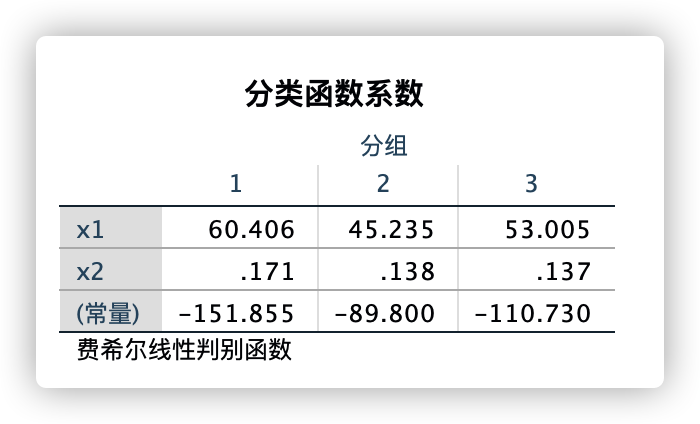
判别变量与标准化典则判别函数之间的汇聚组内相关性变量按函数内相关性的绝对大小排序。每个变量与任何判别函数之间的最大绝对相关性



未标准化的判别函数系数，可以直接把原始数值带入进行计算。把原始数据乘上此处的系数计算得到一个坐标，根据坐标找到点，看点距离哪一组最近，



先验概率，因为假定先验概率相等，所以此处都为33.3%



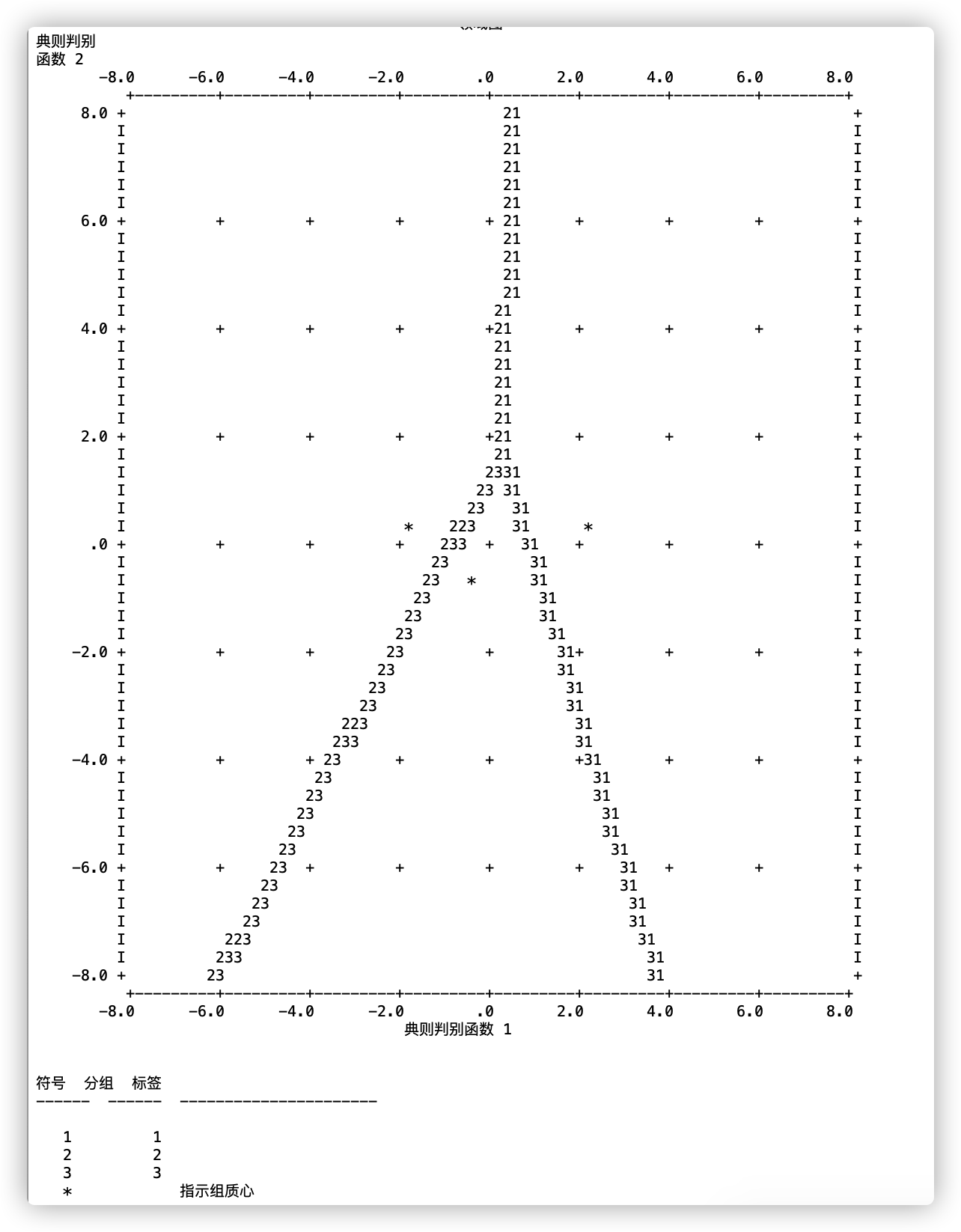
费希尔线性判别函数，可以直接计算出属于那一类的函数，此处的判别函数即：

W1(x) = -151.855 + 60.406x1 + 0.171x2

W2(x) = -89.800 + 45.235x1 + 0.138x2

W3(x) = -110.730 + 53.005x1 + 0.137x2

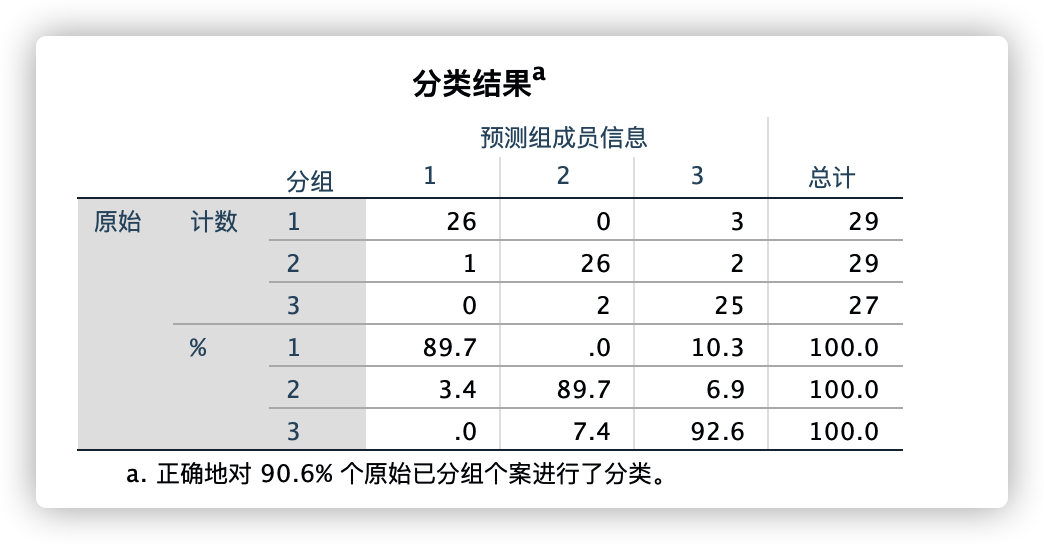
计算出哪个数值更大则判断其属于哪一类。



以第一判断函数为横坐标，第二判别函数为纵坐标进行的一个分类。以未标准化的判别函数计算之后，在这个图中离哪个更近就分为哪一类。

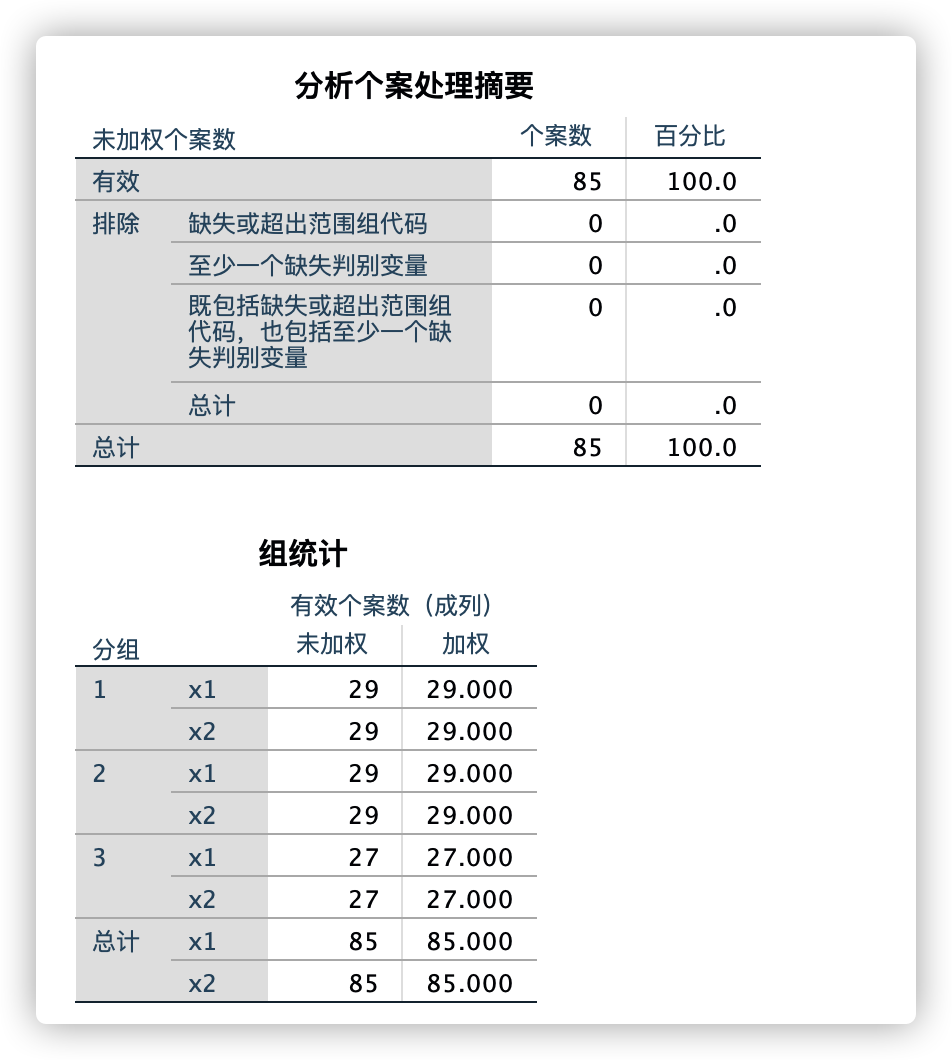


上一张图的可视化处理，将三个类别分为了三处。



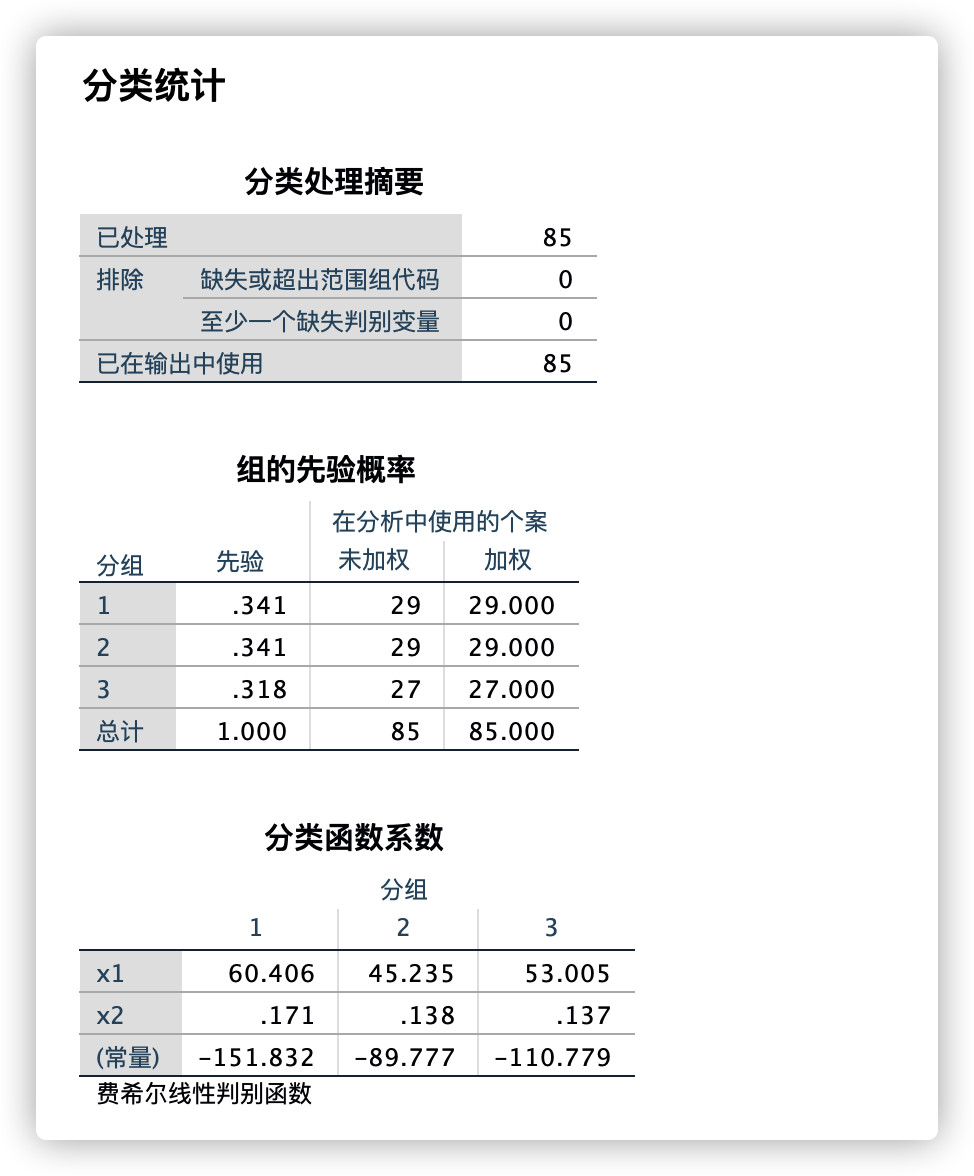
最后的分类结果，从表中可以的知,回代误判率为0。采用交叉确认法得90.6%的判别分析分类与原始个案分类一致,即交叉确认误判率为9.4%,说明该判别分析得符合率还是较高的。

（2）图解释都一样，主要使用分组的先验概率：









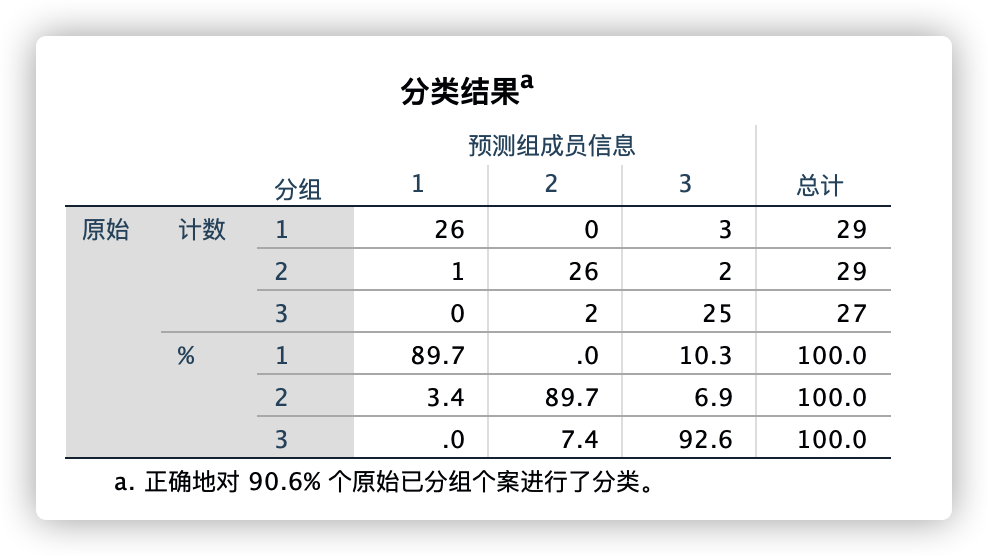
此处的先验概率与之前不同了，是根据组大小计算的

此处的判别函数为：

W1(x) = -151.832 + 60.406x1 + 0.171x2

W2(x) = -89.777 + 45.235x1 + 0.138x2

W3(x) = -110.779 + 53.005x1 + 0.137x2



结果与之前还是一样的，回代误判率为0。采用交叉确认法得90.6%的判别分析分类与原始个案分类一致,即交叉确认误判率为9.4%,说明该判别分析得符合率还是较高的。

（3）x1判为第一类,x2判为第三类,即x1录取,x2待定。