

线性与二次判别分析

判别分析，是一种统计判别和分组技术。已知某种事物有几种类型，现在从各种类型中各取一个样本，由这些样本设计出一套标准，使得从这种事物中任取一个样本，可以按这套标准判别它的类型。

判别分析通常都要设法建立一个判别函数，然后利用此函数来进行判别，线性判别函数（Linear Discriminant Function）是其中一种常用类型，类同于多元回归方程。建立函数必须使用一个训练样品。所谓训练样品就是已知实际分类且各指标的观察值也已测得的样品，用来建立判别函数，对分组（Y）未知但预测变量（X）已知的样本的分组（Y）进行判别。

建立判别函数的方法一般有全模型法和逐步选择法。

1) 全模型法是指将用户指定的全部变量作为判别函数的自变量，而不管该变量是否对研究对象显著或对判别函数的贡献大小。此方法适用于对研究对象的各变量有全面认识的情况。如果未加选择的使用全变量进行分析，则可能产生较大的偏差。

2) 逐步选择法是一种选择最能反映类间差异的变量子集，建立判别函数的方法。它是从模型中没有任何变量开始，每一步都对模型进行检验，将模型外对模型的判别贡献最大的变量加入到模型中，同时也检查在模型中是否存在“由于新变量的引入而对判别贡献变得不太显著”的变量，如果有，则将其从模型中出，以此类推，直到模型中的所有变量都符合引入模型的条件，而模型外所有变量都不符合引入模型的条件为之，则整个过程结束。

二次判别函数，是在线性方程中引入自变量的二次项。

给定分组变量 Y 与一组自变量，本模块自动进行全模型与逐步选择模型的判别分析，自动进行线性判别与二次判别。判别结果输出到数据文件中。给定的数据文件中 Y 已知的记录用于构建模型，对判别结果与实际观察值的一致性进行了统计。输出文件中含对 Y 缺失的数据的判别分组与分组概率。

例1. DEMO 数据对 SBP 的三等分组判别分析，输入界面如下：

判别分析 ?

标题: 判别分析

选择分析对象: 所有数据记录

分类结果变量
sbp.t3: Systolic BP, mmhg 三分组

自变量(连续性)

变量

SEX

Age, years

输出结果:

Discriminant Analysis

Outcome: Systolic BP, mmhg 三分组

Call: SBP.T3 = SEX+AGE

Prior probabilities of groups:

1	2	3
0.32282	0.33922	0.33796

Group means:

	SEX	AGE
1	1.58984	33.63633
2	1.47955	34.06431
3	1.44403	45.77575

Linear discriminants analysis:

Coefficients of linear discriminants:

	LD1	LD2
SEX	-0.40702	1.96999
AGE	0.07552	0.01584

SVD and Proportion of trace:

LD1.svd	LD2.svd	LD1.prop.trace	LD2.prop.trace
8.84324	1.70894	0.964	0.036

- svd: the singular values, the ratio of the between- and within-group standard deviations on the linear discriminant variables.
- LD1 proportion of trace indicates the percentage of between-group variance that the first discriminant function is able to explain from the total amount of between-group variance.
- High trace number = discriminant function plays an important role!

LDA: Original(row) vs Classified(column)

original	1	2	3	NA
1	134 (52.34%)	60 (23.44%)	62 (24.22%)	0 (0%)
2	97 (36.06%)	101 (37.55%)	71 (26.39%)	0 (0%)
3	44 (16.42%)	66 (24.63%)	158 (58.96%)	0 (0%)
NA	11 (29.73%)	12 (32.43%)	14 (37.84%)	0 (0%)

Stepwise Diagonal Discriminant Analysis using LDA

SEX	AGE
TRUE	TRUE

QDA: Original(row) vs Classified(column)

original	1	2	3	NA
1	129 (50.39%)	74 (28.91%)	53 (20.7%)	0 (0%)
2	95 (35.32%)	115 (42.75%)	59 (21.93%)	0 (0%)
3	37 (13.81%)	81 (30.22%)	150 (55.97%)	0 (0%)
NA	11 (29.73%)	13 (35.14%)	13 (35.14%)	0 (0%)

Stepwise Diagonal Discriminant Analysis using QDA

SEX	AGE
TRUE	TRUE

