

多元统计分析 作业 5 判别分析

地信一班 罗皓文 15303096

实验环境：

OS: Windows 7 Pro SP1 x64

CPU: Intel Xeon E3-1241 v3 @ 3.50GHz 3.50GHz

RAM: 8.00 Gb

Soft: SPSS Statistics 19

系统聚类分析

“作业-判别分析数据.sav”为某豆腐干制造厂三种不同种类豆腐干的质量、宽度和长度统计表，每种类别都为 20 个样本，共 60 个样本。根据不同种类豆腐干的特征，建立鉴别不同种类豆腐干的判别方程。

部分数据示例：

	公司编号	固定支出综合率	资产收益率	每千瓦容量成本	每年使用的能源	是否使用核能源
1	1	1.06	9.20	351	9077	0
2	2	.89	13.60	202	5088	1
3	3	1.43	8.90	521	9212	0
4	4	.78	11.20	168	6423	1
5	5	.66	16.30	192	3300	1
6	6	.75	13.50	111	1127	1

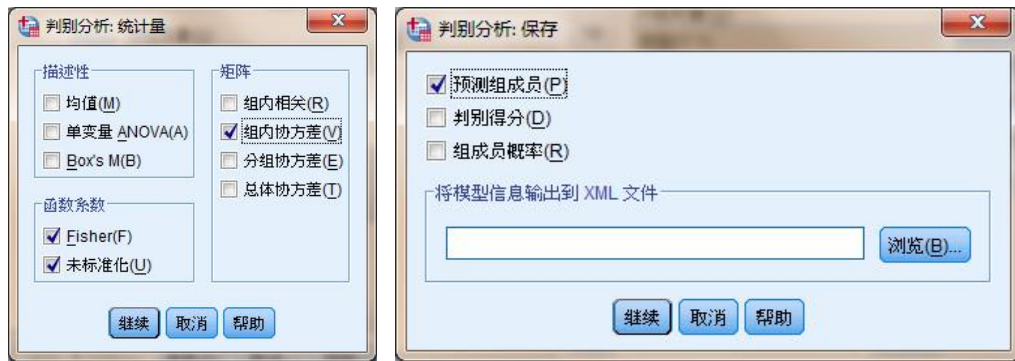
数据变量视图：

	名称	类型	宽度	小数	标签	值	缺失	列	对齐	度量标准	角色
1	编号	数值(N)	8	0		无	无	8	右	度量(S)	输入
2	质量	数值(N)	8	0		无	无	8	右	度量(S)	输入
3	宽度	数值(N)	8	1		无	无	8	右	度量(S)	输入
4	长度	数值(N)	8	0		无	无	8	右	度量(S)	输入
5	类型	数值(N)	8	0		无	无	8	右	度量(S)	输入

(1) 使用 SPSS 判别模块：分析(A) - 分类(F) - 判别...(D)

(2) 选择变量，指定判别参数





结果分析

得到结果包括：

1. 组统计量表

从表 1 可知每一种豆腐干的质量、宽度和长度的均值和标准差，也可以知道总的样本均值和标准差。

表1 组统计量

类型		有效的 N (列表状态)	
		未加权的	已加权的
1	质量	20	20.000
	宽度	20	20.000
	长度	20	20.000
2	质量	20	20.000
	宽度	20	20.000
	长度	20	20.000
3	质量	20	20.000
	宽度	20	20.000
	长度	20	20.000
合计	质量	60	60.000
	宽度	60	60.000
	长度	60	60.000

2. 汇聚的组内矩阵表

从表 2 可知各因素之间的协方差。可以发现，各因素之间的相关性都较小，因此在判别方程中应该不需要剔除变量。

表2 汇聚的组内矩阵^a

		质量	宽度	长度
协方差	质量	10.444	.211	-.303
	宽度	.211	.138	.124
	长度	-.303	.124	9.964

a. 协方差矩阵的自由度为 57。

3. 输入和删除变量情况统计表

从表 3 知，第一步纳入的是质量，到第三步所有变量全部被纳入，显著性为 0，逐步判别没有剔除变量。

表 3 输入的/删除的变量^{a,b,c,d}

步骤	输入的	Wilks's Lambda				Wilks's Lambda 精确 F			
		统计量	df1	df2	df3	统计量	df1	df2	Sig.
1	质量	.038	1	2	57.000	723.540	2	57.000	.000
2	宽度	.003	2	2	57.000	519.862	4	112.000	.000
3	长度	.001	3	2	57.000	476.430	6	110.000	.000

在每个步骤中，输入了最小化整体 Wilk 的 Lambda 的变量。

- 步骤的最大数目是 6。
- 要输入的最小偏 F 是 3.84。
- 要删除的最大偏 F 是 2.71。
- F 级、容差或 VIN 不足以进行进一步计算。

4. 典型判别方程的特征值

从表 4 知，特征值数目为 2，第一个特征值为 75.504，能解释所有变量的 89.9%。

表 4 特征值

函数	特征值	方差的 %	累积 %	正则相关性
1	75.504 ^a	89.9	89.9	.993
2	8.520 ^a	10.1	100.0	.946

- 分析中使用了前 2 个典型判别式函数。

5. 判别方程的有效性检验

从表 5 知，显著性均为 0，因此两个典型方程的判别能力都是显著的。

表 5 Wilks's Lambda

函数检验	Wilks's Lambda	卡方	df	Sig.
1 到 2	.001	369.080	6	.000
2	.105	126.189	2	.000

6. 标准化的典型判别方程

从表 6 可知标准化和非标准化的典型判别方程表达式分别为：

表 6 典型判别式函数系数

	标准化系数		非标准化系数	
	1	2	1	2
质量	.685	.334	.212	.103
宽度	-.613	.816	-1.653	2.197
长度	.642	.288	.203	.091
(常量)			-13.154	-15.393

两个标准化典型判别方程表达式为：

$$\begin{aligned}y_{11} &= 0.685x_1 - 0.613x_2 + 0.642x_3, \\ y_{12} &= 0.334x_1 + 0.816x_2 + 0.288x_3.\end{aligned}$$

两个未标准化典型判别方程表达式为：

$$\begin{aligned}y_{21} &= 0.212x_1 - 1.653x_2 + 0.203x_3 - 13.154, \\ y_{22} &= 0.103x_1 + 2.197x_2 + 0.091x_3 - 15.393.\end{aligned}$$

其中， x_1, x_2, x_3 分别表示变量质量、宽度、长度的值。

7. Bayes 判别方程

从下表可得三个分类方程：

$$\begin{aligned}y_1 &= 2.701x_1 + 16.205x_2 + 2.233x_3 - 94.042, \\ y_2 &= 3.823x_1 + 28.298x_2 + 3.255x_3 - 215.867, \\ y_3 &= 6.996x_1 - 8.532x_2 + 6.331x_3 - 428.650.\end{aligned}$$

表 7 分类函数系数

	类型		
	1	2	3
质量	2.701	3.823	6.996
宽度	16.205	28.298	-8.532
长度	2.233	3.255	6.331
(常量)	-94.042	-215.867	-428.650

Fisher 的线性判别式函数