**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

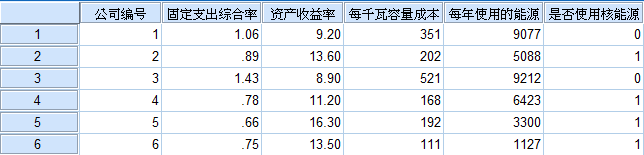
RAM: 8.00 Gb

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。

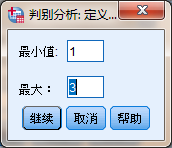
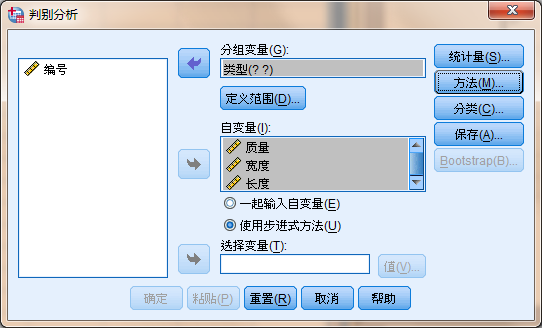
部分数据示例：

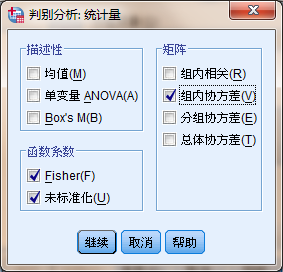
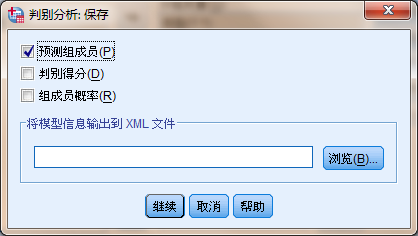


数据变量视图：

（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。 | | | | | | | | | |
| a. 步骤的最大数目是 6。  b. 要输入的最小偏 F 是 3.84。  c. 要删除的最大偏 F 是 2.71。  d. 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 | 75.504a | 89.9 | 89.9 | .993 |
| 2 | 8.520a | 10.1 | 100.0 | .946 |
| a. 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：



两个未标准化典型判别方程表达式为：



其中，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：



| **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 | | | |