

## 2012 高教社杯全国大学生数学建模竞赛

### 承 诺 书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从 A/B/C/D 中选择一项填写）：\_\_\_\_\_ A 题 \_\_\_\_\_

我们的参赛报名号为（如果赛区设置报名号的话）：\_\_\_\_\_

所属学校（请填写完整的全名）：\_\_\_\_\_ 山西大学 \_\_\_\_\_

参赛队员（打印并签名）：1. 许媛 \_\_\_\_\_

2. 鲁花卉 \_\_\_\_\_

3. 赵雅芳 \_\_\_\_\_

指导教师或指导教师组负责人（打印并签名）：\_\_\_\_\_ 吴密景 \_\_\_\_\_

日期：\_\_\_\_\_ 2012 年 9 月 10 日 \_\_\_\_\_

---

赛区评阅编号（由赛区组委会评阅前进行编号）：

## 2012 高教社杯全国大学生数学建模竞赛

### 编号专用页

赛区评阅编号（由赛区组委会评阅前进行编号）：

赛区评阅记录（可供赛区评阅时使用）：

评阅人										
评分										
备注										

全国统一编号（由赛区组委会送交全国前编号）：

全国评阅编号（由全国组委会评阅前进行编号）：

# 葡萄酒的评价

## 摘要

本文利用多元统计的方法分析葡萄及葡萄酒的理化指标及其联系,建立了酿酒葡萄分级模型,并利用得出的几大因子建立了葡萄酒质量评价模型。

针对问题一,首先利用 MATLABR 2008a 进行单因素方差分析,得知红葡萄酒品评检验的  $p$  值为 0.12( $>0.05$ ),白葡萄酒品评检验的  $p$  值为 0.0344( $<0.05$ )。因此两组评酒员对白葡萄酒的评价结果存在显著性差异,而对红葡萄酒的评价结果无显著性差异。然后,对每一组内 10 个评酒员的评价结果做方差,得出第二组评酒员之间的评价差异较小的结论,通过比较两组品酒员的克伦巴赫  $\alpha$  信度系数(内在一致性系数),验证了第二组评价结果更可信的结论。

针对问题二,首先将葡萄的理化指标、葡萄酒的外观分数和口感分数视为变量,进行主成分分析,得出三类主成分:口感因子、外观因子、营养因子,进而得出每个葡萄样品的三个因子的值(归一化后)。其次,利用相关矩阵对芳香物质进行  $R$  型系统聚类分析(聚类方法为类平均法),绘制出聚类树形图并将芳香物质分类,进而从每个类中选取代表性的指标进行计算,得出每种酿酒葡萄样品的香气因子的值(归一化后)。再次,利用上述四个因子(口感因子、外观因子、营养因子、香气因子)建立模糊关系矩阵  $R$ ,由附表一中的各项评分权重定义权重集  $A$ ,选择  $M(\cdot, +)$  为模糊合成算子,建立模糊综合评价模型如下:

$$B = AOR = (a_1, a_2, \dots, a_4)O \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{41} & \cdots & r_{nm} \end{pmatrix}$$

其中矩阵  $B$  的元素  $b_i$  表示第  $i$  个酿酒葡萄样品的评分。最后,根据分级标准将红白葡萄分别分为五级。

针对问题三,首先利用上述主成分分析法及系统聚类分析法将葡萄酒的理化指标同样分为四类因子。然后利用相关矩阵对葡萄酒与葡萄的各因子进行相关性分析,运用 SAS9.1 对酿酒葡萄和葡萄酒的口感因子、外观因子和香气因子建立线性回归模型,并进行拟合优度检验和显著性检验。

针对问题四,我们将葡萄酒的总分数作为被解释变量,把酿酒葡萄和葡萄酒的四大因子(口感因子、外观因子、营养因子、香气因子)视为解释变量,利用 SAS 建立多元线性回归模型,并利用  $\chi^2$  检验来验证此葡萄酒评价模型的正确性,得出了实际值与模型理论值无显著性差异的结论。

**关键词:** 主成分分析 系统聚类分析 糊综合评价 多元线性回归  $\chi^2$  检验

## § 1 问题重述

确定葡萄酒质量时一般是通过聘请一批有资质的评酒员进行品评。每个评酒员在对葡萄酒进行品尝后对其分类指标打分，然后求和得到其总分，从而确定葡萄酒的质量。酿酒葡萄的好坏与所酿葡萄酒的质量有直接的关系，葡萄酒和酿酒葡萄检测的理化指标会在一定程度上反映葡萄酒和葡萄的质量。附件 1 给出了某一年份一些葡萄酒的评价结果，附件 2 和附件 3 分别给出了该年份这些葡萄酒的和酿酒葡萄的成分数据。请尝试建立数学模型讨论下列问题：

1. 分析附件 1 中两组评酒员的评价结果有无显著性差异，哪一组结果更可信？
2. 根据酿酒葡萄的理化指标和葡萄酒的质量对这些酿酒葡萄进行分级。
3. 分析酿酒葡萄与葡萄酒的理化指标之间的联系。
4. 分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响，并论证能否用葡萄和葡萄酒的理化指标来评价葡萄酒的质量？

## § 2 假设及符号说明

### § 2.1 问题假设

- (1) 假设葡萄酒评价结果服从正态分布；
- (2) 假设葡萄酒品质不受年份及环境的影响；
- (3) 假设附表中数据不受人造因素影响；
- (4) 把相关性较弱的元素之间的影响予以忽略。

### § 2.2 符号说明

- $A_i$ ：表示将第  $i$  组评论员的评价结果作为  $A$  因素里的第  $i$  种水平
- $n_i$ ：表示在每个水平下作了第  $i$  次独立试验
- $\bar{X}_i$ ：表示在每个水平下做了第  $i$  次独立试验；
- $S_A$ ：表示各组均值对总方差的偏差平方和，称为组间平方和；
- $S_B$ ：表示各组内的数据对均值偏差平方和的总和。
- $S_E$ ：则表示在同一水平下随机误差的大小。
- $F_{1-\alpha}(r-i, n-r)$ ：表示  $F$  分布的  $1-\alpha$  分位数，
- $P_i$ ：表示红葡萄的第  $i$  主成分
- $Q_i$ ：表示白葡萄的第  $i$  主成分

### § 3 评酒员评价结果显著性和可信度分析(问题一)

#### § 3.1 显著性分析

方差分析是用数理统计分析试验结果鉴别各因素对结果影响程度的方法,在进行显著性差异分析时效果较好,本文通过对方差进行分析来判断显著性差异。将两组评论员作为两种水平 $A_1$ 、 $A_2$ ,在每种水平下分别对白(红)葡萄酒的多个样品进行评价,将组内 10 位评论员评价结果的均值作为因素A,采用单因素方差分析法,假设在试验过程中除A外其它影响指标的因素都保持不变(只有随机因素存在),我们的任务是从试验结果推断,因素A对评价结果有无显著影响,即当A取不同水平时评价结果有无显著差别。

##### § 3.1.1 建立数学模型

将A取两个水平,在水平 $A_i$ 下总体 $x_i$ 服从正态分布 $N(\mu_i, \sigma^2)$ ,其中 $\mu_i, \sigma^2$ 未知, $\mu_i$ 可以互不相同,假定 $x_i$ 有相同的方差。又设在每个水平 $A_i$ 下作了 $n_i$ 次独立试验,即从中抽取容量为 $n_i$ 的样本,记作 $x_{ij}, i=1, \dots, n_i, j=1, \dots, n_i, x_{ij}$ 服从 $N(\mu_i, \sigma^2), i=1, \dots, r, j=1, \dots, n_i$ ,且相互独立.对于白葡萄酒来说,在水平 $A_i$ 下做了28次独立实验,对于红葡萄酒来说,在水平 $A_i$ 下做了27次独立实验。判断A的2个水平对指标有无显著影响,相当于要作以下的假设检验:

$$H_0: \mu_1 = \mu_2 ; H_1: \mu_1 \neq \mu_2$$

由于 $x_{ij}$ 的取值既受不同水平 $A_i$ 的影响,又受 $A_i$ 固定下随机因素的影响,所以将它分解为

$$x_{ij} = \mu_i + \xi_{ij}, i=1,2, \quad (1)$$

其中 $\xi_{ij} \sim N(0, \sigma^2)$ ,且相互独立。记

$$\mu = \frac{1}{n} \sum_{i=1}^r n_i \mu_i, n = \sum_{i=1}^r n_i, \alpha_i = \mu_i - \mu, i=1,2 \quad (2)$$

$\mu$ 是总均值, $\alpha_i$ 是水平 $A_i$ 对指标的效应。由(1)、(2)模型可表为

$$\begin{cases} x_{ij} = \mu + \alpha_i + \xi_{ij} \\ \sum_{i=1}^r \alpha_i = 0 \\ \xi_{ij} \sim N(0, \sigma^2), i=1, \dots, r, j=1, \dots, n_i \end{cases},$$

原假设为(以后略去备选假设)

$$H_0: \alpha_1 = \alpha_2 = 0$$

##### § 3.1.2 统计分析

记

$$\bar{x}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} x_{ij},$$

$\bar{x}_{i\cdot}$ 是第*i*组数据的组平均值,  $\bar{x}$ 是总平均值。考察全体数据对 $\bar{x}$ 的偏差平方和, 则

$$S_T = S_A + S_E,$$

$S_A$ 是各组均值对总方差的偏差平方和, 称为组间平方和;  $S_E$ 是各组内的数据对均值偏差平方和的总和。 $S_A$ 反映A不同水平间的差异,  $S_E$ 则表示在同一水平下随机误差的大小。对 $S_A$ 作进一步分析可得

$$ES_A = (r-1)\sigma^2 + \sum_{i=1}^r n_i \alpha_i^2,$$

以上对 $S_A, S_E, S_r, S_T$ 的分析相当于对组间、组内等方差所谓分析, 所以这种假设检验方法称为方差分析。

### § 3.1.3 方差分析表

表1 单因子方差分析表

方差来源	平方和	自由度	均方	$1-p_r$ 分位数	概率
因素A	$S_A$	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_{1-p_r}(r-1, n-r)$	$p_r$
误差	$S_E$	$n-r$	$\bar{S}_E = \frac{S_E}{n-r}$		
总和	$S_T$	$n-1$			

将试验数据按上述分析、计算的结果排成表1(白葡萄酒), 表2(红葡萄酒)的形式, 即为单因素方差分析表(Matlab中给出的方差分析表)。

表2 白葡萄酒单因素方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	86.75	1	86.7516	4.71	0.0344
Error	994.61	54	18.4187		
Total	1081.36	55			

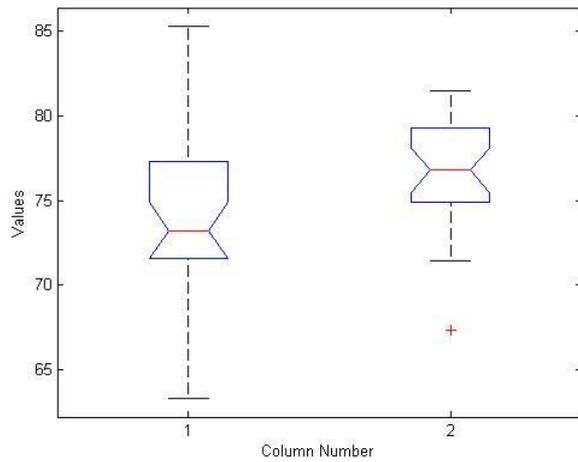


图 1 白葡萄酒方差分析 box 图

由 *MATLAB* 得  $p=0.0344 < 0.05$ , 说明两个小组对白葡萄酒的评价差异性显著。

表 3 红葡萄酒单因素方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	87.15	1	87.1474	2.5	0.12
Error	1813.2	52	34.8692		
Total	1900.35	53			

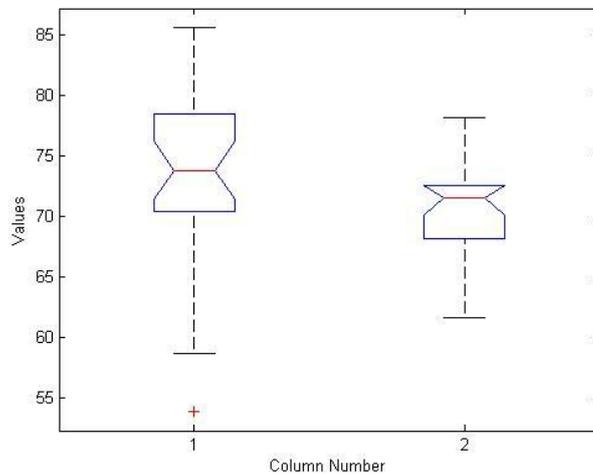


图 2 白葡萄酒方差分析 box 图

由 *MATLAB* 得,  $p=0.1200 > 0.05$ , 说明两个小组对红葡萄的评价差异性不显著。

### § 3.2 可信度

#### § 3.2.1 可信度分析

由于评论员本身之间存在差异, 于是对不同评论员对同一种酒品种的评价效

果做方差，结果如表 4 所示：

表 4 两组评论员对两种酒的评价效果方差表

	第一组 方差	第二组方 差		第一组 方差	第二组方 差
白葡萄酒 1	83	23.29	红葡萄酒 1	83.61	73.69
白葡萄酒 2	180.96	44.16	红葡萄酒 2	35.81	14.6
白葡萄酒 3	328.61	128.24	红葡萄酒 3	41.24	27.64
白葡萄酒 4	40.24	37.89	红葡萄酒 4	97.24	37.16
白葡萄酒 5	113.8	23.65	红葡萄酒 5	55.81	12.29
白葡萄酒 6	146.44	20.45	红葡萄酒 6	53.76	19.01
白葡萄酒 7	35.25	37.96	红葡萄酒 7	93.25	56.41
白葡萄酒 8	165.24	28.01	红葡萄酒 8	39.61	58.6
白葡萄酒 9	83.49	95.64	红葡萄酒 9	29.65	23.16
白葡萄酒 10	191.41	63.36	红葡萄酒 10	27.36	32.56
白葡萄酒 11	159.41	79.04	红葡萄酒 11	63.69	34.24
白葡萄酒 12	104.21	126.04	红葡萄酒 12	71.69	22.61
白葡萄酒 13	153.69	42.09	红葡萄酒 13	40.44	13.76
白葡萄酒 14	102.8	14.29	红葡萄酒 14	32.4	20.84
白葡萄酒 15	118.44	48.64	红葡萄酒 15	77.01	37.21
白葡萄酒 16	160.2	74.01	红葡萄酒 16	16.29	18.09
白葡萄酒 17	129.76	34.61	红葡萄酒 17	79.21	8.25
白葡萄酒 18	140.89	27.21	红葡萄酒 18	42.49	45.24
白葡萄酒 19	41.76	23.44	红葡萄酒 19	42.64	49.64
白葡萄酒 20	57.96	45.04	红葡萄酒 20	23.44	35.16
白葡萄酒 21	155.44	57.96	红葡萄酒 21	104.49	31.96
白葡萄酒 22	124.8	48.24	红葡萄酒 22	45.56	21.84
白葡萄酒 23	39.29	10.44	红葡萄酒 23	29.24	22.29
白葡萄酒 24	100.01	34.69	红葡萄酒 24	67.4	9.65
白葡萄酒 25	30.49	95.85	红葡萄酒 25	58.16	39.36
白葡萄酒 26	65.61	92.61	红葡萄酒 26	28.16	37.4
白葡萄酒 27	129.96	32	红葡萄酒 27	44.8	18.45
白葡萄酒 28	72.41	22.84			
方差总和	3255.57	1411.69	方差总和	1424.45	821.11

将每一组评论员对红葡萄酒和白葡萄酒的评价的方差进行加和平均，得到结果如表 5：

表 5 方差表

	第一组	第二组
红葡萄酒方差(27)	1424.45	821.11
白葡萄酒方差(28)	3255.57	1411.69
平均方差	85.09	40.59

由于第二组的平均方差小于第一组,说明第二组的数据比较集中,即相对第一组来说第二组的评论员之间没有太大差异,所以可以认为第二组可信。

同时我们用 *MATLAB* 程序得出各组内的方差分析结果,从而对上述结论进行进一步说明,见附录一。

### § 3. 2. 2 内在一致性

为了说明结果的准确性,我们采用了克伦巴赫 *a* 信度系数,*a* 信度系数是目前最常用的信度系数。其公式为:

$$a = \left( \frac{k}{k-1} \right) \times \left( 1 - \left( \sum S_{i2} \right) / S_{T2} \right)$$

其中,*K*为量表中题项的总数,*S<sub>i2</sub>*为第*i*题得分的题内方差,*S<sub>T2</sub>*为全部题项总得分的方差。从公式中可以看出,*a*系数评价的是量表中各题项得分间的一致性,属于内在一致性系数。用 *SPSS11.5* 分析的结果如下图:

```
Reliability Coefficients    21 items
Alpha =    .9675           Standardized item alpha =    .9695
```

图 3 第一组白葡萄酒

```
Reliability Coefficients    21 items
Alpha =    .9406           Standardized item alpha =    .9474
```

图 4 第二组白葡萄酒

```
Reliability Coefficients    20 items
Alpha =    .8089           Standardized item alpha =    .8179
```

图 5 第一组红葡萄酒

```
Reliability Coefficients    21 items
Alpha =    .9309           Standardized item alpha =    .9264
```

图 6 第二组红葡萄酒

用信度系数来表示信度的大小。信度系数越大,表明测量的可信程度越大。学者 *DeVellis*(1991)认为,0.60~0.65(最好不要);0.65~0.70(最小可接受值);0.70~0.80(相当好);0.80~0.90(非常好)。由图可知,第二组更可信。

## § 4 对酿酒葡萄进行分级(问题二)

### § 4.1 酿酒葡萄的主成分分析

主成分分析是一种多变量分析方法,也称为矩阵数据分析法,它是在保证数据信息丢失最少的原则下,选出比原始变量个数少,能解释大部分资料的变异的几个新变量,即所谓主成分(Z),也就是对高维变量空间进行降维处理。

#### § 4.1.1 关系矩阵的建立

考虑到除了各酿酒葡萄样品中化学成分的含量存在规律外,葡萄样品的各理化指标之间也存在着相互影响,我们首先对酿酒葡萄的理化指标做相关性分析。

(1) 首先,确定进行相关性分析的变量。为了综合考虑酿酒葡萄的理化指标和葡萄酒的质量,除了附表三中的红(白)葡萄的30个理化指标,我们引入附表二中红(白)葡萄样品的外观分数和口感分数,定义变量如下表6。

表6 变量定义

编号	1	2	3	4	5	6	7	8
变量名称	氨基酸	蛋白质	VC含量	花色苷	酒石酸	苹果酸	柠檬酸	多酚氧化酶
编号	9	10	11	12	13	14	15	16
变量名称	褐变度	DPPH自由基	总酚	单宁	葡萄总黄酮	白藜芦醇	黄酮醇	总糖
编号	17	18	19	20	21	22	23	24
变量名称	还原糖	可溶性固形物	PH值	可滴定酸	固酸比	干物质含量	果穗质量	百粒质量
编号	25	26	27	28	29	30	31	32
变量名称	果梗比	出汁率	果皮质量	果皮L*值	果皮a*值	果皮b*值	外观分数	口感分数

(2) 然后,对于任意两个变量 $X$ 、 $Y$ ,定义相关系数如下:

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2}}$$

其中: $X_i, Y_i$ 是第 $i$ 个红(白)葡萄样品的 $X$ 和 $Y$ 变量, $N=27$ (28)表示红(白)葡萄样品的数量。

(3) 最后,以相关系数作为元素建立相关矩阵 $R$ :

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

其中： $p$ 为进行相关性分析的变量数。

利用 *matlab* 统计工具箱中的 *CORRCOEF* 函数（程序见附录二），分别得到红白葡萄的相关矩阵（32 阶矩阵）。

#### § 4. 1. 2 分析过程

依据求得的相关矩阵，利用主成分分析中的变量分类可以进一步减少红（白）酿酒葡萄的变量数。

（1）首先，利用 *Jacobi* 法求解特征方程  $|\lambda I - R| = 0$ ，得出相关矩阵的特征值及特征向量，然后计算主成分贡献率：

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (i = 1, 2, \dots, p)$$

随后得出累计贡献率：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} \quad (i = 1, 2, \dots, p)$$

其中： $p=32$  为进行相关性分析的变量数。

利用 *Matlab* 计算结果如下表（程序见附录二，完整结果见附录三）：

表 7 红葡萄各主成分贡献率(部分)

主成分 z	特征值	贡献率	累计贡献率
1	7.42203	0.2319	0.2319
2	5.32539	0.1664	0.3983
3	4.02317	0.1257	0.6240
4	3.1123	0.0973	0.6513
5	2.06328	0.0645	0.6858
6	1.77279	0.0554	0.7412
7	1.46704	0.0458	0.7870
8	1.28859	0.0403	0.8273
...	...	...	...

表 8 白葡萄各主成分贡献率(部分)

主成分 z	特征值	贡献率	累计贡献率
1	6.49557	0.2030	0.2030
2	5.11356	0.1598	0.4628
3	3.89841	0.1218	0.6546
4	2.17347	0.0679	0.6825
5	1.91843	0.0600	0.7125
6	1.61295	0.0504	0.7629
7	1.52205	0.0476	0.8105
8	1.41188	0.0441	0.8546
...	...	...	...

由于红（白）葡萄的前三个主成分  $P1$ 、 $P2$ 、 $P3(Q1$ 、 $Q2$ 、 $Q3)$ 的累积贡献率均已达到 60%，因此我们采用前三个主成分反映原数据的全部信息。

(2) 然后，定义因子载荷量（主成分与原始变量的相关系数）如下：

$$F_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} \quad (i, j = 1, 2 \dots 32)$$

其中为对应的单位正交特征向量。

利用 *Matlab* 计算得红（白）葡萄各变量的载荷量如下（程序见附录二，完整结果见附录三）：

表 9 红葡萄各变量的载荷量（部分）

变量编号	第一主成分 P1	第二主成分 P2	第三主成分 P3
1	0.30756	-0.58305	-0.13416
2	0.671986	0.352063	0.226537
3	-0.09387	0.399084	0.16667
4	0.839878	-0.03359	0.071682
5	0.363109	-0.18519	0.234208
6	0.278809	-0.40599	0.319519
7	0.235411	-0.2595	0.437545
8	0.27806	-0.12032	-0.04044
...	...	...	...
32	0.452073	0.034996	-0.55716
特征根	7.42203	5.32539	4.02317

表 10 白葡萄各变量的载荷量（部分）

变量编号	第一主成分 Q1	第二主成分 Q2	第三主成分 Q3
1	-0.5164	-0.22546	0.240116
2	-0.10426	-0.68015	0.194599
3	0.13946	0.098897	-0.6176
4	0.356972	0.454926	0.097383
5	-0.40101	0.431597	0.0731
6	-0.07645	-0.39563	-0.01351
7	-0.20477	0.060883	0.322942
8	0.356619	0.46784	-0.08302
...	...	...	...
32	-0.3571323	-0.0663887	-0.5265025
特征根	7.42203	5.32539	4.02317

(3) 定义因子载荷向量  $F_i = (F_{i1}, F_{i2} \dots F_{i32})$  ( $i=1,2\dots32$ ), 分别以  $F_1, F_2$  的分量为横纵坐标, 利用 *Matlab* 绘制出红(白)葡萄各变量的散布图(第二因子载荷向量对第一因子载荷向量的散布图)如下(程序见附录二):

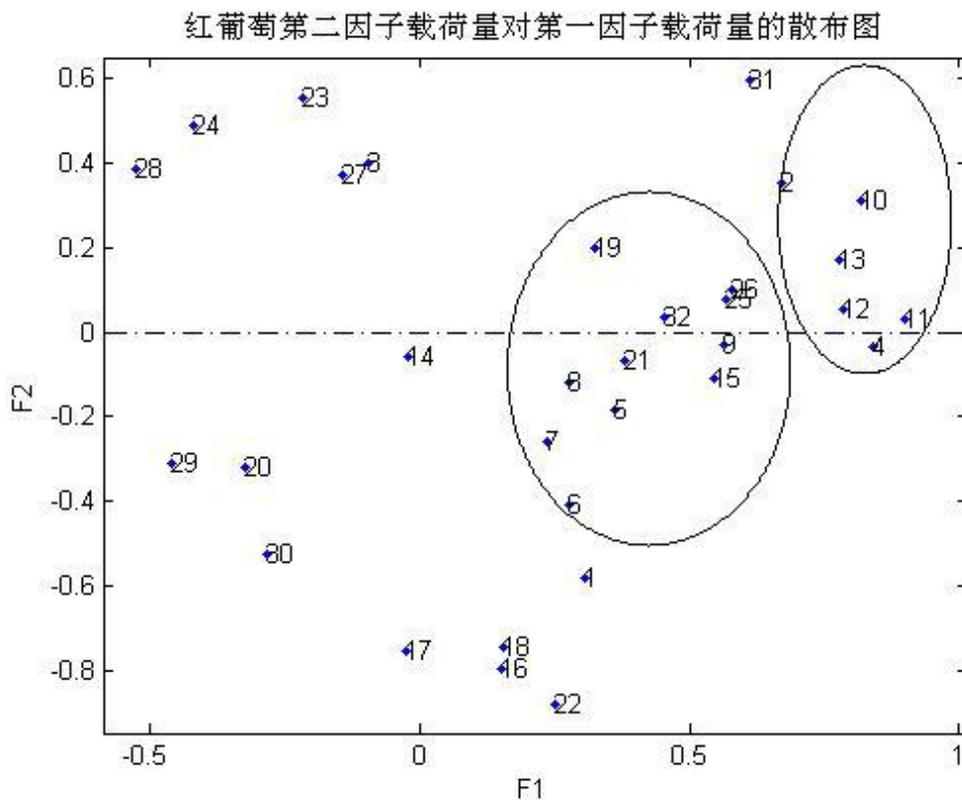


图 7 红葡萄第二因子载荷量对第一因子载荷量的散布图

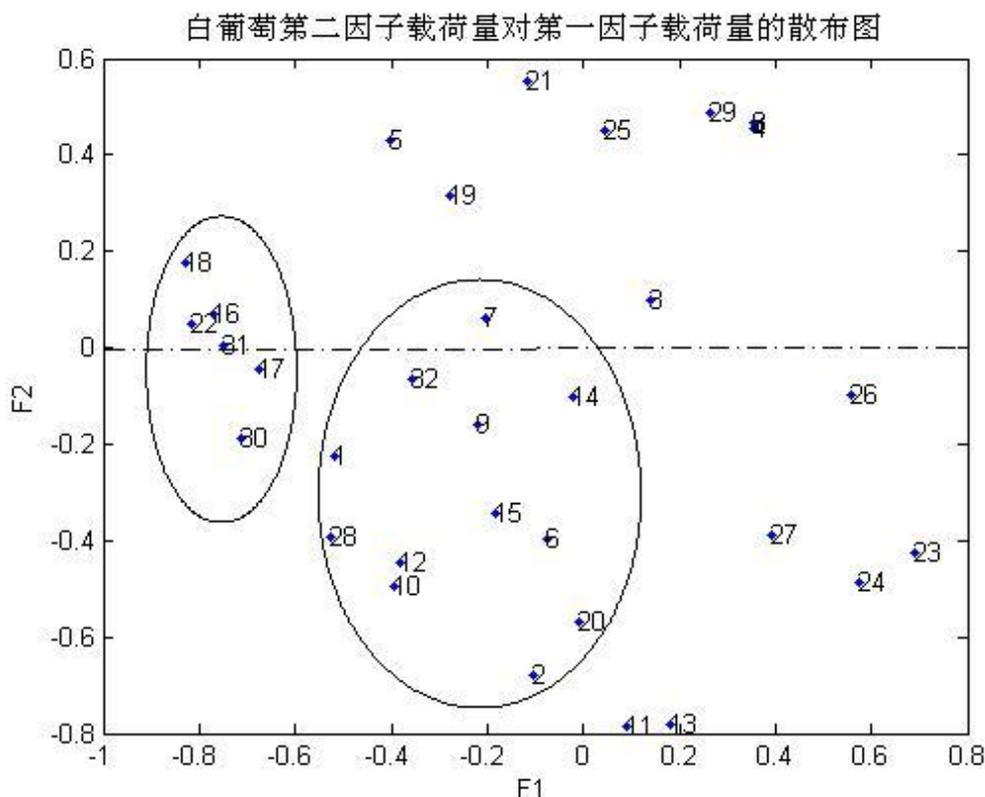


图 8 白葡萄第二因子载荷量对第一因子载荷量的散布图

#### § 4. 1. 3 酿酒葡萄理化指标分类

根据上述主成分分析中红（白）葡萄散布图的特征，我们将红（白）葡萄的理化指标分为三大因子：

**口感因子：**VC 含量（3），酒石酸（5），苹果酸（6），柠檬酸（7），多酚氧化酶活力（8），褐变度（9），总酚（11），单宁（12），黄酮醇（15），总糖（16），还原糖（17），可溶性固形物（18），PH 值（19），可滴定酸（20），固酸比（21），果梗比（25），出汁率（26）；

**外观因子：**花色苷（4），DPPH 自由基（10），果皮质量（27），果皮颜色 L\*值（28），果皮颜色 a\*值（29），果皮颜色 b\*值（30）；

**营养因子：**氨基酸总量（1），蛋白质（2），葡萄总黄酮（13），白藜芦醇（14），干物质含量（22），果穗质量（23），百粒质量（24）。

#### § 4. 1. 4 红（白）葡萄三类因子的结果

由红（白）葡萄各变量的载荷量可以看出，前三个主成分的累计贡献率均已超过 60%，因此只需用三个主成分就能很好地概括每个酿酒葡萄样本的数据。

由最大的三个特征值对应的特征向量（列向量）可以写出 27(28)个红（白）葡萄样品的第一主成分（口感因子）、第二主成分（外观因子）、第三主成分（营养因子）：

$$\text{红葡萄: } P_{ij} = X_j \cdot \alpha_i \quad (i = 1,2,3 \quad j = 1,2, \dots, 27),$$

$$\text{白葡萄: } Q_{ij} = Y_j \cdot \beta_i \quad (i = 1,2,3 \quad j = 1,2, \dots, 28),$$

其中： $X_j$  ( $Y_j$ )表示第  $j$  个红（白）葡萄样本的 32 个变量组成的行向量； $\alpha_i$  ( $\beta_i$ )表示红（白）葡萄相关矩阵的第  $i$  个最大特征值对应的特征向量（列向量）。

红（白）葡萄的三种因子计算出结果如下：

表 11 红葡萄样品的口感因子

红葡萄样品	1	2	3	4	5	6	7
口感因子	0.023	0.0619	0.0641	0.0327	0.0489	0.0208	0.0251
红葡萄样品	8	9	10	11	12	13	14
口感因子	0.0111	0.0576	0.0197	0.0025	0.0446	0.0154	0.0414
红葡萄样品	15	16	17	18	19	20	21
口感因子	0.0176	0.037	0.0543	0.0338	0.0424	0.0673	0.05
红葡萄样品	22	23	24	25	26	27	
口感因子	0.0457	0.0511	0.0392	0.0132	0.0327	0.0468	

表 12 白葡萄样品的口感因子

白葡萄样品	1	2	3	4	5	6	7
口感因子	0.0564	0.0174	0.297	0.605	0.707	0.0297	0.0051
白葡萄样品	8	9	10	11	12	13	14
口感因子	0.010	0.0317	0.0728	0.0018	0.0037	0.0174	0.0399
白葡萄样品	15	16	17	18	19	20	21
口感因子	0.042	0.0014	0.081	0.0543	0.399	0.0276	0.0646
白葡萄样品	22	23	24	25	26	27	28
口感因子	0.0317	0.0317	0.0481	0.0605	0.0035	0.0174	0.0584

表 13 红葡萄样品的外观因子

红葡萄样品	1	2	3	4	5	6	7
外观因子	0.0417	0.0368	0.0391	0.0375	0.0422	0.0312	0.0249
红葡萄样品	8	9	10	11	12	13	14
外观因子	0.0391	0.0433	0.0412	0.0222	0.0312	0.0354	0.0417
红葡萄样品	15	16	17	18	19	20	21
外观因子	0.038	0.038	0.361	0.264	0.407	0.296	0.391
红葡萄样品	22	23	24	25	26	27	
外观因子	0.349	0.454	0.386	0.0396	0.0438	0.0375	

表 14 白葡萄样品的外观因子

白葡萄样品	1	2	3	4	5	6	7
外观因子	0.0342	0.0498	0.0263	0.040	0.044	0.0126	0.0303
白葡萄样品	8	9	10	11	12	13	14
外观因子	0.0322	0.0518	0.0146	0.0146	0.040	0.0185	0.0361
白葡萄样品	15	16	17	18	19	20	21
外观因子	0.0342	0.0070	0.0381	0.0361	0.0263	0.0420	0.0440
白葡萄样品	22	23	24	25	26	27	28
外观因子	0.0498	0.0498	0.0420	0.0440	0.0538	0.0498	0.0518

表 15 红葡萄酒样品的营养因子

红葡萄酒样品	1	2	3	4	5	6	7
营养因子	0.0165	0.0068	0.0095	0.0360	0.0663	0.0349	0.0338
红葡萄酒样品	8	9	10	11	12	13	14
营养因子	0.0384	0.0220	0.0459	0.0147	0.0410	0.0258	0.0213
红葡萄酒样品	15	16	17	18	19	20	21
营养因子	0.0359	0.011	0.0701	0.0488	0.0338	0.0969	0.0098
红葡萄酒样品	22	23	24	25	26	27	
营养因子	0.0135	0.0529	0.0599	0.0652	0.063	0.0261	

表 16 白葡萄酒样品的营养因子

白葡萄酒样品	1	2	3	4	5	6	7
营养因子	0.0718	0.031	0.0246	0.0025	0.0086	0.0305	0.0354
白葡萄酒样品	8	9	10	11	12	13	14
营养因子	0.0795	0.0036	0.118	0.0402	0.0541	0.1068	0.0018
白葡萄酒样品	15	16	17	18	19	20	21
营养因子	0.1606	0.0508	0.0018	0.0129	0.004	0.0108	0.0102
白葡萄酒样品	22	23	24	25	26	27	28
营养因子	0.0027	0.0086	0.0723	0.0126	0.0057	0.0347	0.0038

#### § 4.1.5 基于系统聚类分析的葡萄芳香物质探究

我们利用附表三中给出的红（白）葡萄酒中各类芳香物质的数据（剔除了缺失较多的几种芳香物质，指标编号为 30），并此引入附表一中葡萄酒的香气分数，进一步利用相关矩阵进行  $R$  型系统聚类分析（聚类距离为欧氏距离，聚类方法为类平均法  $UPGMA$ ），最终从每个类中选取代表性的指标。如果有两个样本类和，则定义类与类之间的相似性度量为：

$$D(G_1, G_2) = \frac{1}{n_1 n_2} \sum_{x_i \in G_1} \sum_{x_j \in G_2} d(x_i, x_j)$$

它等于  $G_1, G_2$  中两两样本点距离的平均，式中  $n_1, n_2$  分别为  $G_1, G_2$  中的样本点个数。利用 *matlab* 中的 *linkage* 画出红（白）葡萄酒芳香物质的聚类树形图如下：

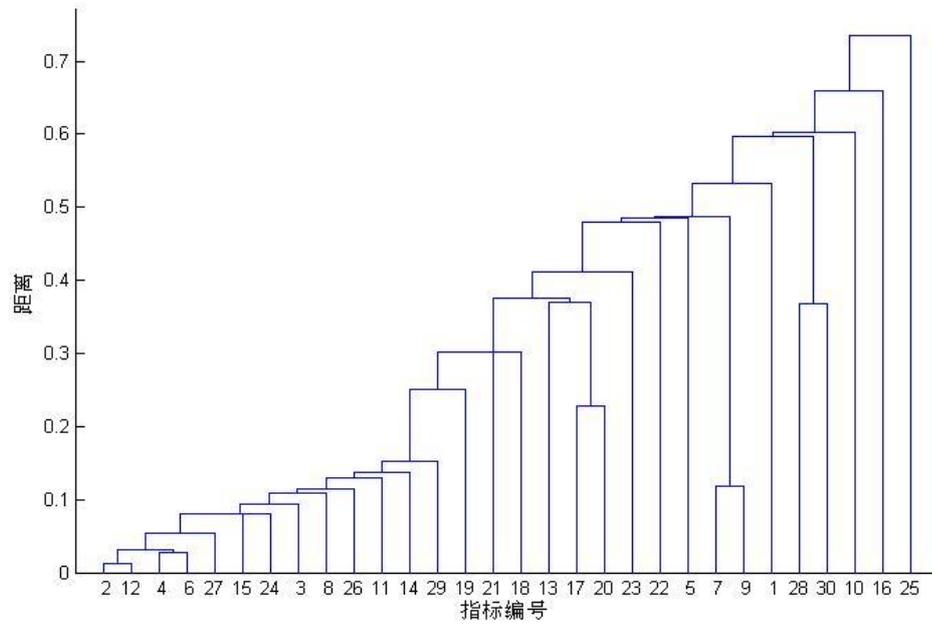


图9 红葡萄芳香物质聚类树形图

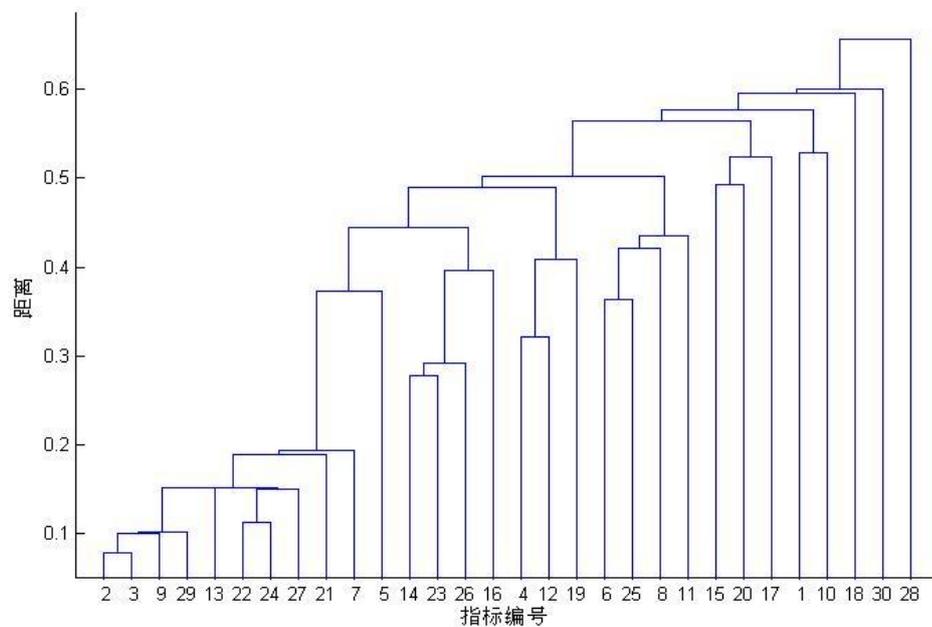


图10 白葡萄芳香物质聚类树形图

**分析：**由聚类树形图可以看出，当最长距离分别为 0.7、0.6 时，红（白）葡萄的均可分为两类，从中分别选取变量 30（香气评分）、25（1-壬醇），和 30（香气评分）、28（(E)-3,7-二甲基-2,6-辛二烯-1-醇）。由此，我们计算出红（白）葡萄的第四类因子——香气因子如下（归一化后）：

表 17 红葡萄酒样品的香气因子

红葡萄酒样品	1	2	3	4	5	6	7
香气因子	0.0341	0.0379	0.0388	0.0386	0.0358	0.0345	0.0345
红葡萄酒样品	8	9	10	11	12	13	14
香气因子	0.0333	0.0437	0.0360	0.0355	0.0340	0.0383	0.0374
红葡萄酒样品	15	16	17	18	19	20	21
香气因子	0.0324	0.0355	0.0400	0.0321	0.0386	0.0427	0.0367
红葡萄酒样品	22	23	24	25	26	27	
香气因子	0.0379	0.0431	0.0377	0.0362	0.0379	0.0369	

表 18 白葡萄酒样品的香气因子

白葡萄酒样品	1	2	3	4	5	6	7
香气因子	0.0359	0.0347	0.0351	0.0337	0.0392	0.0362	0.0356
白葡萄酒样品	8	9	10	11	12	13	14
香气因子	0.0336	0.0402	0.0389	0.0330	0.0354	0.0348	0.0357
白葡萄酒样品	15	16	17	18	19	20	21
香气因子	0.0379	0.0292	0.0369	0.0344	0.0359	0.0359	0.0363
白葡萄酒样品	22	23	24	25	26	27	28
香气因子	0.0391	0.0354	0.0354	0.0369	0.0336	0.0363	0.0368

#### § 4.2 利用模糊综合评价法对葡萄分级

模糊综合评价方法，是应用模糊关系合成的原理，从多个因素(指标)对被评价事物的隶属等级状况进行综合性评判的一种方法。下面进行评价论域的确定和权重向量的选择。

##### § 4.2.1 因素集的确定和权重向量的选取

我们首先定义因素集、评价集、和权重集如下：

因素集： $U = \{u_1, u_2, \dots, u_n\}$  ( $u_i$  表示第  $i$  项指标， $i=1, 2, \dots, n$ )

权重集：对于各因素间的权数分配为  $A = (a_1, a_2, a_3, \dots, a_n)$ ，其中  $a_i \geq 0$  且

$$\sum_{i=1}^n a_i = 1$$

根据前面主成分分析中的指标分类，我们给出实际的因素集  $U$  如下：

$$U = \{\text{口感因子、外观因子、营养因子、香气因子}\}$$

根据附表一中的各项评分权重，我们定义权重集  $A$  如下：

$$A = (0.44, 0.15, 0.11, 0.30)$$

##### § 4.2.2 模糊综合评价模型的建立

(1) 建立模糊关系矩阵

由于每个葡萄样品的四种因子均进行了归一化处理（无量纲化），因此可直

接将模糊关系矩阵的每一列元素直接取为四种因子，得到模糊关系矩阵定义如下：

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}, \quad 0 \leq r_{ij} \leq 1$$

其中  $n=4$  为因素集  $U$  中元素个数， $m$  为红（白）葡萄样品数量。

(2) 建立模糊综合评价模型

选择  $M(\cdot, +)$  为模糊合成算子，建立模糊综合评价模型如下：

$$B = AOR = (a_1, a_2, \dots, a_4)O \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{41} & \cdots & r_{4m} \end{pmatrix},$$

得到的矩阵  $B$  的元素为第  $i$  个红（白）葡萄样品的评分。

(3) 分级标准

最后，根据下面的葡萄分级标准表，将酿酒葡萄样品分级。

表 18 葡萄分级标准表

级别	优秀	优良	合格	不合格	劣质
得分区间	0.085-0.10	0.070-0.085	0.055-0.07	0.040-0.055	0.025-0.04

#### § 4.2.3 最终红（白）葡萄分级结果

通过模糊综合评价模型将红（白）葡萄分级，得到如下结果：

表 19 红葡萄分级结果

级别	优秀	优良	合格	不合格	劣质
红葡萄 编号	9、20、23	2、3、5、 14、17、19、 21、26、	1、4、10、 12、13、16、 22、24、25、 27	6、7、8、15、18	11

表 20 白葡萄分级结果

级别	优秀	优良	合格	不合格	劣质
白葡萄 编号	5、9、17	1、10、14、 15、21、22、 23、25、27、 28	4、7、8、11、 12、13、18、 20、26	2、3、6、19、 24	16

## § 5 葡萄与葡萄酒理化指标的关系模型（问题三）

### § 5.1 对葡萄酒相关指标的分析

#### § 5.1.1 主成分分析

利用 *Matlab* 绘制出红（白）葡萄酒各变量的散布图（第二因子载荷向量对第一因子载荷向量的散布图）如下（程序见附录二）：

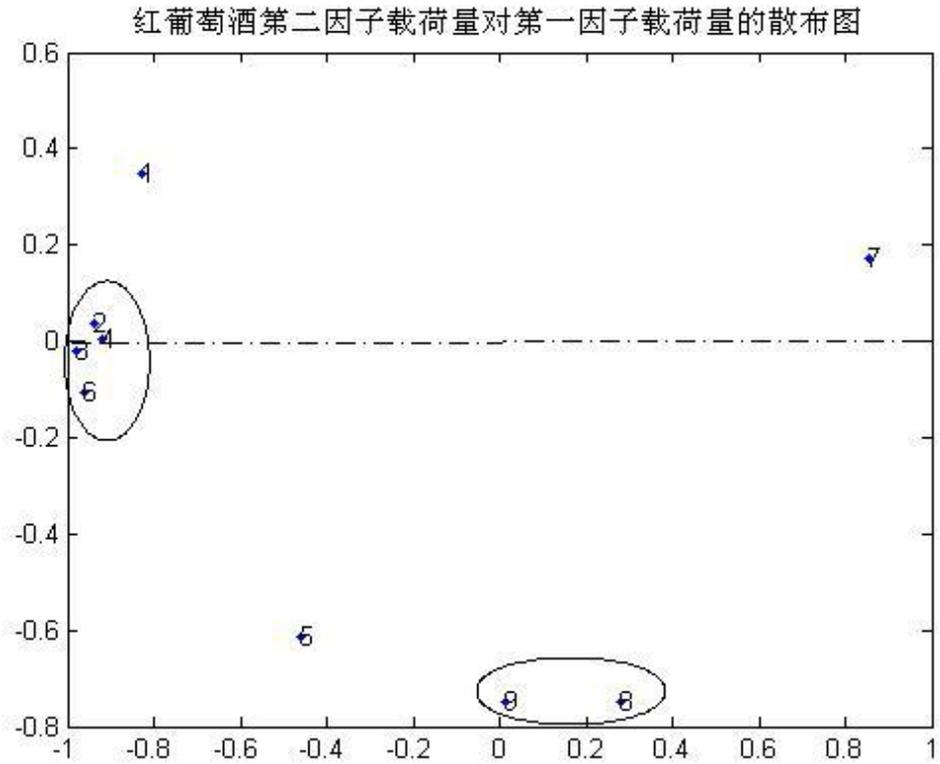


图 11 红葡萄酒第二因子载荷量对第一因子载荷量的散布图

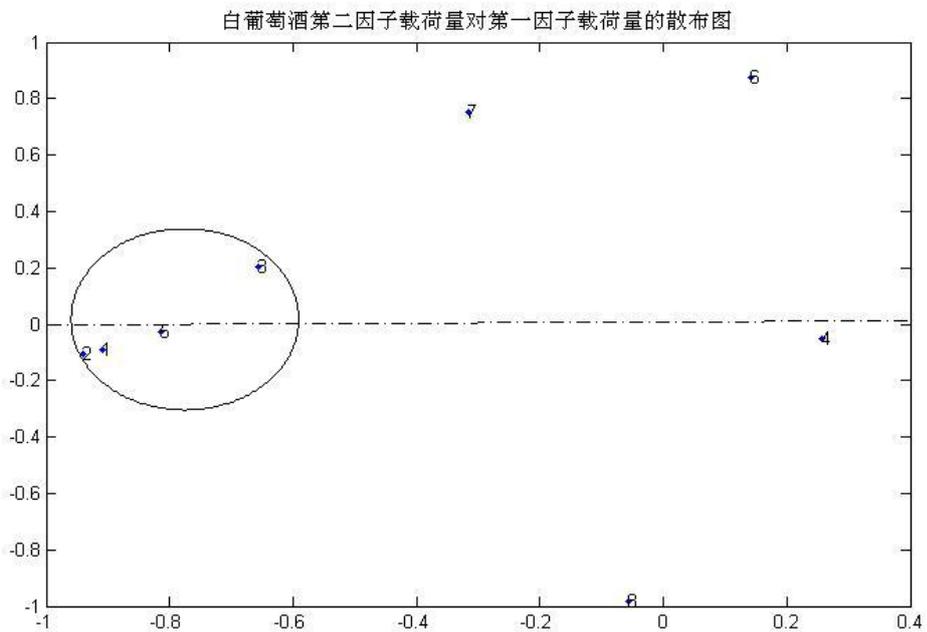


图 12 白葡萄酒第二因子载荷量对第一因子载荷量的散布图

**分析:** 根据上述主成分分析中红(白)葡萄酒散布图的特征,我们将红(白)葡萄酒的理化指标分为三大因子:

**口感因子:** 单宁(2), 总酚(3), 酒总黄酮(4), DPPH 半抑制体积(6)

**外观因子:** 花色苷(1), 果皮颜色 L\*值(7), 果皮颜色 b\*值(8), 果皮颜色 b\*值(9);

营养因子：白藜芦醇（5）。

### § 5. 1. 2 系统聚类分析

对于红（白）葡萄酒中的芳香物质，我们采用系统聚类分析的方法，得到聚类树形图如下（程序见附录二）：

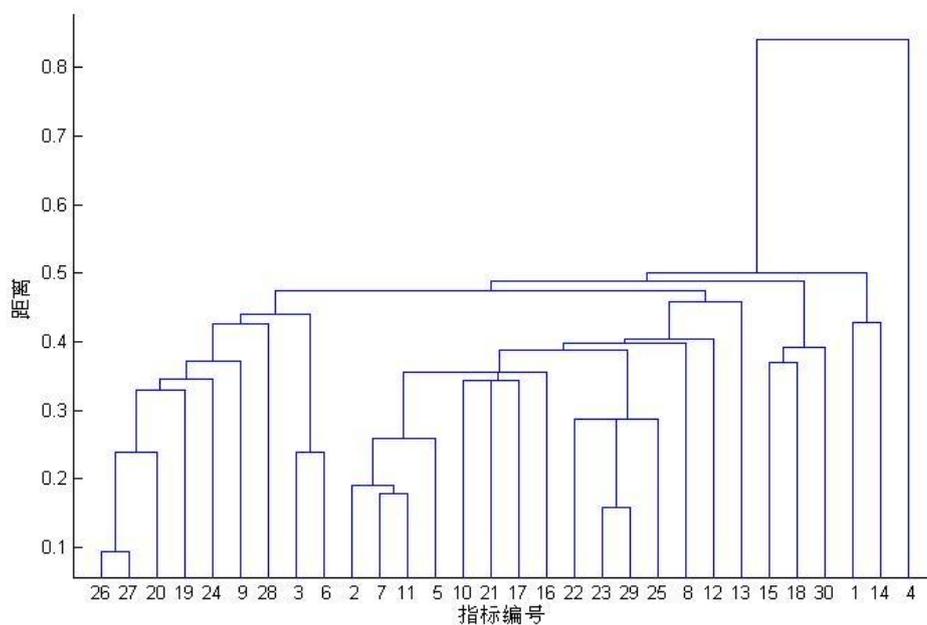


图 13 红葡萄酒芳香物质聚类树形图

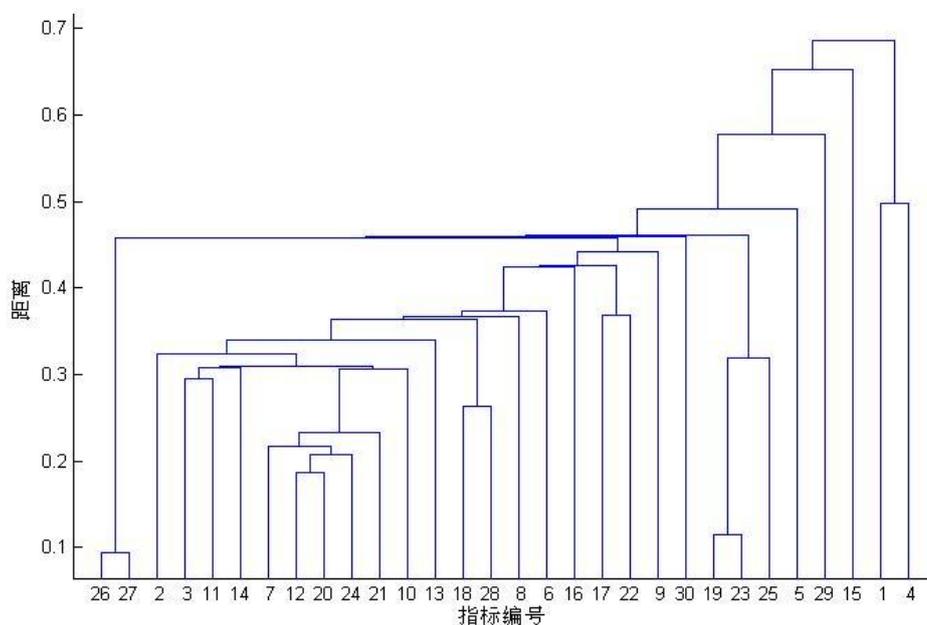


图 14 白葡萄酒芳香物质聚类树形图

**分析：**由此可见红（白）葡萄酒也可以分为四类因子（口感因子、外观因子、营养因子、香气因子），计算出的每种葡萄酒样品的四类因子结果见附录三。

## § 5.2 葡萄酒与葡萄的因子相关性分析

对红（白）葡萄和葡萄酒的因子分别做相关性分析，利用 *matlab* 统计工具箱中的 *CORRCOEF* 函数（程序见附录二），得到相关矩阵  $R_1, R_2$  如下：

$$R_1 = \begin{bmatrix} 1 & 0.212647 & 0.059542 & 0.005166 & 0.816692 & 0.308846 & 0.077827 & 0.619441 \\ 0.212647 & 1 & -0.10802 & 0.330045 & 0.223467 & 0.789025 & -0.1116 & 0.35552 \\ 0.059542 & -0.10802 & 1 & 0.158231 & 0.091508 & 0.035557 & 0.99399 & 0.201418 \\ 0.005166 & 0.330045 & 0.158231 & 1 & 0.140344 & 0.036895 & 0.153894 & 0.405599 \\ 0.816692 & 0.223467 & 0.091508 & 0.140344 & 1 & 0.276216 & 0.0848 & 0.62911 \\ 0.308846 & 0.789025 & 0.035557 & 0.3685 & 0.276216 & 1 & 0.020333 & 0.373246 \\ 0.077827 & -0.1116 & 0.99339 & 0.153894 & 0.0848 & 0.020333 & 1 & 0.209654 \\ 0.619441 & 0.35552 & 0.201418 & 0.405599 & 0.629011 & 0.373246 & 0.209654 & 1 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 1 & 0.98797 & 0.144534 & 0.214148 & -0.12477 & -0.31403 & 0.510986 & 0.087249 \\ 0.98797 & 1 & 0.146934 & 0.201978 & -0.08136 & -0.26649 & 0.526639 & 0.106366 \\ 0.144534 & 0.146934 & 1 & 0.969246 & -0.43373 & -0.32795 & 0.355924 & 0.174009 \\ 0.214148 & 0.201978 & 0.969246 & 1 & -0.43373 & -0.32795 & 0.355924 & 0.174009 \\ -0.12477 & -0.08136 & -0.43373 & -0.44528 & 1 & 0.862206 & -0.04879 & -0.01869 \\ -0.31403 & -0.26649 & -0.32795 & -0.364 & 0.862206 & 1 & -0.21134 & -0.05585 \\ 0.510986 & 0.526639 & 0.355924 & 0.367966 & -0.04879 & -0.21134 & 1 & 0.801818 \\ 0.087249 & 0.106366 & 0.174009 & 0.212576 & -0.01869 & -0.05585 & 0.801818 & 1 \end{bmatrix}$$

## § 5.3 葡萄与葡萄酒的回归方程模型

由上述相关性分析可知，红（白）葡萄和葡萄酒的四个因子（口感因子、外观因子、营养因子、香气因子）对应的相关系数比较大，因此我们利用统计软件 *sas9.1* 对葡萄酒与葡萄的四个因子建立回归方程，外观因子的线性回归分析见图 1（其他因子的回归分析见附录三），回归方程如下：

(1) 红葡萄与红葡萄酒

$$\text{口感因子: } Y_1 = 25.76463P_1 - 4.39770$$

$$\text{外观因子: } Y_2 = 32.83118P_1 - 4.53681$$

营养因子: 不成线性关系

$$\text{香气因子: } Y_4 = 28.38850P_1 - 4.54785$$

(2) 白葡萄与白葡萄酒

$$\text{口感因子: } Z_1 = 44.18843Q_1 - 5.33955$$

$$\text{外观因子: } Z_2 = 35.34291Q_1 - 4.66600$$

营养因子: 不成线性关系

$$\text{香气因子: } Z_4 = 33.35802Q_1 - 4.53671$$

其中,  $Y_i$  表示红葡萄酒第  $i$  个因子的值;  $P_i$  表示红葡萄第  $i$  个因子的值;  $Z_i$  表示白葡萄酒第  $i$  个因子的值;  $Q_i$  表示白葡萄第  $i$  个因子的值。

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 27  
Number of Observations Used 27

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.96869	0.96869	59.34	<.0001
Error	25	0.40812	0.01632		
Corrected Total	26	1.37681			

Root MSE 0.12777 R-Square 0.7036  
Dependent Mean -3.32084 Adj R-Sq 0.6917  
Coeff Var -3.84748

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.53681	0.15976	-28.40	<.0001
x	1	32.83118	4.26205	7.70	<.0001

图 15 红葡萄与红葡萄酒外观因子的线性回归分析

**分析：**在线性函数的回归分析结果的中间部分，可以找到拟合优度检验的判定系数 $R^2 = 0.7036$ ( $R - Square$ )，修正 $R^2 = 6917$ ( $Adj R - Sq$ )，说明拟合程度高；在方差分析表“*Analysis of Variance*”中，对应 $P$ 值(“ $Pr > |F|$ ”)小于 0.0001，所以在 $\alpha = 0.05$ 的条件下非常显著，因此回归方程整体上是显著的；在参数估计表中，可以看到用普通最小二乘法(*OLS*)对回归模型的参数估计结果及对应回归系数显著性检验的 $P$ 值(“ $Pr > |t|$ ”)小于 0.0001(通常不对截距项进行显著性检验)，即在给定的显著性水平条件下非常显著。

## § 6. 葡萄酒的质量评价模型（问题四）

### § 6.1 葡萄酒质量评价模型模型的建立

我们将葡萄酒的总分数 $J(K)$ 作为被解释变量，酿酒葡萄以及葡萄酒的四大因子（口感因子、外观因子、营养因子、香气因子） $P_i(Q_i)$ 、 $Y_i(Z_i)$ 作为解释变量，利用 SAS 进行多元线性回归，得到结果如图 16、17。

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 27  
Number of Observations Used 27

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	409.43429	51.17929	460.86	<.0001
Error	18	1.99979	0.11110		
Corrected Total	26	411.43407			

Root MSE 0.33332 R-Square 0.9951  
Dependent Mean 70.51481 Adj R-Sq 0.9930  
Coeff Var 0.47269

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	37.28589	1.00232	37.20	<.0001
x1	1	113.84352	7.19881	15.81	<.0001
x2	1	218.62321	22.28315	9.81	<.0001
x3	1	8.93353	28.11265	0.32	0.7543
x4	1	555.91670	34.82754	15.96	<.0001
x5	1	-0.26831	7.51410	-0.04	0.9719
x6	1	0.36603	15.85520	0.02	0.9818
x7	1	-5.98105	28.74630	-0.21	0.8375
x8	1	5.74735	13.87343	0.41	0.6836

图 16 红葡萄酒评价模型多元线性回归分析图

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 28  
Number of Observations Used 28

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	265.74309	33.21789	109.99	<.0001
Error	19	5.73798	0.30200		
Corrected Total	27	271.48107			

Root MSE 0.54954 R-Square 0.9789  
Dependent Mean 76.53214 Adj R-Sq 0.9700  
Coef Var 0.71806

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	46.84548	2.76086	16.97	<.0001
x1	1	81.22582	35.59877	2.28	0.0342
x2	1	-10.56198	32.45507	-0.33	0.7484
x3	1	104.62510	48.08186	2.18	0.0424
x4	1	-49.15055	44.61907	-1.10	0.2844
x5	1	-2.83592	6.81197	-0.42	0.6818
x6	1	1.11539	6.82347	0.16	0.8719
x7	1	677.01186	173.45820	3.90	0.0010
x8	1	31.52405	98.53633	0.32	0.7525

图 17 白葡萄酒评价模型多元线性回归分析图

分析图表明拟合优度检验的判定系数为 0.9951 和 0.9789，拟合度较高，因此建立模型如下：

(1) 红葡萄酒质量评价模型模型

$$J = 113.84352P_1 + 218.62321P_2 + 8.93353P_3 + 555.91670P_4 - 0.26831Y_1 - 0.36603Y_2 - 5.98105Y_3 + 5.74735Y_4 + 37.28589$$

(2) 白葡萄酒质量评价模型模型

$$K = 81.22582Q_1 - 10.56198Q_2 + 104.62510Q_3 - 49.15055Q_4 - 2.83592Z_1 + 1.11539Z_2 - 677.01186Z_3 + 31.52405Z_4 + 46.84548$$

## § 6.2 葡萄酒质量评价模型的检验

我们利用 $\chi^2$ 检验来验证评价模型的正确性：

$$\chi = \sum_{i=1}^n \frac{(x_i - x_i^*)^2}{x_i^*}$$

其中 $x_i$ 为实际值， $x_i^*$ 为理论值。若 $\chi < \chi_n(0.05)$ ，则实际值与模型得出的理论值无显著性差异。利用 7.1 中模型，计算葡萄酒样品的理论分数值及 $\chi$ 值，如下表所示：

表 21 红葡萄酒质量评价模型的卡方检验

红葡萄酒 质量分数	实际值	评价模型值
	68.1	67.9
	74	71.3
	74.6	74.9
	...	...
	72	71.5
71.5	72.3	
$\chi$	33.8	
$\chi_{27}(0.05)$	40.1	

表 22 白葡萄酒质量评价模型的卡方检验

白葡萄酒 质量分数	实际值	评价模型值
	77.9	77.6
	75.8	76.2
	75.6	75.3
	...	...
	77	76.5
79.6	80.1	
$\chi$	36.6	
$\chi_{28}(0.05)$	41.3	

利用 matlab 绘制出图形如图 18、19。

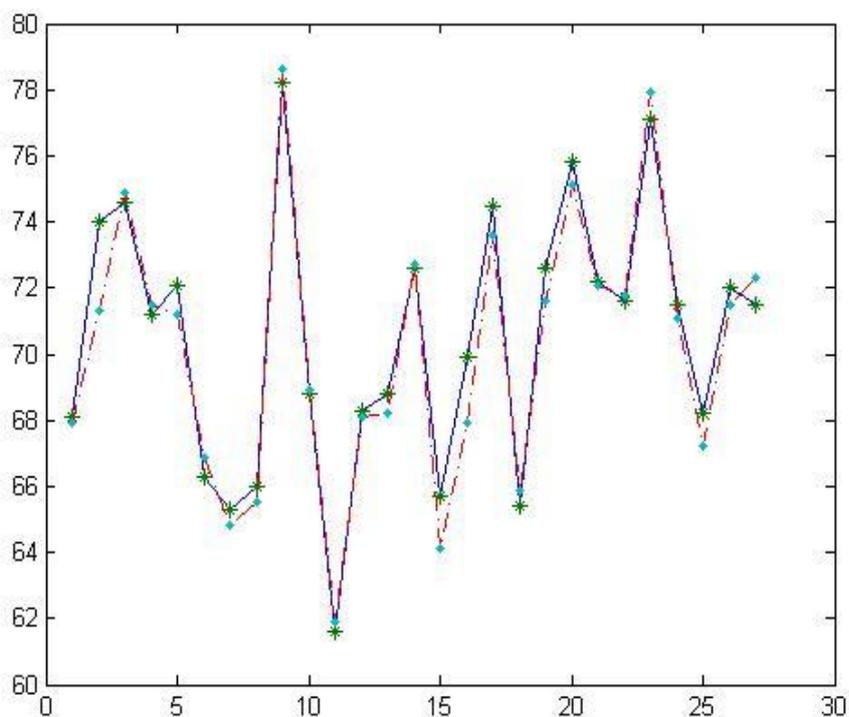


图 18 红葡萄酒质量评价模型分数与实际分数

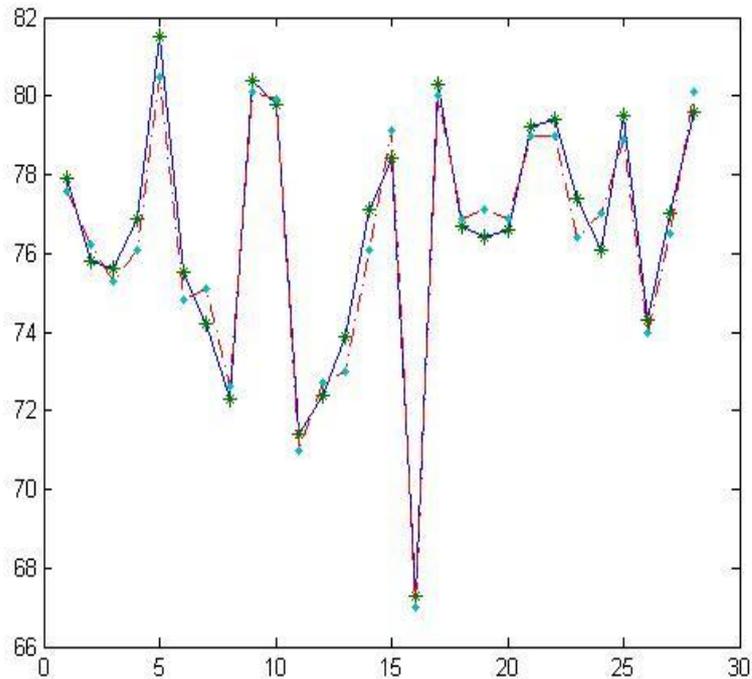


图 19 白葡萄酒质量评价模型分数与实际分数

**结论：**由上表可知，红（白）葡萄酒质量评价模型的计算分数与实际分数无显著差异，因此可以用葡萄和葡萄酒的理化指标来评价葡萄酒的质量。

## § 7. 模型评价与推广

### § 7.1 模型评价

**优点：**建立模型时融合了主观因素和客观因素，使结果更加准确、全面。此外，在处理数据时，将庞大的数据根据其相关性进行归类简化，使处理方便。

**缺点：**第一，模糊数学模型中模糊集的隶属函数主要根据人们的实践经验来确定，具有一定的主观性。第二，方差分析模型假设了总体呈正态分布且方差相等，由于酒样品数量比较少，所以并不一定趋向于正态分布，且没有进行方差齐性检验，需要进一步说明。

### § 7.2 模型推广

- 1、主成分分析模型还可以用于人口统计，数量地理学，分子动力学模拟等方面。
- 2、模糊数学模型还可以在经济管理、社会管理等方面运用。

## 参考文献

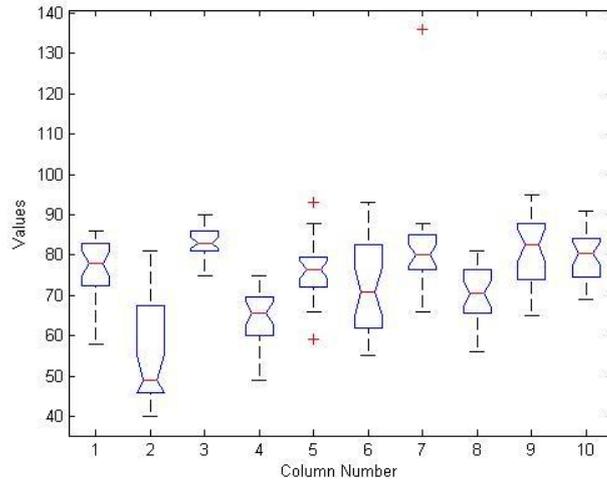
- [1] 吴有炜. 试验设计与数据处理[M]. 苏州: 苏州大学出版社, 2002.
- [2] 王栋, 等. 食品感官评价原理与技术[M]. 中国轻工业出版社, 1997. 402-426.
- [3] 姜起源, 谢金星, 叶俊, 数学模型 (第三版)》, 北京: 高等教育出版社, 2003 年.
- [4] 周璐, 陈渝等译, 数值方法(Matlab 版) [M], 北京: 电子工业出版社, 2011. 460-468.
- [5] 李运, 李记明, 姜忠军统计分析在葡萄酒质量评价中的应用[J] 酿酒科技, 2009, (4): 79-82.
- [6] 李华. 葡萄酒品尝学[M]. 北京: 科学出版社, 2006. 97~106

附录一

方 差 分 析 表

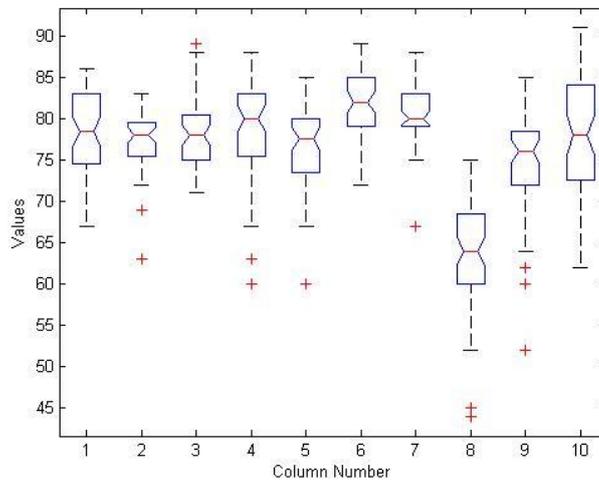
第一组白葡萄方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	18023.9	9	2002.66	24.76	0
Error	21836	270	80.87		
Total	39860	279			



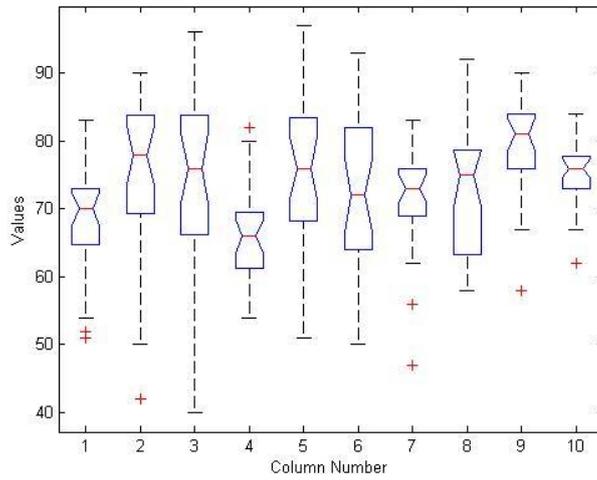
第二组白葡萄方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	13450.2	9	1494.47	40.66	0
Error	20213.2	550	36.75		
Total	33663.4	559			



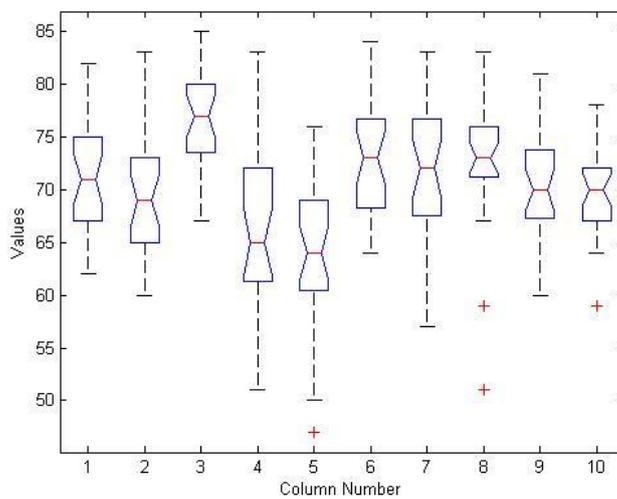
第一组红葡萄方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	3172.5	9	352.504	3.65	0.0003
Error	25089.6	260	96.499		
Total	28262.2	269			



第二组红葡萄方差分析表

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	3060.8	9	340.086	9.54	1.44895e-012
Error	9264.7	260	35.633		
Total	12325.4	269			



## 附录二

主成分分析

```
%cwfac.m
function result=cwfac(vector)
fprintf(' 相关系数矩阵:\n')
std=CORRCOEF(vector) %计算相关系数矩阵
fprintf(' 特征向量(vec)及特征值(val): \n')
[vec, val]=eig(std) %求特征值(val)及特征向量(vec)
newval=diag(val) ;
[y, i]=sort(newval) ; %对特征根进行排序, y 为排序结果, i 为索引
fprintf(' 特征根排序: \n')
for z=1:length(y)
    newy(z)=y(length(y)+1-z);
end
fprintf(' %g\n', newy)
rate=y/sum(y) ;
fprintf(' \n 贡献率: \n')
newrate=newy/sum(newy)
sumrate=0;
newi=[];
for k=length(y):-1:1
    sumrate=sumrate+rate(k);
    newi(length(y)+1-k)=i(k);
end
fprintf(' 主成分数: %g\n\n', length(newi));
fprintf(' 主成分载荷: \n')
for p=1:length(newi)
    for q=1:length(y)
        result(q, p)=sqrt(newval(newi(p)))*vec(q, newi(p));
    end
end
end %计算载荷
disp(result)
```

红葡萄第二因子载荷量对第一因子载荷量的散布图

%picture.m

```
s1=[0.307560164 0.671985726 -0.093865021 0.83987841 0.363108706 0.278808884
0.235411349 0.278059927 0.564427253 0.816544956 0.899232905 0.783413813
0.779276059 -0.021869785 0.544419765 0.151036566 -0.025210927
0.157345202 0.326060682 -0.321744988 0.379614402 0.250780045
-0.216856813 -0.417742689 0.569304535 0.578686866 -0.141219349
-0.52392513 -0.456985882 -0.282611681 0.61192535 0.452073125];
s2=[-0.583046712 0.352063164 0.399084383 -0.033589053 -0.185185828
-0.405987312 -0.25949801 -0.120315376 -0.028809327 0.314236193
0.032373983 0.054134066 0.170593458 -0.055491453 -0.111117839
```

```

-0.796440492    -0.756695638    -0.746453989    0.197929794 -0.318684438
-0.066221177    -0.880846681    0.557250382 0.488024861 0.078033653
0.10128492  0.373461569 0.387656901 -0.312152417    -0.524548481
0.59863396  0.034996106];
plot(s1,s2,'b.')
```

**%散点图**

```

for i=1:32
    text(s1(1,i),s2(1,i),num2str(i));
end
```

**%%%%%%%%%%%%白葡萄第二因子载荷量对第一因子载荷量的散布图%%%%%%%%**

**%picture.m**

```

s1=[-0.516400425    -0.104260087    0.13946 0.35697203    -0.401007509
-0.076453426    -0.204767746    0.356619014 -0.221050721    -0.393738847
0.091742456 -0.380708314    0.180665994 -0.022345362    -0.180901734
-0.769352017    -0.676743247    -0.828874408    -0.27586784 -0.008191632
-0.118518921    -0.817381446    0.69022 0.574169426 0.043854263 0.558020913
0.391294132 -0.525423453    0.263329076 -0.711845253    -0.751569005
-0.357132281];
s2=[-0.225460086    -0.680149519    0.098896607 0.454925722 0.431596976
-0.395634401    0.060882807 0.467840171 -0.159804066    -0.496416809
-0.78754938 -0.44727258 -0.783219704    -0.103854591    -0.344966914
0.067918633 -0.04558032 0.177318275 0.317329069 -0.569777194    0.554081887
0.047524644 -0.423720072    -0.487130576    0.449452181 -0.097565508
-0.388479973    -0.392005912    0.486707508 -0.188454408    0.005903579
-0.066388733];
plot(s1,s2,'b.')
```

**%散点图**

```

for i=1:32
    text(s1(1,i),s2(1,i),num2str(i));
end
```

**%%%%%%%%%%%%R 型聚类分析%%%%%%%%**

```

load gj.txt    %把原始数据保存在纯文本文件 gj.txt 中
r=corrcoef(gj); %计算相关系数矩阵
d=tril(r);    %取出相关系数矩阵的下三角元素
for i=1:10    %对角线元素化成零
    d(i,i)=0;
end
d=d(:);
d=nonzeros(d); %取出非零元素
d=d';d=1-d;
z=linkage(d)
dendrogram(z)
```

一元线性回归

```
data a;
input x z;
y=log(z);
cards;
0.022970221 0.0234
0.061911428 0.0563
0.064074828 0.0567
0.032705523 0.0345
0.048931026 0.0438
0.020806821 0.0349
0.025133622 0.0129
0.011071519 0.0078
0.057584627 0.0589
0.019725121 0.0178
0.002545177 0.0179
0.044604225 0.0457
0.01539832 0.0129
0.041359124 0.0583
0.017561721 0.0219
0.037032324 0.0389
0.054339527 0.0478
0.033787223 0.038
0.042440825 0.0489
0.067319929 0.0589
0.050012726 0.0508
0.045685925 0.0397
0.051094426 0.0677
0.039195724 0.0389
0.01323492 0.0178
0.032705523 0.0389
0.046767625 0.01
```

```
;
proc reg;
model y=x;
run;
```

第四问的多元线性回归

```
data pcontent;
input x1-x8 y;
cards;
0.056351821 0.0577 0.034175863 0.0362 0.071800406 0.0432 0.035877863 0.0327
77.9
0.017393167 0.0132 0.049839801 0.0526 0.031008303 0.023 0.034656489 0.0367
75.8
```

0.0296959 0.028 0.026343895 0.0238 0.024616085 0.0234 0.035114504 0.0362  
 75.6  
 0.060452732 0.0629 0.04004984 0.046 0.002503728 0.0026 0.033740458 0.0326  
 76.9  
 0.07070501 0.0687 0.043965824 0.052 0.008648377 0.0078 0.039236641 0.0398  
 81.5  
 0.0296959 0.0274 0.012637949 0.0162 0.030520543 0.0324 0.036183206 0.0387  
 75.5  
 0.005090434 0.0064 0.030259879 0.0294 0.03542381 0.0376 0.035572519 0.0372  
 74.2  
 0.001020127 0.0017 0.032217871 0.0328 0.079527544 0.0829 0.033587786 0.0352  
 72.3  
 0.031746356 0.0301 0.049797793 0.0475 0.003605368 0.0036 0.040152672 0.0423  
 80.4  
 0.072755465 0.0732 0.014595942 0.0183 0.117957864 0.0124 0.038931298 0.0368  
 79.8  
 0.001785222 0.0029 0.013595942 0.0126 0.040224391 0.0398 0.032977099 0.034  
 71.4  
 0.003723464 0.0036 0.04004984 0.042 0.054087032 0.0527 0.035419847 0.0373  
 72.4  
 0.017393167 0.0128 0.016511926 0.0183 0.10684208 0.12 0.03480916 0.0342  
 73.9  
 0.039948178 0.0478 0.036133855 0.0357 0.001845414 0.0014 0.035725191 0.0358  
 77.1  
 0.041998633 0.056 0.034175863 0.0311 0.160572648 0.1638 0.037862595 0.0392  
 78.4  
 0.001428178 0.0021 0.006941972 0.0069 0.050775401 0.0538 0.029160305 0.027  
 67.3  
 0.080957287 0.0834 0.030091848 0.0321 0.001802684 0.0016 0.036946565 0.0352  
 80.3  
 0.054301366 0.0528 0.036133855 0.0324 0.012858512 0.0153 0.034351145 0.0311  
 76.7  
 0.039948178 0.0372 0.026343895 0.0286 0.003976154 0.004 0.035877863 0.0372  
 76.4  
 0.027645445 0.0271 0.042003832 0.0502 0.010753444 0.0145 0.035877863 0.0382  
 76.6  
 0.064553643 0.0619 0.043165824 0.0462 0.01018867 0.0121 0.036335878 0.0356  
 79.2  
 0.031746356 0.0361 0.049839801 0.0498 0.002704026 0.0032 0.039083969 0.0392  
 79.4  
 0.031746356 0.0341 0.049839801 0.0478 0.008597034 0.0083 0.035419847 0.0339  
 77.4  
 0.048149999 0.0439 0.042007832 0.045 0.072339508 0.0765 0.033435115 0.031  
 76.1

```

0.060452732 0.0628 0.043965824 0.04 0.012627468 0.0143 0.036946565 0.0349
79.5
0.003519439 0.0042 0.053755785 0.0527 0.0057085 0.0053 0.033587786 0.0318
74.3
0.017393167 0.0183 0.049839801 0.0453 0.034679335 0.0327 0.036335878 0.0343
77
0.058402277 0.0559 0.051717793 0.0528 0.003805666 0.0039 0.036793893 0.0329
79.6
;
proc reg data=pcontent; /*reg 语句调用回归分析*/
model y=x1-x8; /*model 因变量=自变量;*/
run;

```

### 附录三

%%%%%%%%%%红葡萄酒与红葡萄酒四个因子之间的关系 sas 图%%%%%%%%%%

SAS 系统		2012年09月10日 星期一 上午			
The REG Procedure					
Model: MODEL1					
Dependent Variable: y					
		Number of Observations Read		27	
		Number of Observations Used		27	
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.31311	5.31311	31.28	<.0001
Error	25	4.24629	0.16985		
Corrected Total	26	9.55940			
Root MSE		0.41213	R-Square	0.5558	
Dependent Mean		-3.44345	Adj R-Sq	0.5380	
Coeff Var		-11.96854			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.39770	0.18815	-23.37	<.0001
x	1	25.76463	4.60664	5.59	<.0001

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 27  
Number of Observations Used 27

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.96869	0.96869	59.34	<.0001
Error	25	0.40812	0.01632		
Corrected Total	26	1.37681			

Root MSE 0.12777 R-Square 0.7036  
Dependent Mean -3.32084 Adj R-Sq 0.6917  
Coeff Var -3.84748

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.53681	0.15976	-28.40	<.0001
x	1	32.83118	4.26205	7.70	<.0001

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 27  
Number of Observations Used 27

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.11321	0.11321	4.67	0.0405
Error	25	0.60631	0.02425		
Corrected Total	26	0.71952			

Root MSE 0.15573 R-Square 0.1573  
Dependent Mean -3.30962 Adj R-Sq 0.1236  
Coeff Var -4.70542

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.12579	0.37894	-10.89	<.0001
x	1	22.03646	10.19945	2.16	0.0405

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 27  
Number of Observations Used 27

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	10.73565	10.73565	185.29	<.0001
Error	25	1.44847	0.05794		
Corrected Total	26	12.18413			

Root MSE 0.24071 R-Square 0.8811  
Dependent Mean -3.49643 Adj R-Sq 0.8764  
Coeff Var -6.88431

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.54785	0.09007	-50.49	<.0001
x	1	28.38850	2.08552	13.61	<.0001

%%%%%%%%%%白葡萄与白葡萄酒四个因子之间的关系 sas 图%%%%%%%%%%

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 28  
Number of Observations Used 28

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31.11507	31.11507	126.77	<.0001
Error	26	6.38179	0.24545		
Corrected Total	27	37.49686			

Root MSE 0.49543 R-Square 0.8298  
Dependent Mean -3.78139 Adj R-Sq 0.8233  
Coeff Var -13.17152

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-5.33955	0.16856	-31.68	<.0001
x	1	44.18843	3.92471	11.26	<.0001

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 28  
Number of Observations Used 28

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.15204	0.15204	48.10	<.0001
Error	26	0.08219	0.00316		
Corrected Total	27	0.23423			

Root MSE 0.05622 R-Square 0.6491  
Dependent Mean -3.34535 Adj R-Sq 0.6356  
Coeff Var -1.68067

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-4.53671	0.17212	-26.36	<.0001
x	1	33.35802	4.81004	6.94	<.0001

## 白葡萄酒样品各因子

样品	口感因子	外观因子	营养因子	香气因子
1	0.0577	0.0362	0.0432	0.0327
2	0.0132	0.0526	0.023	0.0367
3	0.028	0.0238	0.0234	0.0362
4	0.0629	0.046	0.0026	0.0326
5	0.0687	0.052	0.0078	0.0398
6	0.0274	0.0162	0.0324	0.0387
7	0.0064	0.0294	0.0376	0.0372
8	0.0017	0.0328	0.0829	0.0352
9	0.0301	0.0475	0.0036	0.0423
10	0.0732	0.0183	0.0124	0.0368
11	0.0029	0.0126	0.0398	0.034
12	0.0036	0.042	0.0527	0.0373
13	0.0128	0.0183	0.12	0.0342
14	0.0478	0.0357	0.0014	0.0358
15	0.056	0.0311	0.1638	0.0392
16	0.0021	0.0069	0.0538	0.027
17	0.0834	0.0321	0.0016	0.0352
18	0.0528	0.0324	0.0153	0.0311
19	0.0372	0.0286	0.004	0.0372
20	0.0271	0.0502	0.0145	0.0382
21	0.0619	0.0462	0.0121	0.0356

22	0.0361	0.0498	0.0032	0.0392
23	0.0341	0.0478	0.0083	0.0339
24	0.0439	0.045	0.0765	0.031
25	0.0628	0.04	0.0143	0.0349
26	0.0042	0.0527	0.0053	0.0318
27	0.0183	0.0453	0.0327	0.0343
28	0.0559	0.0528	0.0039	0.0329

红葡萄酒样品各因子

样品	口感	外观	其他	香气
1	0.0234	0.0348	0.0187	0.0329
2	0.0563	0.0398	0.0075	0.0335
3	0.0567	0.0378	0.0108	0.0328
4	0.0345	0.0325	0.0356	0.0419
5	0.0438	0.0389	0.072	0.0378
6	0.0349	0.0298	0.0307	0.0328
7	0.0129	0.0209	0.0376	0.0318
8	0.0078	0.0432	0.0428	0.0378
9	0.0589	0.0483	0.0241	0.0419
10	0.0178	0.0372	0.0421	0.0291
11	0.0179	0.0256	0.0129	0.0389
12	0.0457	0.0279	0.0397	0.0291
13	0.0129	0.0378	0.0276	0.0427
14	0.0583	0.0532	0.0197	0.0392
15	0.0219	0.0372	0.0378	0.0328
16	0.0389	0.0352	0.0101	0.0428
17	0.0478	0.0367	0.0684	0.0326
18	0.038	0.0278	0.0467	0.0378
19	0.0489	0.037	0.0307	0.031
20	0.0589	0.0256	0.0943	0.0483
21	0.0508	0.0367	0.0105	0.0319
22	0.0397	0.0345	0.0153	0.0319
23	0.0677	0.0502	0.0527	0.0493
24	0.0389	0.0327	0.0628	0.0328
25	0.0178	0.0361	0.063	0.0409
26	0.0389	0.0506	0.0597	0.0316
27	0.01	0.052	0.0262	0.0541