

## 摘要

针对问题 1: 评酒员评价是否有显著性差异问题, 我们建立了双样本假设检验模型, 在验证评酒员的打分服从正态分布后, 对所有酒类分别使用  $t$  检验法分析, 得出了两组评酒员分数的均值并无显著性差异。对于评酒员的可信度问题, 我们将其转化为比较评酒员评分的一致性与稳定性问题, 通过对比两组评酒员评分的标准差, 得到第二组评酒员的评分标准差更小, 即波动更小, 故第二组品酒员的评分更可信。

针对问题 2: 葡萄分级问题, 我们建立了多元统计分析模型, 首先使用主成分分析法将葡萄样品的理化指标进行降维, 分别得到酿酒红葡萄的 8 个主成分与酿酒白葡萄的 10 个主成分, 结合各主成分的贡献率, 对酿酒葡萄进行量化评分。其次使用系统聚类法, 根据葡萄的评分将类别分为好、较好、一般、较差四个等级, 并结合葡萄酒质量, 以好葡萄酿出好葡萄酒的原则, 对分级结果的真实性与可靠性进行判定。

针对问题 3: 葡萄与葡萄酒理化指标之间的联系, 建立了多元统计分析模型, 采用典型性关联分析法, 对第二问中提取出的葡萄主成分与附录中所给葡萄酒的理化指标进行关联性分析, 定性地得出了:

- (1) 红葡萄与红葡萄酒理化指标间的 3 组联系;
- (2) 白葡萄与白葡萄酒理化指标间的 2 组联系。

针对问题 4: 我们考虑芳香物质也是影响葡萄酒质量的因素, 结合葡萄与葡萄酒的理化指标, 建立多元回归分析模型, 分别将以上三个因素与葡萄酒评分进行拟合, 分析各因素与葡萄酒的拟合程度, 得出在白葡萄酒中芳香物质对评分的影响最大, 同时在红葡萄酒中, 芳香物质对评分的影响也排在第二位, 由此得出芳香物质对葡萄酒的评分有着不可忽视的影响。因此仅用葡萄和葡萄酒的理化指标来评价葡萄酒的质量是不合理的。

关键字: 假设检验模型 主成分分析 系统聚类 典型性关联分析 多元线性回归

## 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，并论证能否用葡萄酒和葡萄酒的理化指标来评价葡萄酒的质量？

## 2 问题分析

### 2.1 问题一的分析

一方面，由于问题一需要判断两组评酒师的评酒结果是否存在明显差异，因此非常适合使用假设性检验。由于两组评酒师的评分都是根据自身感觉随机给出的，因此评酒师对各个葡萄酒样本的打分都是相互独立的。只需验证每个样本近似满足正态分布，就可以使用 T 检验法对两组评分的差异性做出评价。而在红葡萄酒 27 个样本和白葡萄酒 28 个样本中，若绝大多数样本均没有显著性差异，则可表明两组评酒师的评价结果是相似的，反之则存在差异。

另一方面，评酒更可信的一组评酒师对同一样品的评价必然更加接近，存在的差异也更小，因此对同一样品打分也与此样品所得均分更加接近。只需比较两组评酒师评分的标准差，标准差越小，则说明越可信。

### 2.2 问题二的分析

由于酿酒葡萄的好坏与所酿葡萄酒的质量有直接的关系，故先分析酿酒葡萄的理化指标，我们选取了 28 个酿酒葡萄的一级理化指标，同时为了简化问题，使用主成分分析法，对 28 个指标降维归并得到若干主成份。根据不同葡萄酒各主成份贡献度的差异对葡萄酒进行量化评分，随后使用系统聚类法对葡萄酒进行分级。

### 2.3 问题三的分析

从数据中可以看到，酿酒葡萄与葡萄酒的理化指标个数并不相同，因此并不能简单的采用成对数据的相关性的比较。为此本文采用典型性相关分析，将酿酒葡萄理化指标与葡萄酒理化指标两组变量，采用类似于主成分分析的方法，在其中分别选取若干有代表性的变量组成有代表性的综合指标，通过研究这两组综合指标之间的相关关系，来代替这两组变量间的相关关系。从而，即可得到酿酒葡萄与葡萄酒的理化指标之间联系的大小。

### 2.4 问题四的分析

对于葡萄酒来说，品质的好坏不单由原料即葡萄的品质决定，气味给人们感官上带来的直接刺激也是影响因素之一。为了找出各因素对于葡萄酒品质的影响大小，利用葡萄与葡萄酒的主要理化指标以及附录中所给出的芳香物质对与第一问中

得到的葡萄酒评分进行拟合，通过拟合的程度分析各因素对葡萄酒好坏的影响情况。

3 模型假设

- 1. 评酒员的评分是客观的，不受其他因素影响；
- 2. 评酒员都是有一定资质的，评酒员的分组是随机的；
- 3. 葡萄是新鲜干净的，不含其他杂质，葡萄酒也是纯净的；
- 4. 测量酿酒葡萄与葡萄酒理化指标时的操作是标准的，没有操作上的失误；
- 5. 忽略测量过程中人为因素造成的误差。

4 符号说明

符号	意义
$X_i$	酿酒葡萄的第 $i$ 个一级理化指标
$Y_i$	葡萄酒的第 $i$ 个一级理化指标
$F_k$	第 $k$ 个主成分
$v_{ij}^n$	第 $j$ 个评酒师对第 $n$ 个样本的第 $i$ 个指标的评分
$W_i$	评酒师对第 $i$ 个葡萄酒酒样品评价的平均分
$\delta_{ij}^n$	第 $j$ 个评酒师对第 $n$ 个样本的第 $i$ 个指标的评分相对与平均值的偏差
$s_{ij}^2$	第 $j$ 号评酒师对第 $i$ 个指标的评分偏差的方差
$f_i$	酿酒葡萄第 $i$ 个样本的评分
$U_i$	酿酒葡萄第 $i$ 个指标变量
$V_i$	葡萄酒第 $i$ 个指标变量

5 问题 1 的模型建立与求解

5.1 模型的建立

T 检验法是根据 t 分布理论来求出太差异发生的概率，从而判断两个均值的差异是否显著。由于检验红葡萄酒与检验白葡萄酒的方法和模型是一样的，我们以红葡萄酒为例，给出检验红葡萄酒的模型。

5.1.1 正态分布的检验

由于使用 t 分布的前提是数据服从正态分布。我们利用 Excel 算出两组评酒员对于各红葡萄酒的评分均值如下表所示：

表 1 各红葡萄酒评分均值

红葡萄酒样品	1	2	3	4	5	6	7
第一组	62.7	80.3	80.4	68.6	73.3	72.2	71.5
第二组	68.1	74	74.6	71.2	72.1	66.3	65.3
红葡萄酒样品	8	9	10	11	12	13	14
第一组	72.3	81.5	74.2	70.1	53.9	74.6	73
第二组	66	78.2	68.8	61.6	68.3	68.8	72.6
红葡萄酒样品	15	16	17	18	19	20	21
第一组	58.7	74.9	79.3	60.1	78.6	79.1	77.1
第二组	65.7	69.9	74.5	65.4	72.6	75.8	72.2
红葡萄酒样品	22	23	24	25	26	27	
第一组	77.2	85.6	78	69.2	73.8	73	
第二组	71.6	77.1	71.5	68.2	72	71.5	

利用 SPSS 对两组红葡萄酒评价的均值进行正态分布检验，置信度为95%，并绘制出正态分布你和曲线图。

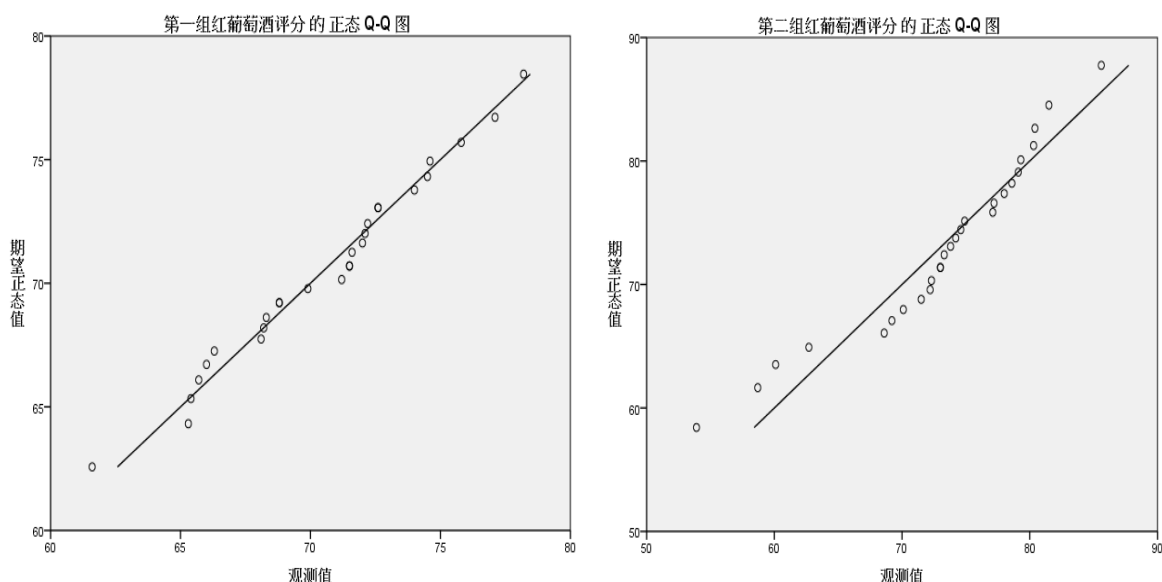


图 1 正态分布拟合曲线图

从图中我们可以看出评分分布曲线近似为一条直线，因此我们认为评酒员对于红葡萄酒以及白葡萄酒的评分均值都服从正态分布。

### 5. 1. 2 假设检验模型的建立与求解

$$\text{设 } T = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{n_1 s_{1n_1}^2 + n_2 s_{2n_2}^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2) \quad (1)$$

对于给定的显著性水平  $\alpha = 0.05$  我们利用 MatLab 软件进行求解。得到如下结果：

(1) 27 个红葡萄酒样本中有 20 个样本的显著性水平小于 0.05，说明两组评酒师对这 20 个酒样的评价结果没有显著性差异，从而可以表明两组评酒师对红葡萄酒的评价结果没有明显差别。

(2) 28 个白葡萄酒中有 22 个样本的显著性水平小于 0.05，说明两组评酒

师对这 22 个酒样的评价结果没有显著性差异，从而表明两组评酒师对白葡萄酒的评价结果没有明显差别。

综上，可以得到两组评酒师的评价结果无显著性差异。

5. 1. 3 评分结果的可信度分析

分别计算两组评酒员对于葡萄酒评分的方差并比较方差大小。

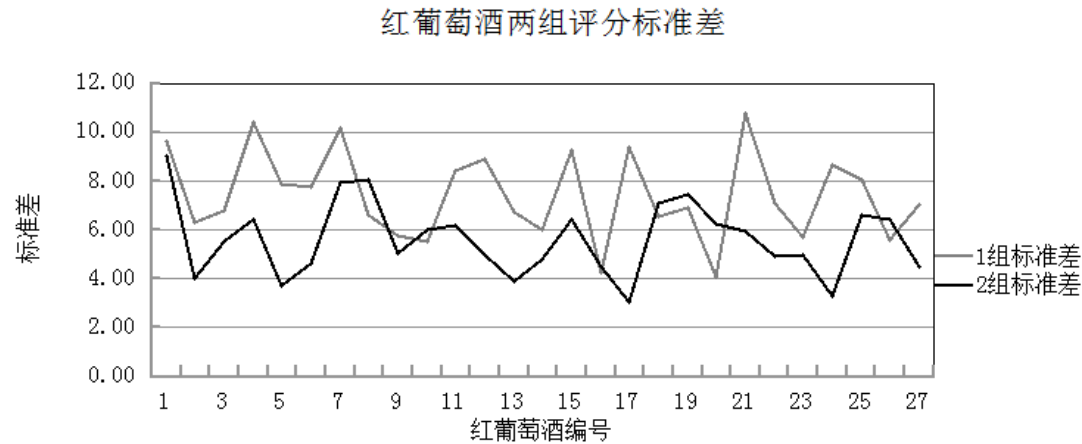


图 2 红葡萄酒两组评分标准差

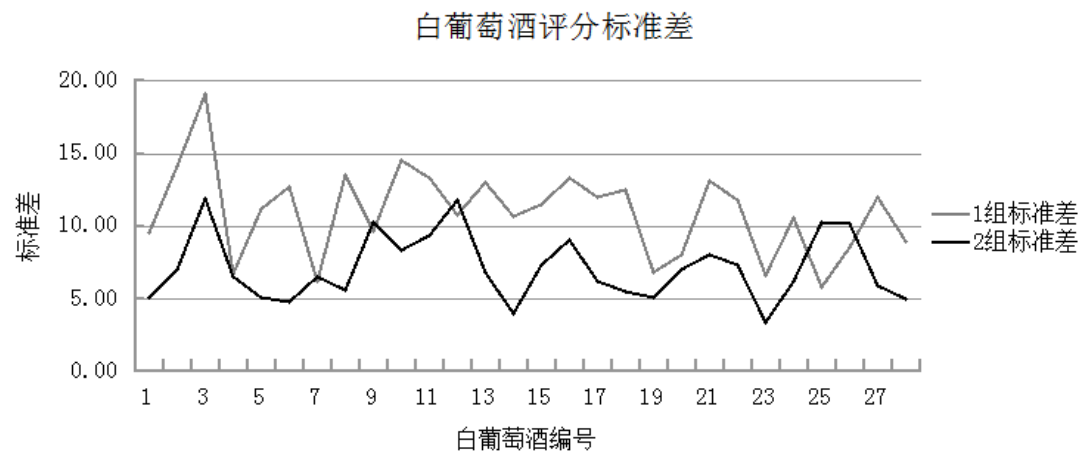


图 3 白葡萄酒两组评分标准差

从图中我们可以明显的看出第二组评酒员无论是对红葡萄酒还是白葡萄酒就的评价都要比第一组评酒员稳定和一致，因此第二组评酒员更可信。

6 问题 2 模型的建立与求解

6. 1 模型的建立

6. 1. 1 主成份分析法的使用

根据主成分分析的方法，分析酿酒葡萄理化指标的检测数据。步骤如下：

(1)为了消除不同变量量纲的影响，首先需要对所有变量进行标准化处理：设数据中酿酒葡萄品种共有  $n$  个，主要一级指标共有  $p$  个，分别为  $X_1, X_2, \dots, X_p$ ，令  $x_{ij}(i = 1, 2, \dots, n; j = 1, 2, \dots, p)$  为第  $i$  个样本的第  $j$  个指标的数值。做变换：

$$Y_j = \frac{x_j - E(x_j)}{\sqrt{var(x_j)}} \quad (j = 1, 2, \dots, p) \quad (2)$$

得到标准化的数据:

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (3)$$

其中  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ ,  $s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$

(2) 计算相关系数矩阵: 在标准化数据  $Y = (y_{ij})_{n \times p}$  的基础上计算  $p$  个原始值, 标相关系数矩阵  $R = (r_{ij})_{p \times p}$ , 其中:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (4)$$

(3) 求相关系数矩阵  $R$  的特征值  $R$  的特征值并排序  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , 再求出  $R$  的特征值相应的正则特征化向量  $l_i = (l_{1i}, l_{2i}, \dots, l_{pi})$ , 则第  $i$  个主成分表述为各个指标  $X_k$  的线性组合

$$Z_i = \sum_{k=1}^p l_{ki} X_k \quad (5)$$

(4) 计算综合得分:

首先得到第  $i$  个样本中第  $k$  个成分的得分为

$$F_{ik} = \sum_{j=1}^p l_{jk} X_j \quad (6)$$

再以  $m$  个主成分的特征值贡献率为权重, 求得第  $i$  个样本的综合得分:

$$f_i = \sum_{k=1}^m F_{ik} \times \lambda_k \quad (i = 1, 2, \dots, n) \quad (7)$$

## 6.2 模型的求解

通过 SPSS 得到酿酒红葡萄与酿酒白葡萄的成分碎石图, 如图:

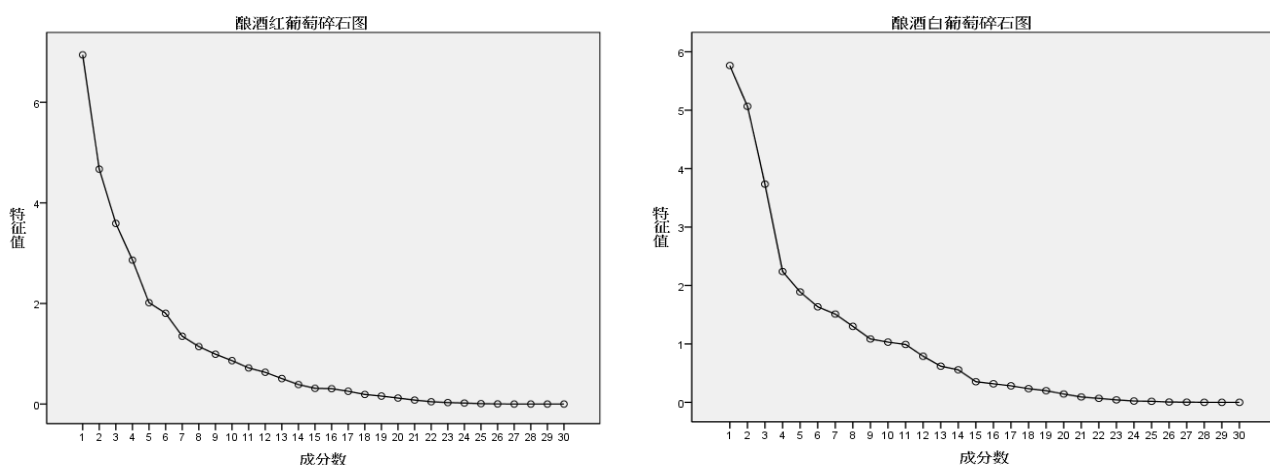


图 4 酿酒红葡萄与酿酒白葡萄成分碎石图

从图中我们可以看出酿酒红葡萄的理化指标被分成了 8 个主成份, 而酿酒白葡萄被分成了 10 个主成份。具体两种酿酒葡萄的主成份及贡献率见附录中的成分表。为了给葡萄分级, 我们分为如下几步:

(1) 接合附录表中酿酒葡萄中的各成分贡献率得出各葡萄样本的评分:

表2 酿酒红葡萄样本的评分

红葡萄样品	1	2	3	4	5	6	7
评分	98.57	95.05	173.91	-82.58	12.27	35.94	-12.50
红葡萄样品	8	9	10	11	12	13	14
评分	39.14	97.15	-157.78	-56.11	29.59	-35.49	29.79
红葡萄样品	15	16	17	18	19	20	21
评分	7.39	-38.52	51.04	-23.26	5.93	-51.37	61.18
红葡萄样品	22	23	24	25	26	27	
评分	0.18	58.96	-6.33	-130.89	-102.33	-98.93	

表3 酿酒白葡萄样本的评分

白葡萄样品	1	2	3	4	5	6	7
评分	-92.03	-6.07	25.60	-34.25	35.90	26.35	13.85
白葡萄样品	8	9	10	11	12	13	14
评分	-109.48	3.31	17.44	-19.01	16.49	-49.42	-37.52
白葡萄样品	15	16	17	18	19	20	21
评分	20.27	-117.8	-61.21	-1.05	-50.70	-3.50	-20.42
白葡萄样品	22	23	24	25	26	27	28
评分	-22.26	33.80	115.71	10.97	41.76	230.01	33.26

(2) 根据葡萄样品的评分使用 SPSS 进行系统聚类分析，得到对酿酒葡萄的分级

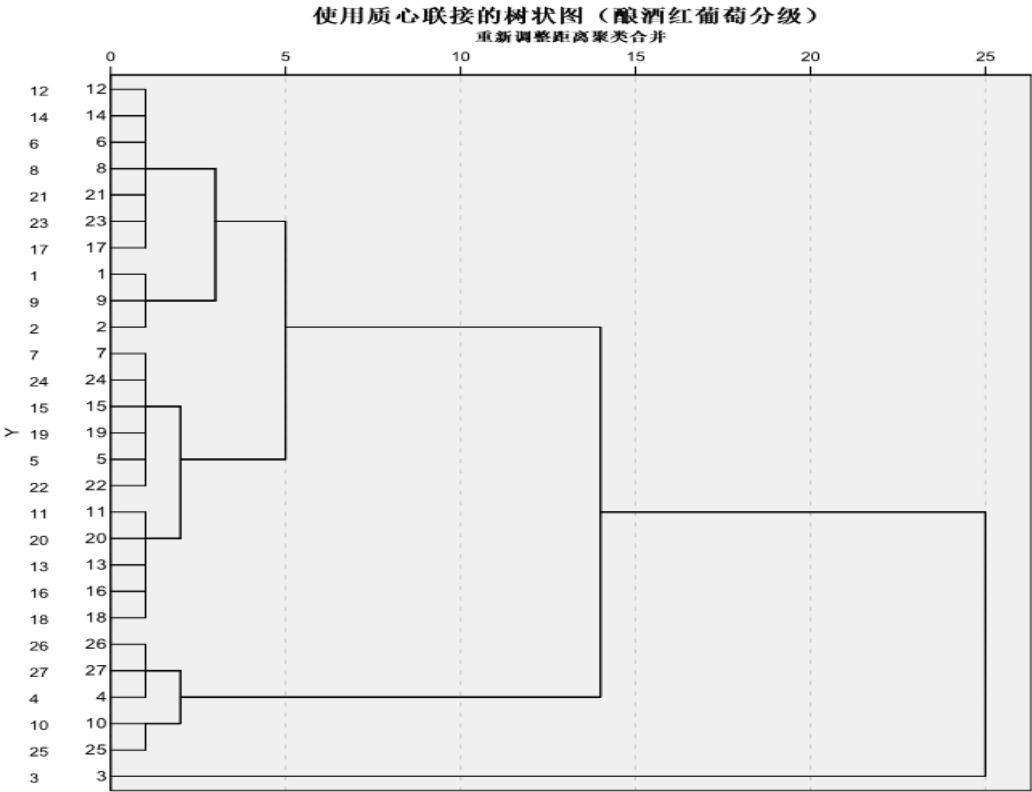


图5 酿酒红葡萄分级树状图

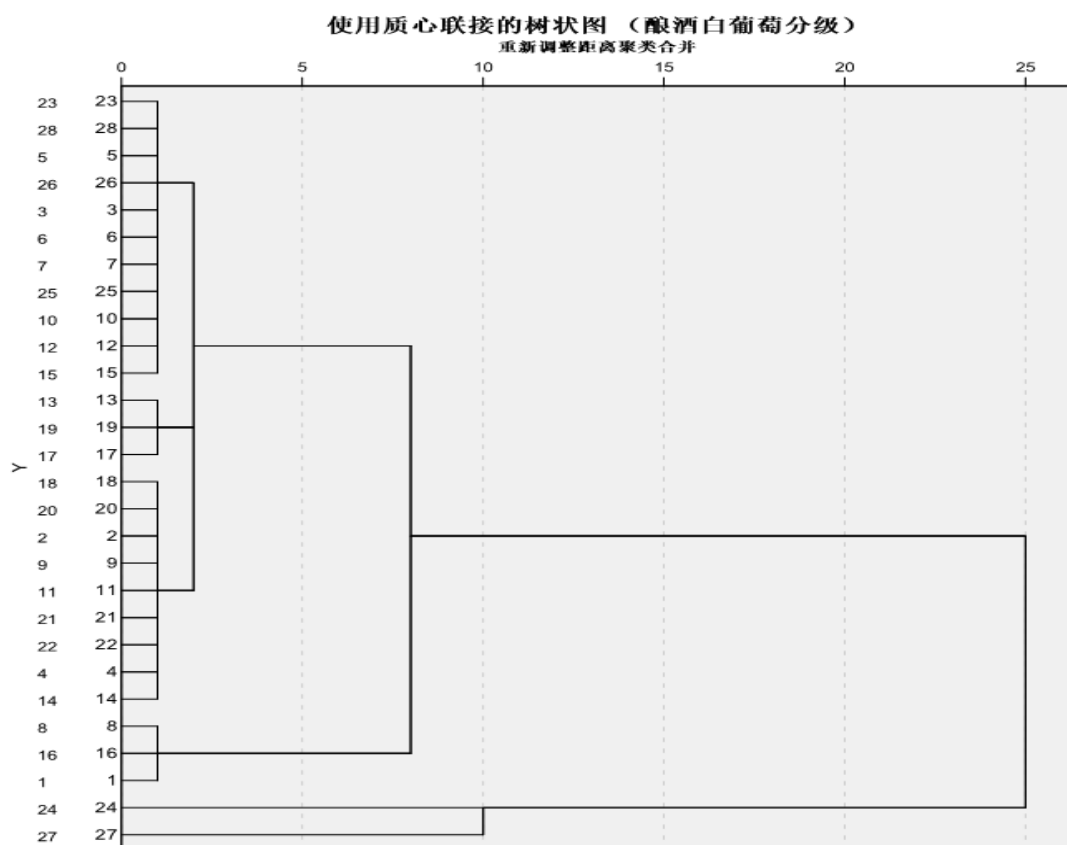


图 6 酿酒白葡萄分级树状图

最后，得到酿酒红葡萄与酿酒白葡萄的分级见表 4 和表 5：

表 4 酿酒红葡萄分级表

酿酒红葡萄等级	好	较好	一般	较差
样品编号	3	12、14、6、8、 21、23、17、 1、9、2	7、24、15、 19、5、22、 11、20、13、 16、18	26、27、4、 10、25

表 5 酿酒白葡萄分级表

酿酒白葡萄等级	好	较好	一般	较差
样品编号	24、27	23、28、5、 26、3、6、7、 25、10、12、 15	13、19、17、 18、20、2、9、 11、21、22、 4、14	8、16、1

由于第一问得到组 2 评酒师的评价结果更可信，因此通过与其对比，可以发现一个普遍的认知：好的葡萄酿出好的葡萄酒。可见分级结果是符合现实依据的，是可信的。

## 7 问题 3 模型的建立与解决



7.1 模型的建立

典型相关分析的基本原理是：为了从总体上把握两组指标之间的相关关系，分别在两组变量中提取有代表性的两个综合变量 U 和 V（分别为两个变量组中各变量的线性组合），利用这两个综合变量之间的相关关系来反映两组指标之间的整体相关性。

分析酿酒葡萄和葡萄酒的理化指标之间的联系，就是研究酿酒葡萄的理化指标与葡萄酒的理化指标之间的相互关联程度。具体方法如下：（由于两类葡萄与葡萄酒的模型与解法一致，故只以红葡萄与红葡萄酒举例）

- (1) 根据问题 2 中的理化指标与贡献度的分析，得出 6 个红葡萄酒的主要因子与 8 个酿酒红葡萄的主要因子。
- (2) 在此假设酿酒红葡萄的理化指标用变量  $X = (x_1, x_2, \dots, x_i)$  表示，红葡萄酒的理化指标用变  $Y = (y_1, y_2, \dots, y_i)$  表示。分别在两组变量中选取具有代表性的综合变量  $U_i, V_i$ ，使得每一个综合变量是原变量的线性组合，即：

$$U_i = \sum_{j=1}^8 a_{ij} \cdot x_j \tag{8}$$

$$V_i = \sum_{j=1}^6 b_{ij} \cdot y_j \tag{9}$$

- (3) 为了确保典型变量的唯一性，只考虑方差为 1 的 X、Y 的线性函数  $a_{ij} \cdot x_j$  与  $b_{ij} \cdot y_j$  求使得他们相关系数达到最大的这一组。

若存在常数向量  $a_{ij} \cdot x_j$   $b_{ij} \cdot y_j$  在  $D(a \ X) = 1 = D(b \ Y)$  的条件下，使得相关系数  $\rho(a^{(1)}X, b^{(1)}Y)$  达到最大，则  $a^{(1)}X$ 、 $b^{(1)}Y$  是 X、Y 的第一典型相关向量，他们之间的相关向量就叫做典型相关系数。

求出第一对典型相关变量之后，可以类似的求出各对之间互不联系的第二对典型相关变量、第三对典型相关变量等。这些典型相关变量就反映了 X、Y 之间的相关情况。

7.2 模型的求解

通过 SPSS 自带的典型性相关分析工具包，对酿酒葡萄与葡萄酒之间的关系做简要分析。得到酿酒红葡萄指标变量 U 的系数矩阵 A 为：

表 6 酿酒红葡萄指标变量 U 的系数矩阵

主要因子	1	2	3	4	5	6	理化指标
x1	-0.929	0.912	0.602	-0.243	-0.743	-0.011	花色苷鲜重
x2	-0.176	0.202	-0.131	0.631	0.241	-0.338	单宁
x3	0.327	-0.504	-1.993	-0.47	1.527	1.025	总酚
x4	-0.256	-0.617	0.907	-0.492	-2.368	-0.231	葡萄总黄酮
x5	0.281	0.55	-1.284	0.188	-1.12	-0.141	蛋白质
x6	-0.336	-0.485	1.195	0.707	1.368	0.136	DPPH 自由基
x7	0.222	-0.155	-0.052	-0.57	0.713	-0.899	出汁率
x8	-0.105	-0.249	0.528	0.613	0.306	-0.47	酒石酸

红葡萄酒指标变量 V 的系数矩阵 B 为：

表 7 红葡萄酒指标变量 V 的系数矩阵

主要因子	1	2	3	4	5	6	理化指标
y1	-0.698	1.088	0.484	-0.868	0.694	-0.023	花色苷
y2	0.067	0.386	-1.416	1.421	1.426	-2.074	单宁
y3	-0.737	-0.678	1.343	2.088	-3.394	0.544	总酚
y4	-0.153	0.013	-1.534	-1.46	-0.707	-1.969	酒总黄酮
y5	-0.173	-0.159	0.487	-0.262	0.578	-1.273	白藜芦醇
y6	0.571	-0.893	0.617	-1.22	2.012	4.064	DPPH 半抑制 体积

表 8 酿酒红葡萄与红葡萄酒理化指标间的相关性

U	V	Corr (U, V)
花色苷	花色苷、DPPH 自由基、总酚	0.973
总酚、葡萄总黄酮、蛋白质	花色苷、总酚、DPPH 半抑制 体积	0.918
总酚、蛋白质、DPPH 自 由基、	单宁、总酚、酒总黄酮	0.651

酿酒白葡萄指标变量 U 的系数矩阵 A 为：

表 9 酿酒白葡萄指标变量 U 的系数矩阵

主要因子	1	2	3	4	5	理化指标
x1	-0.095	-0.46	0.216	0.807	0.57	花色苷鲜重
x2	0.171	0.416	-0.192	0.594	0.667	单宁
x3	-0.784	-1.916	-0.389	1.199	-1.372	总酚
x4	0	1.725	0.021	-1.348	0.833	葡萄总黄酮
x5	-0.544	-0.152	0.673	-0.025	0.558	蛋白质
x6	0.019	0.38	-0.067	0.586	-0.25	DPPH 自由基
x7	0.128	-0.064	0.19	0.809	0.356	出汁率
x8	-0.318	0.378	0.924	0.227	-0.344	酒石酸

红葡萄酒指标变量 V 的系数矩阵 B 为：

表 10 白葡萄酒指标变量 V 的系数矩阵

主要因子	1	2	3	4	5	理化指标
y1	-0.231	0.799	1.345	0.59	1.465	单宁
y2	0.107	0.536	-1.444	-1.521	-0.999	总酚
y3	-0.852	-0.727	-0.005	0.248	0.456	葡萄总黄酮
y4	0.335	0.151	-0.681	0.353	0.554	蛋白质
y5	-0.05	-0.206	-0.217	1.187	-0.856	DPPH 自由基

表 11 酿酒白葡萄与白葡萄酒理化指标间的相关性

U	V	Corr (U, V)
---	---	-------------

总酚 蛋白质 酒石酸	葡萄总黄酮 蛋白质	0.854
花色苷 丹宁 总酚	单宁 葡萄总黄酮	0.710

## 8 问题 4 模型的建立与求解

对于葡萄酒来说，品质的好坏不单由原料即葡萄的品质决定，气味给人们感官上带来的直接刺激也是影响因素之一。分析各因素对于葡萄酒品质的影响其实就是分析各因素与葡萄酒品质的拟合程度，拟合程度越大则影响程度越深。

### 8.1 模型的建立

#### 8.1.1 多元线性回归模型的建立

为了简化线性拟合的复杂度，在考虑酿酒葡萄的理化指标对葡萄酒品质的影响时，使用第二问所求得的两酿酒葡萄理化指标的主成份作为理化指标的代表。同时使用附件中所给的葡萄酒的理化指标以及芳香物质共三项影响因素，建立如下多元回归线性模型：

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m + \varepsilon \quad (10)$$

其中， $b_0, b_1, \dots, b_m$  是  $m+1$  个未知参数， $b_0$  称回归常数， $b_1, \dots, b_m$  称回归系数， $m \geq 2$  时，即为多元线性回归模型。 $\varepsilon$  是随机误差（残差）。

令  $n$  组相互独立的实验数据为  $(x_{i1}, x_{i2}, \dots, x_{im}, y_i), i = 1, 2, \dots, n$ ，则线性回归模型可用矩阵形式表示为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (11)$$

运用 MATLAB 求解得到多元线性经验回归方程

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m + \varepsilon$$

### 8.2 模型求解

将葡萄理化指标主成份、葡萄酒理化指标与芳香物质的数据代入上述公式进行处理，得到多元线性回归方程。

最后得到各因素与葡萄酒品质的拟合程度（即  $b$  值）如下表所示：

表 12 影响因素与葡萄酒的拟合程度

影响因素	红葡萄酒拟合程度	白葡萄酒拟合程度
红葡萄理化指标	0.7446	0.5148
红葡萄酒理化指标	0.5785	0.1529
芳香物质	0.6588	0.7286

根据上表可知在白葡萄酒中芳香物质的影响程度最大，而在红葡萄酒中芳香物质的影响程度也排在了第二位。由此可知芳香物质是一种不可忽略的影响因素，故只用葡萄和葡萄酒的理化指标来评价葡萄酒的质量是不合理的。

## 9 模型的优点与缺点

### 9.1 模型的优点

(1) 问题二建立的主成分分析法用较少的新指标代替原来的大量指标,使少数新指标尽可能多的保留原来的指标所反映的信息,使问题变的简化。

(2) 在研究两组变量之间的相关关系时,巧妙的引入了典型相关分析,避免了使用多远线性回归而带来的繁琐求解过程。

(3) 在处理数据时尽可能的使用 SPSS,对问题中的大量数据处理带来了方便。

### 9.1 模型的缺点

(1) 在分析酿酒葡萄与葡萄酒理化指标之间的联系时,忽略了葡萄酒色泽的影响。

(2) 在解决问题时,仅考虑了数据中的一级指标,忽略了二级指标带来的影响。

(3) 酿酒葡萄的一些指标并不一定是越大越好,比如 PH 值。因此,对酿酒葡萄的分级存在一定的误差。

## 参考文献

- [1]魏宗舒.概率论与数理统计(第二版).北京:高等教育出版社,2012
- [2]Mark M.Meerschaert.Mathematical Modeling(Third Edition).北京:机械工业出版社,2010
- [3]Frank R.Giordano 等.A First Course In Mathematical Modeling(Fourth Edition).北京:机械工业出版社,2010
- [4]胡晓东 董辰辉.MATLAB 从入门到精通.北京:人民邮电出版社,2010
- [5]游玲 王涛 李华兰.葡萄酒芳香物质研究进展.四川食品与安全.第 44 卷 第 2 期:29-33,2008
- [6]李华 刘曙东 王华.葡萄酒感官评价结果的统计分析方法研究.中国食品学报,2006
- [7]杨小平.统计分析方法与 spss 应用教程.北京:清华大学出版社,2008
- [8]张文彤 钟云飞.IBM SPSS 数据分析与挖掘实战案例精粹.北京:清华大学出版社,2013

## 附录

表 酿酒红葡萄的主成份表

理化指标	1	2	3	4	5	6	7	8
氨基酸总量	0.36	0.57	0.02	0.43	-0.14	-0.36	0.11	0.04
蛋白质	0.65	-0.46	-0.22	0.24	0.21	-0.10	0.07	0.00
VC 含量	-0.12	-0.38	-0.14	-0.07	-0.59	-0.09	-0.03	0.18
花色苷	0.82	-0.11	0.20	-0.30	0.03	0.23	-0.08	0.04
酒石酸	0.40	0.11	-0.36	0.34	0.39	-0.14	0.29	-0.40
苹果酸	0.37	0.31	-0.03	-0.68	0.00	0.40	-0.05	0.05
柠檬酸	0.31	0.22	-0.28	-0.40	0.45	-0.02	0.43	-0.27
多酚氧化酶活力	0.29	0.11	0.34	-0.54	0.34	-0.16	-0.13	0.20
褐变度	0.58	-0.08	0.10	-0.72	0.03	-0.05	-0.09	0.04
DPPH 自由基	0.77	-0.44	0.02	0.20	-0.10	0.04	0.19	0.09
总酚	0.85	-0.16	0.19	0.25	-0.11	0.15	-0.01	-0.07
单宁	0.71	-0.12	0.33	-0.03	-0.24	0.14	0.31	-0.15
葡萄总黄酮	0.71	-0.29	0.18	0.29	-0.12	0.26	0.16	-0.05
白藜芦醇	0.14	0.03	-0.78	-0.06	-0.27	-0.01	0.28	0.32
黄酮醇	0.56	0.08	0.10	-0.09	-0.04	-0.55	0.32	0.39
总糖	0.17	0.77	0.21	0.11	0.05	-0.06	-0.13	0.30
还原糖	0.05	0.75	0.17	0.15	0.16	0.01	-0.16	0.17
可溶性固形物	0.31	0.61	0.24	0.39	0.16	-0.02	-0.01	0.00
PH 值	0.35	-0.26	-0.26	0.64	0.14	-0.05	-0.32	0.23
可滴定酸	-0.36	0.43	0.59	0.10	-0.36	0.09	0.29	-0.15
固酸比	0.45	-0.13	-0.46	0.03	0.52	0.08	-0.27	0.13
干物质含量	0.32	0.84	0.29	0.12	0.11	0.06	0.03	0.03
果穗质量	-0.34	-0.51	0.18	0.12	0.50	0.19	0.22	0.28
百粒质量	-0.54	-0.42	0.39	0.18	0.13	0.25	0.18	0.29
果梗比	0.60	-0.13	-0.06	-0.27	-0.28	-0.51	-0.02	0.07
出汁率	0.52	-0.18	0.24	0.18	-0.16	0.42	-0.17	0.05
果皮质量	-0.28	-0.24	0.68	-0.03	0.26	-0.04	0.39	0.27
果皮颜色 l	-0.57	-0.20	-0.08	-0.08	0.13	-0.30	0.11	-0.04
果皮颜色 a	-0.27	0.39	-0.72	-0.02	-0.11	0.20	0.22	0.19
果皮颜色 b	0.03	0.49	-0.52	-0.01	-0.07	0.53	0.27	0.21

表 酿酒白葡萄的主成份表

理化指标	1	2	3	4	5	6	7	8	9	10
氨基酸总量	0.54	0.18	0.27	-0.49	0.28	0.06	0.10	0.13	-0.15	-0.20
蛋白质	0.20	0.69	0.12	0.29	-0.11	-0.03	-0.35	0.02	-0.13	-0.01
VC 含量	-0.20	-0.13	-0.62	0.33	0.08	0.01	0.43	0.06	0.08	0.05
花色苷	-0.37	-0.39	0.09	-0.13	0.38	-0.08	0.27	-0.43	0.08	-0.42
酒石酸	0.35	-0.46	0.14	-0.09	0.51	-0.02	0.27	0.39	0.09	0.06
苹果酸	0.12	0.37	0.04	-0.39	0.69	-0.10	-0.20	0.01	-0.22	0.04

柠檬酸	0.20	-0.06	0.32	0.07	0.42	-0.37	-0.22	0.11	0.30	0.15
多酚氧化酶活力	-0.41	-0.43	-0.03	0.17	0.03	-0.47	-0.14	-0.25	-0.04	0.12
褐变度	0.28	0.13	-0.20	0.62	0.01	-0.02	-0.42	0.29	0.17	-0.20
DPPH 自由基	0.41	0.46	-0.18	0.02	-0.19	0.00	0.40	-0.03	0.06	-0.16
总酚	0.00	0.81	0.26	0.29	0.25	0.03	0.20	-0.05	-0.09	0.14
单宁	0.44	0.50	0.19	0.05	-0.11	-0.06	0.37	-0.31	-0.07	0.31
葡萄总黄酮	-0.08	0.82	0.31	0.19	0.22	-0.02	0.25	-0.15	-0.06	0.03
白藜芦醇	0.02	0.11	0.22	-0.27	0.09	0.79	-0.22	-0.11	0.17	-0.10
黄酮醇	0.26	0.35	0.38	0.57	0.35	0.15	-0.07	-0.01	0.30	0.08
总糖	0.79	-0.06	0.02	-0.19	-0.26	-0.13	0.02	0.09	0.08	0.18
还原糖	0.72	-0.01	0.10	-0.27	-0.20	-0.05	-0.26	-0.20	0.23	-0.06
可溶性固形物	0.74	-0.23	0.22	-0.07	-0.09	-0.32	0.12	0.16	0.10	-0.05
PH 值	0.20	-0.35	0.36	0.09	-0.19	0.45	0.23	0.40	-0.28	0.23
可滴定酸	-0.01	0.52	-0.61	-0.38	0.01	-0.02	-0.04	0.00	0.26	0.33
固酸比	0.15	-0.49	0.67	0.35	0.03	-0.05	0.03	0.07	-0.23	-0.23
干物质含量	0.83	-0.11	0.21	-0.19	0.02	0.02	-0.12	-0.28	-0.01	0.06
果穗质量	-0.65	0.47	0.17	-0.20	0.03	0.14	-0.10	0.07	0.09	-0.17
百粒质量	-0.55	0.52	0.25	-0.16	-0.28	-0.03	0.03	0.23	0.22	-0.18
果梗比	-0.09	-0.50	-0.43	-0.01	0.30	0.24	0.16	0.08	0.40	0.11
出汁率	-0.57	0.09	-0.32	-0.14	0.24	-0.08	-0.23	0.21	-0.36	0.24
果皮质量	-0.33	0.45	0.37	-0.26	-0.13	-0.32	0.21	0.43	0.22	-0.11
果皮颜色 1	0.49	0.24	-0.63	0.34	0.16	0.16	0.07	-0.06	-0.03	-0.19
果皮颜色 a	-0.29	-0.42	0.61	0.09	0.00	0.22	-0.09	-0.12	0.16	0.34
果皮颜色 b	0.69	0.07	-0.57	-0.03	0.12	0.06	-0.12	0.17	-0.16	-0.14