

葡萄酒的评价

摘要

本文运用多种相关分析、综合评价和线性回归等方法解决了葡萄酒质量的评价问题。

对于问题一，首先通过单样本 K-S 检验等方法确定了各葡萄酒样本评分数据的概率分布，从而确定了显著性差异模型的建立，接着考虑两组评分数据的配对关系约束，引入 Wilcoxon 符号秩检验法来进行显著性差异的假设检验。结果显示对于红、白葡萄酒，两个品酒组的评价结果均存在显著性差异。最后利用秩相关分析，引入肯德尔和谐系数法评定评酒组的评分信度，评价结果显示对于红葡萄酒，第一组品酒员的品尝得分更为可信，而对于白葡萄酒则是第二组品酒员在可信度方面占优。

问题二，运用主成分分析法进行指标遴选，构建酿酒葡萄质量的综合评价指标体系，并利用该指标体系建立基于综合评价的酿酒葡萄分级模型，对酿酒葡萄进行分级。结果发现样本葡萄大多集中在二、三级，红葡萄样本中样本 23 质量最优，为特级葡萄；样本 12 质量相对欠缺，属六级葡萄。

问题三中，采用研究两组变量之间相关关系的多元统计方法——典型相关分析，识别并量化两组变量——酿酒葡萄与葡萄酒的理化指标——之间的关系。分析结果如下：第一，增大酿酒葡萄果皮的含量对葡萄酒中 DPPH 半抑制体积含量的增加有重要影响；第二，酿酒葡萄中的苹果酸不仅能促发酵，还能给对红葡萄酒起主要呈色作用的花色苷和对花色苷起中等辅色作用的单宁物质起保护作用，使得红葡萄酒呈色亮丽；第三，在葡萄总黄酮消除自由基的抗氧化作用和总酚保护清除自由基的共同作用下，酿酒葡萄中的 DPPH 自由基转化为葡萄酒中的 DPPH 半抑制体积。

对于问题四，首先在问题三分析酿酒葡萄与葡萄酒的理化指标间联系的基础上，在保留葡萄酒指标的前提下，剔除酿酒葡萄指标中某些认为可以被用于表示对应葡萄酒指标的部分。接着，利用筛选后的指标建立多元线性回归模型，探究酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响。经检验样本组的线性回归模型评价值与评分值的显著性差异检验，用葡萄和葡萄酒的理化指标来评价葡萄酒的质量是可行的。

本文综合秩相关分析评价、基于层次分析法的综合评价、典型相关分析、多元线性回归等模型，结合 MATLAB、SPSS、SAS 和 EXCEL 等软件，对葡萄酒质量的评价问题进行了多角度的分析，并给出了利用理化指标评价葡萄酒质量的模型。在文章的最后对模型的适用范围做出了推广，在实际应用中有一定的参考价值。

关键词：秩相关 主成分分析 层次分析综合评价 典型相关分析 多元线性回归

一、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？

二、问题分析

2.1 问题一的分析

问题一要求比较两组评价结果的是否存在差异，并建立合理的评价模型以判断两组结果在可信程度方面的优劣。首先，我们从问题分析可以得出品酒员对葡萄酒样本的品尝评分是属于感官评价，具有较大的主观性。因此，我们先从问题所给的数据入手，分析四组品酒结果中对不同样本打分分布。依靠葡萄酒样本评分的概率分布，建立显著性差异模型。由于品酒员间存在评价尺度、评价位置和评价方向等方面的差异，不同组别的品酒员对同一酒样的评价结果存在着差异。此时不适用参数检验的方法，而只能用非参数统计方法来处理。

对主观评分结果合理性的评价，仅仅局限于评分之间表面的数值关系是不够的。因此，考虑采取秩相关分析法建立评价模型，将评分结果的具体数值部分予以丢弃，只保留各评分秩大小关系的信息，以给出数据中最稳固、最一般的关系，度量整体评分结果在可信度方面的优劣。

2.2 问题二的分析

酿酒葡萄，是指以酿造葡萄酒为主要生产目的的葡萄品种[1]。问题二要求分析确定合理的评价指标体系，并运用该评价指标体系对酿酒葡萄进行分级。显而易见，该问题要求我们建立一个评价模型。

评价体系主要包含两方面指标：

第一个方面是葡萄酒的质量。这包括外观、香气、口感、整体四方面的评分。外观包括澄清度和色调，香气包括纯正度、浓度和质量，口感则通过纯正度、浓度、持久性和质量体现。

第二个方面酿酒葡萄自身的理化指标。如附加二中的葡萄总黄酮、总酚、单宁、果皮质量等 27 个指标。对于这 27 个酿酒葡萄自身的理化指标，根据多个样本得到的数据分析出其内在的关系，将相关性显著的指标合并，则可以使得计算简单。

那么由以上的分析可以构建综合评价指标体系，建立模型进行多指标综合评价。基于综合评价的结果，即可对酿酒葡萄进行分级。

2.3 问题三的分析

问题三中，题目要求分析酿酒葡萄与葡萄酒的理化指标之间的联系。酿酒葡萄和葡萄酒分别存在多个理化指标，若采用简单相关分析的方法，只是孤立考虑了单个 x 与单个 y 间的相关，而没有考虑 x 、 y 变量组内部各变量间的相关。酿酒葡萄经发酵酿成葡

萄酒的化学过程，使得两组变量间有许多简单相关系数，使问题显得复杂，难以从整体描述。因此，考虑采用研究两组变量之间相关关系的多元统计方法——典型相关分析，识别并量化酿酒葡萄与葡萄酒的理化指标两组变量之间的关系，考虑两组变量的线性组合，并研究它们之间的相关系数 $p(u,v)$ 。

2.4 问题四的分析

问题四中，需要我们通过酿酒葡萄和葡萄酒的理化指标，得到对葡萄酒的质量的评价，并论证是否可行。因此，首先考虑在问题三的基础上，针对酿酒葡萄与葡萄酒理化指标之间的联系和它们与葡萄酒质量之间的相关性进行指标的筛选。随后，期望建立一个线性回归模型，通过该模型来得到对葡萄酒质量的评价。

由于要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，初步认为在建立线性回归模型时对样本进行随机遴选，选中的样本作为示例样本组建立线性回归方程，未选中的样本作为检验样本组对模型的可行性进行验证。

三、模型假设

1. 假设各样本能真实客观地反映酿酒葡萄与葡萄酒的情况；
2. 葡萄酒的质量只与酿酒葡萄的好坏有关，忽略酿造过程中的温度、湿度、人为干扰等其他因素的影响；
3. 不考虑理化性质的二级指标；
4. 每组评酒员的打分不受上个酒样品的影响，即各评分数据间独立；

四、符号说明

序号	符号	符号说明
1.	m	品酒员个数
2.	n	样本数
3.	j	样本序数
4.	i	指标序数
5.	$r_{ii'}$	第 i 个指标与第 i' 个指标的相关系数
6.	p	一级评价指标中的指标序数
7.	q	二级评价指标中的指标序数
8.	y	酿酒葡萄质量综合评价
9.	B	每一酿酒葡萄样本所在级别
10.	X	酿酒葡萄理化指标
11.	Y	葡萄酒的理化指标
12.	β	线性回归系数
13.	V	典型变量
14.	W	解释变量

五、模型建立与求解

5.1 问题一的模型建立与求解

问题一要求分析两组评酒员的评价结果有无显著性差异，并判断两组结果在可信程度方面的优劣。我们认为由以下三个步骤组成：

步骤一：葡萄酒样本评分概率分布的确定，其目的是确定显著性差异模型的类型；

步骤二：两组评酒员评价结果的显著性差异模型的建立，主要通过 Wilcoxon 符号秩检验法进行显著性差异的假设检验；

步骤三：建立秩相关分析评价模型，并通过该模型判断两组品酒员评价结果在可信度方面的优劣。

5.1.1 数据的预处理

经过对数据的查找，我们发现部分原始数据存在异常，另外有些类型数据存在缺失，在此我们将其正常化处理。

(1) 缺失数据的处理

对于数据中存在的缺失现象，本文采用均值替换法对这种缺失数据进行处理。

均值替换法就是将该项目剔除异常数据后取整剩余数据的平均值来替换异常或缺失数据的方法，即：

$$x_m^* = \frac{1}{9} \left[\sum_{k=1, k \neq m}^{10} x_k \right] \quad (m=1, 2, \dots, 10)$$

其中， x_m^* 为缺失值。

由于不同品酒师对同一样本相同项目的打分值差别不大，所以认为采用均值替换法来处理缺失数据是可行的。以“酒样品 20”色调数据为例进行修补，得到修正后的数据如下表所示。

表 1 红葡萄酒样品 20 色调数据修补

品酒员	1 号	2 号	3 号	4 号	5 号	6 号	7 号	8 号	9 号	10 号
修补前	6	6	4	---	6	6	8	6	6	8
修补后	6	6	4	6	6	6	8	6	6	8

注：表中“---”代表数据缺失

(2) 异常数据的修正

原始数据中，有的数据明显比两侧的数据过大或过小，显然是不合理数据。

例如，第一组白葡萄酒品尝评分的数据中，可能由于手工输入的误差，品酒员 7 对样品 3 持久性评分的数据相对于相邻各品酒员的评分发生了明显的突变现象。这种数据异常有可能对数据挖掘的结果产生不利影响。

表 2 第一组白葡萄酒品尝评分样本 3 持久性数值异常

品酒员	1 号	2 号	3 号	4 号	5 号	6 号	7 号	8 号	9 号	10 号
持久性	7	5	7	5	6	7	77	5	6	7

对于类似的异常数据采取“先剔除，后替换”的策略，对异常数据进行修正。

5.1.2 各葡萄酒样本评分数据概率分布的确定

对两组品酒员差异性评价的假设检验一般要求数据符合正态分布。统计规律表明，正态分布有极其广泛的实际背景，生产与科学实验中很多随机变量的概率分布都可以近似地用正态分布来描述[2]。因此，对葡萄酒质量的评分进行正态性检验有助于我们分析得出该评分是否科学、合理。

首先，计算针对每一个样本 10 个品酒员的评分均值，即

$$\bar{x} = \frac{\sum_{m=1}^{10} x_{mn}}{10} \quad (m=1,2,\dots,10 \quad n=1,2,\dots,10)$$

其次，利用 SPSS 统计软件中的 P-P 图和单样本 K-S 检验，对数据集两组品酒员分别对红、白葡萄酒品尝得到的四组评价结果（见附录 8.1.2）进行了正态分布检验，若样点在正态分布 P-P 图上呈直线散布，则被检验数据基本上成一条直线[3]。

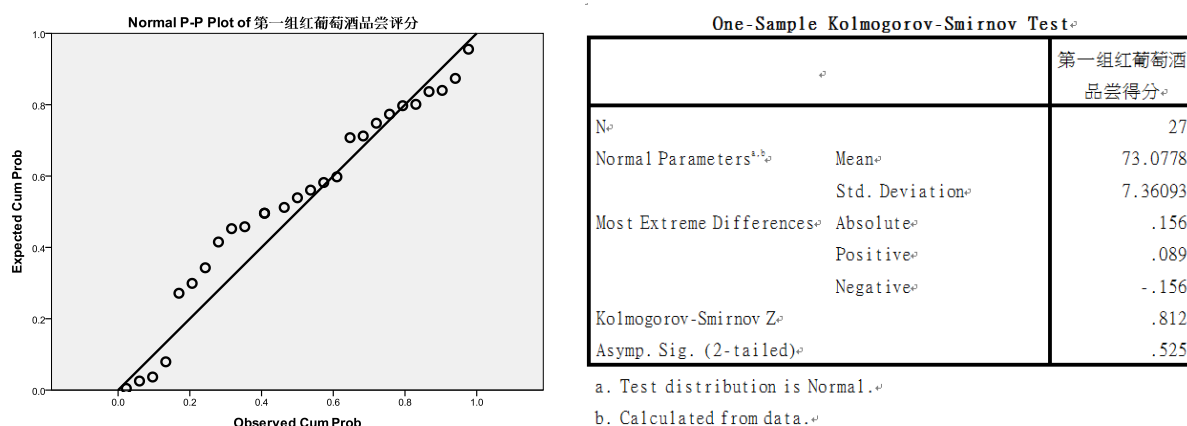


图 1 第一组红葡萄酒评价结果的正态 P-P 图和 K-S 检验结果

从图 1 可以看出第一组（其余三组见附录 8.1-图 8.1）数据的散点分别近似为一条直线，且与对角线大致重叠；双边检验结果 $p = 0.525 > 0.05$ 。因此可以认为品酒员对葡萄酒的评分服从正态分布。

5.1.3 两组评价结果的显著性差异评价

上述检验显示各类葡萄酒得分情况属于正态总体，为了进一步说明品酒员评分的科学性以及两个评分组评分的可信度，需要检查两组给出的评分是否有显著性差异，即对数据进行显著性检验。

两配对样本非参数检验一般用于同一研究对象分别给予两种不同处理的效果比较[4]。因为两组品酒员分别对同一样本组进行评分，故两组数据为配对数据。对于两组配对数据的检验，需要引入适用于 T 检验中的成对比较，但并不要求成对数据之差 D_i 服从正态分布，只要求对称分布即可[5]的 Wilcoxon 符号秩检验法，用来决定两个样本是否来自相同的或相等的总体。其检验步骤（以红葡萄为例）如下：

Step1. 提出假设： H_0 ：两组品酒员对酒样本的评价结果是相同的；

H_1 ：两组品酒员对酒样本的评价结果是不同的。

Step2. 选定显著性水平 $\alpha = 0.05$ ， $n_1 = n_2 = 27$

Step3. 根据样本值计算成对观测数据之差 D_i ，并将 D_i 的绝对值按大小顺序编上等级。最小的数据等级为 1，第二小的数据等级为 2，以此类推（若有数据相等的情形，则取这几个数据排序的平均值作为其等级）（见附录 8.1.3）。

Step4. 等级编号完成后恢复正负号，分别求出正等级之和 T^+ 和负等级之和 T^- ，选择 T^+ 和 T^- 中较小的一个作为威尔科克森检验统计量 T 。

Step5. 统计量 T 的均值和方差分别为 $E(T)$ 和 $D(T)$ ，确定检验统计量

$$z = \frac{T - E(t)}{\sqrt{D(t)}} \sim N(0,1)$$

近似服从于标准正态分布。

Step6. 查正态分布表可得 $-z_{\alpha/2} = -z_{0.05/2} = -1.96$ 的值，确定 H_0 的拒绝域为

$$z = \frac{T - E(t)}{\sqrt{D(t)}} = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \leq -1.96$$

根据样本值计算的检验统计量的观测值

$$z = \frac{T - E(t)}{\sqrt{D(t)}} = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx -2.53 < -1.96$$

所以应拒绝 H_0 ，即在显著性水平 $\alpha = 0.05$ 下，认为两个品酒组对红葡萄酒的评价结果是不同的，即存在显著性差异。

类似地，对于两个品酒组白葡萄酒的评价结果（见附录 8.1.3），可以得到

$$z = \frac{T - E(t)}{\sqrt{D(t)}} \approx -2.23 < -1.96$$

因此，两组品酒员对红、白葡萄酒的评价结果均存在显著性差异。

5.1.4 秩相关分析评价模型的建立

感官分析（特别是描述分析）是评价葡萄酒质量高低的重要方法[6-7]。然而，在葡萄酒的感官评价中，由于品酒员间存在评价尺度、评价位置和评价方向等方面的差异，不同组别的品酒员对同一酒样的评价结果存在显著性差异。因此，判断不同组别品酒员对葡萄酒质量的评价结果在可信程度方面的优劣尤为重要。

评鉴主评的评分信度主要采用肯德尔和谐系数法来评定。肯德尔和谐系数是指“以等级次序排列或以等级次序表示的变量之间的相关”，它常用来刻画多个评价者对某一事物评判或评分的一致性程度[8]。设有 m 个品酒员对 j 个样本评分，用肯德尔和谐系数法分析评分的一致性，步骤如下：

Step1. 对附件 1 葡萄酒品尝评分表中的数据进行秩变换（见附录 8.1.4）；

Step2. 计算和谐系数 ω ，建立秩相关分析评价模型，其目的是度量品酒组的整体评分效果；

令和谐系数

$$\omega = Q / \left[\frac{1}{12} \times m^2 (j^3 - j) \right]$$

其中， Q 为每个被评样本所得等级秩数之和 R_{mj} 与所有等级分的平均数 \bar{R} 之差的平方和，

即： $Q = \sum_{m=1}^{10} (R_{mj} - \bar{R})^2$ 。

$$\text{为计算方便，可将上式化为： } Q = \sum_{m=1}^{10} R_{mj}^2 - \frac{\left(\sum_{m=1}^{10} R_{mj} \right)^2}{j}$$

Step3. 进行显著性检验。

本文令 10 个品酒员为一个品酒组，分别对 27 个红葡萄酒样本和 28 个白葡萄酒样本进行评分，第 m 个品酒员对第 j 个样本的评分秩为 r_{mj} ，则 $R_{mj} = \sum_{m=1}^{10} r_{mj}$ 为品酒组对第 j

个样本的评秩之和。显然，当品酒组意见比较一致时，各个秩和 R_1, R_2, \dots, R_n 之间差距较大，而当品酒组意见分歧较大时， R_1, R_2, \dots, R_n 差距较小。

解得四组评分结果 ω 值如下：

表 3 各组品酒员对葡萄酒品尝得分 ω 值

组别	第一组红葡萄酒品尝得分	第二组红葡萄酒品尝得分	第一组白葡萄酒品尝得分	第二组白葡萄酒品尝得分
ω 值	0.044366788	0.043314652	0.040039039	0.041112055

本文样本数 $j > 7$ ，检验统计量为 χ^2 ， $\chi^2 = m(j-1)\omega$ 。 χ^2 服从自由度为 j 的 χ^2 分布。 χ^2 的计算值如表 4 所示，如果 $\chi^2 > \chi_{(m)\alpha}$ ，那么有 $100(1-\alpha)\%$ 的把握可以断定 10 个品酒员的评分存在相关。

表 4 各组品酒员对葡萄酒品尝得分检验结果

组别	第一组红葡萄酒品尝得分	第二组红葡萄酒品尝得分	第一组白葡萄酒品尝得分	第二组白葡萄酒品尝得分
χ^2	11.53536508	11.26180952	10.81054064	11.10025493

由表 4 可知，对于红葡萄酒，第一组品尝得分存在相关的概率大于第二组品尝得分存在相关的概率，即第一组品酒员的评价结果更为可信；而对于白葡萄酒，第二组存在相关的概率大于第一组，即是第二组品酒员的评价结果更为可信。因此，本文后续分析中，对于红葡萄酒质量评价结果选用第一组品尝得分，对于白葡萄酒质量评价结果选用第二组品尝得分。

5.2 问题二的模型建立与求解

问题二要求我们建立模型，可以根据酿酒葡萄自身的理化指标和酿造后葡萄酒的质量情况，对酿酒葡萄进行分级。为解决该问题，我们通过以下步骤来评价与分级酿酒葡萄。

步骤一：酿酒葡萄 27 种指标之间的关系研究，目的是构建评价模型的指标体系；

步骤二：建立综合评价模型，并通过该模型对步骤一得到的指标进行多指标综合评价，以对酿酒葡萄进行分级。

5.2.1 分级综合评价指标体系的构建

1. 酿酒葡萄 27 种指标的遴选

对于酿酒葡萄而言，虽然每种指标在成因上互不相同，但是不同的指标之间往往具有相关性，其产生的原因是有潜在的因素对酿酒葡萄的各指标起支配作用。为了找到这些潜在因素以及相应的支配作用，本文选用主成分分析法[9]对这些问题做以解决，步骤如下：

Step1. 为消除不同变量的量纲的影响，首先需要对变量进行标准化中处理；

本部分涉及到的指标共 27 个，样本对象 27 个，第 j 个样本的第 i 个指标值为 F_{ij} ，将各标准化值按如下方式进行标准化为 \tilde{F}_{ij} ：

$$\tilde{F}_{ij} = \frac{F_{ij} - \bar{F}_i}{s_i} \quad (1)$$

其中 \bar{F}_i 和 s_i 分别为 i 指标的均值和标准差。标准化的目的在于消除不同变量的量纲的影响，而且标准化转化不会改变变量的相关系数。

Step2. 计算标准化数据的相关系数阵，求出相关系数矩阵的特征值和特征向量。

记第 i 个指标与第 i' 个指标的相关系数为 $r_{ii'}$ ，其计算方法为：

$$r_{i'i'} = \frac{\sum_{k=1}^{27} \tilde{F}_{ik} \tilde{F}_{i'k}}{27-1}, i, i' = 1, 2, \dots, 27$$

则相关系数矩阵为 $R = (r_{i'i'})_{27 \times 27}$ ，其中， $r_{ii} = 1$ ， $r_{i'i'} = r_{i'i}$ 。

Step3: 计算特征值与特征向量；

计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{27} \gg 0$ ，及其对应的特征向量 $\xi_1, \xi_2, \dots, \xi_{27}$ ，其中 $\xi_i = (\mu_{1i}, \mu_{2i}, \dots, \mu_{27i})^T$ ，由特征向量组成 27 个新的指标变量：

$$\begin{cases} Y_1 = \mu_{1,1}\tilde{F}_1 + \mu_{2,1}\tilde{F}_2 + \dots + \mu_{27,1}\tilde{F}_{27} \\ Y_2 = \mu_{1,2}\tilde{F}_1 + \mu_{2,2}\tilde{F}_2 + \dots + \mu_{27,2}\tilde{F}_{27} \\ \dots\dots\dots \\ Y_{27} = \mu_{1,27}\tilde{F}_1 + \mu_{2,27}\tilde{F}_2 + \dots + \mu_{27,27}\tilde{F}_{27} \end{cases}$$

其中， Y_i 为第 i 主成分， $i = 1, 2, \dots, 27$ 。

Step4. 确定 p 个主成分，进行统计分析。

根据以上步骤，本文利用 SPSS 统计软件，首先求得各指标的相关性系数表（见附件 5），从表中可以发现，某些指标具有很强的相关性，如果直接用这些指标对酿酒葡萄质量进行分级，不仅会使得运算量过大，同时还会造成信息的重叠，影响分级的客观性。主成分分析可以把多个指标转化成少数几个不相关的综合指标。以红葡萄为例，相应主成分的特征值和累计贡献率如下表：

表 5 红葡萄特征值和累计贡献率

主成分	提取平方和载入			旋转平方和载入		
	合计	方差的%	累计%	合计	方差的%	累计%
1	6.966	23.221	23.221	5.196	17.318	17.318
2	4.940	16.467	39.687	4.458	14.859	32.177
3	3.737	12.457	52.144	3.135	10.451	42.629
4	2.840	9.467	61.611	2.712	9.039	51.668
5	1.999	6.663	68.274	2.690	8.968	60.636
6	1.742	5.808	74.082	2.565	8.552	69.187
7	1.418	4.728	78.810	2.257	7.523	76.711
8	1.270	2.234	83.044	1.900	6.333	83.044

在累计方差为 83.044% 的前提下分析得到八个主成分，这八个主成分提供了附件 2 酿酒红葡萄的理化指标中 83.044% 的信息，满足主成分分析原则。从表 5 还可以看到，主成分 1 和 2 的累计贡献率较大，这就可以解释为主成分 1 与主成分 2 可能是酿酒葡萄分级最重要的指标。

由以上分析利用 SPSS 统计软件计算得到主成分分析正交解见附录 8.2.1。

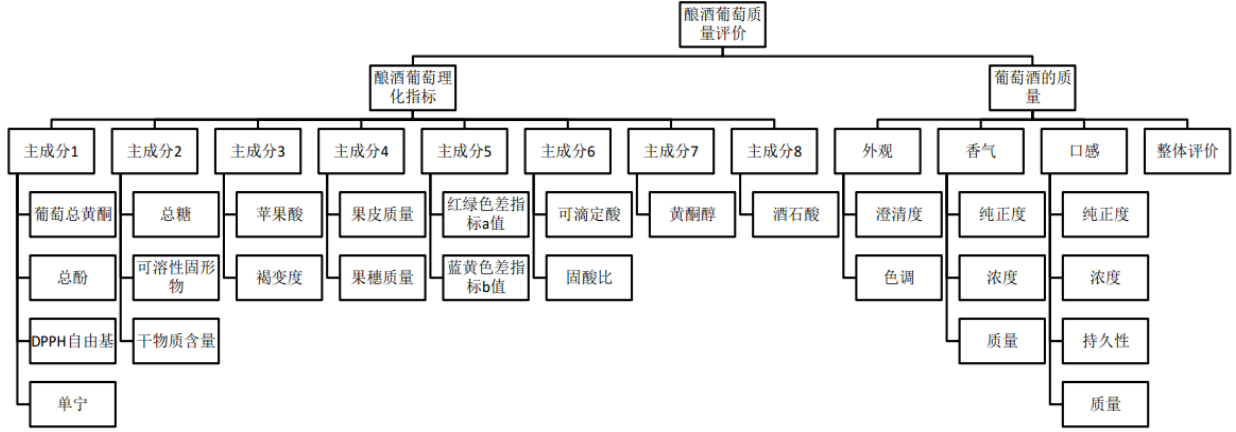
正交解说明，红葡萄理化指标当中，主成分 1 为葡萄总黄酮、总酚、DPPH 自由基和单宁的组合，主成分 2 为总糖、可溶性固形物和干物质含量的组合，主成分 3 为苹果酸和褐变度的组合，主成分 4 为果皮质量与果穗质量的组合，主成分 5 为红绿色差指标 a 值和黄蓝色差指标 b 值的组合，主成分 6 为可滴定酸和固酸比的组合，主成分 7 为黄酮醇，主成分 8 为酒石酸。这组合说明葡萄总黄酮、总酚、DPPH 自由基和单宁，总糖、可溶性固形物和干物质含量，苹果酸和褐变度，果皮质量与果穗质量，红绿色差指标 a 值、黄蓝色差指标 b 值和白藜芦醇，可滴定酸和固酸比可能在同一方面对酿酒葡萄分级起重要作用，而黄酮醇、酒石酸分别在不同角度影响酿酒葡萄的质量与分级。

2. 指标体系的初步建立

根据酿酒葡萄指标遴选分析与已知葡萄酒质量评分规则，以红葡萄为例，酿酒葡萄理化指标 R_1 、葡萄酒质量 R_2 是评价葡萄质量的一级指标，其中一级指标——酿酒葡萄理化指标 R_1 进一步分为主成分 1 至主成分 8 的八个二级指标，葡萄酒质量 R_2 分为外观、香气、口感、整体的四个二级指标讨论。八个二级指标和四个二级指标下面又分别进一步分为 17 个和 10 个三级指标，故三级指标共 27 个。

综上所述，酿酒葡萄质量评价指标体系构架为：

图 2 酿酒葡萄质量评价指标体系构架表



5. 2. 2 基于综合评价的酿酒葡萄分级模型的建立

1. 数据的预处理

(1) 评价指标类型的一致化处理

在已建立的指标体系中，指标集可能同时含有“极大型”和“极小型”指标，我们分别称之为优质因子和劣质因子，也存在“中间型”指标。因此在评价之前必须将评价指标的类型进行一致化处理，即要统一化为极大型指标。

极小型指标：对于某个极小型指标 x_i ，通过平移变换 $x'_i = M_i - x_i (x_i > 0, i = 1, 2, \dots, 27)$ ，其中 M_i 为指标 x_i 可能取到的最大值。

中间型指标：对于某个中间型指标 x_i ，要将其化为极大型指标，令

$$x'_i = \begin{cases} \frac{2(x_i - m_i)}{M_i - m_i}, m_i \leq x_i \leq \frac{1}{2}(M_i + m_i) \\ \frac{2(M_i - x_i)}{M_i - m_i}, \frac{1}{2}(M_i + m_i) \leq x_i \leq M_i \end{cases}$$

其中 M_i 、 m_i 分别为 x_i 可能取值的最大值和最小值。

(2) 评价指标的无量纲化处理

本文的各个评价指标之间由于各自的度量单位及数量级的差别，而存在着不可公度性，这就为确定综合评价指标带来了困难和问题。此时，可以通过一些常用的数学变换，即数据的无量纲化，来消除原始指标数据的差异影响。

本文采用极差化的方法，对 n 个样本 27 项三级指标的指标值 $x_{ij} (j = 1, 2, \dots, 27; i = 1, 2, \dots, n)$ 做无量纲化处理，令

$$M_i = \max \{x_{1,i}, x_{2,i}, \dots, x_{n,i}\}, m_i = \min \{x_{1,i}, x_{2,i}, \dots, x_{n,i}\}$$

则新的指标为

$$x_{ij}^* = \frac{x_{ij} - m_i}{M_i - m_i} \in [0,1]$$

即 x_{ij}^* ($j=1,2,\dots,n$; $i=1,2,\dots,27$) 为无量纲化的指标值。

2.运用层次分析法（AHP）确定评价指标权重[9]

考虑到酿酒葡萄的分级不仅仅是由葡萄或葡萄酒内的一种成分决定的，并且每一种成分对分级的影响也不一样。为了确定各指标对酿酒葡萄分级影响的权值，本文采用层次分析法和综合评价法[9]进行酿酒葡萄的评价。

对于层次分析法中的判断矩阵，根据不同理化性质在样本中的分布情况以及不同样本的评分结果，确定各个指标之间的相对重要程度，可以得到如下判断矩阵表（以红葡萄为例）：

表 6 判断矩阵表

	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6	因子 7	因子 8
因子 1	1.0000	1.1655	1.6571	1.9159	1.9311	2.0250	2.3020	2.7346
因子 2	0.8580	1.0000	1.4218	1.6439	1.6569	1.7375	1.9751	2.3463
因子 3	0.6035	0.7033	1.0000	1.1562	1.1654	1.2221	1.3892	1.6502
因子 4	0.5219	0.6083	0.8649	1.0000	1.0079	1.0569	1.2015	1.4273
因子 5	0.5178	0.6035	0.8581	0.9921	1.0000	1.0486	1.1921	1.4161
因子 6	0.4938	0.5755	0.8183	0.9461	0.9536	1.0000	1.1368	1.3504
因子 7	0.4344	0.5063	0.7198	0.8323	0.8389	0.8797	1.0000	1.1879
因子 8	0.3657	0.4262	0.6060	0.7006	0.7062	0.7405	0.8418	1.0000

得到判断矩阵后，求得其最大特征向量，将该特征向量归一化处理后即可得到各影响分级程度指标的权向量（以红葡萄为例，白葡萄见附录 8.2.1）：

$$w = (0.025338, 0.050677, 0.037031, 0.049375, 0.098749, 0.024966, 0.033288, 0.033288, 0.091543, 0.055744, 0.027149, 0.026999, 0.025413, 0.02471, 0.03016, 0.029704, 0.029601, 0.031552, 0.031373, 0.028495, 0.025929, 0.027401, 0.026595, 0.02642, 0.025072, 0.045296, 0.038131)$$

表 7 各影响分级程度指标的权重汇总表

澄清度	色调	纯正度	浓度	质量	纯正度	浓度
0.025338	0.050677	0.037031	0.049375	0.098749	0.024966	0.033288
持久性	质量	整体评价	葡萄总黄酮	总酚	DPPH 自由基	单宁
0.033288	0.091543	0.055744	0.027149	0.026999	0.025413	0.02471
总糖	可溶性固形物	干物质含量	苹果酸	褐变度	果皮质量	果穗质量
0.03016	0.029704	0.029601	0.031552	0.031373	0.028495	0.025929
a*(+红; -绿)	b*(+黄;-蓝)	可滴定酸	固酸比	黄酮醇	酒石酸 (g/L)	
0.027401	0.026595	0.02642	0.025072	0.045296	0.038131	

此时判断矩阵的最大特征根为 8，随机一致性指标 $CI^{(1)} = 0.013$ ， $CR^{(2)} = CR^{(3)} = 0 < 0.10$ ，组合一致性比率指标为： $CR = CR^{(1)} + CR^{(2)} + CR^{(3)} = 0.029 < 0.1$ ，表明判断矩阵具有满意一致性，可以作为评价因子的权向量。

3. 综合评价模型的建立

酿酒葡萄综合评价模型是通过一定的数学模型或算法将多个指标的评价值“合成”一个综合评价值，实现对酿酒葡萄质量的综合评价以分级。线型加权求和法因计算简单而广泛采用。设 x_p 为第 p 个一级评价指标所得评价值， x_{pq} 为第 p 个一级评价指标的第 q 个二级评价指标所得的评价值， x_{pqi} 为第 p 个一级评价指标，第 q 个二级评价指标的第 i 个三级评价指标所得的评价值。将酿酒葡萄分级综合评价指标所得的评价值以相应的权重系数来加权，其加权和作为酿酒葡萄质量的综合评价值 y ：

$$\begin{aligned} y &= \sum w_p (x_p) x_p ; \\ x_p &= \sum w_{pq} (x_{pq}) x_{pq} ; \\ x_{pq} &= \sum w_{pqi} (x_{pqi}) x_{pqi} ; \end{aligned}$$

综上所述，酿酒葡萄综合评价模型为：

$$\begin{cases} y = \sum w_p (x_p) x_p \\ x_p = \sum w_{pq} (x_{pq}) x_{pq} \\ x_{pq} = \sum w_{pqi} (x_{pqi}) x_{pqi} \\ x_p = f(R_p) \\ x_{pq} = f(R_{pq}) \\ x_{pqi} = f(R_{pqi}) \end{cases}$$

其中， $w_p (x_p)$ 为第 p 个一级评价指标的权重值， $w_{pq} (x_{pq})$ 为第 p 个一级评价指标的第 q 个二级评价指标的权重值， $w_{pqi} (x_{pqi})$ 为第 p 个一级评价指标，第 q 个二级评价指标的第 i 个三级评价指标的权重值，其值见各影响分级程度指标的权重汇总表（表 7）。

根据酿酒葡萄质量评价指标体系构架，有

$$\begin{aligned} y &= w_1 (x_1) f(R_1) + w_2 (x_2) f(R_2) \\ &= \sum_{q=1}^{10} w_{1q} (x_{1q}) f(R_{1q}) + \sum_{q=1}^{17} w_{2q} (x_{2q}) f(R_{2q}) \\ &= \sum_{i=1}^4 w_{11i} (x_{11i}) f(R_{11i}) + \sum_{i=1}^3 w_{12i} (x_{12i}) f(R_{12i}) + \sum_{i=1}^2 w_{13i} (x_{13i}) f(R_{13i}) + \sum_{i=1}^2 w_{14i} (x_{14i}) f(R_{14i}) \\ &\quad + \sum_{i=1}^2 w_{15i} (x_{15i}) f(R_{15i}) + \sum_{i=1}^2 w_{16i} (x_{16i}) f(R_{16i}) + w_{17i} (x_{17i}) f(R_{17i}) + w_{18i} (x_{18i}) f(R_{18i}) \\ &\quad + \sum_{i=1}^2 w_{21i} (x_{21i}) f(R_{21i}) + \sum_{i=1}^3 w_{22i} (x_{22i}) f(R_{22i}) + \sum_{i=1}^4 w_{23i} (x_{23i}) f(R_{23i}) + w_{24i} (x_{24i}) f(R_{24i}) \end{aligned}$$

综合评价值 y 的大小与酿酒葡萄的等级高低呈正相关关系，即综合评价值 y 越大，酿酒葡萄质量越好，等级越高。

4. 酿酒葡萄的分级阶梯模型的建立

就葡萄的质量的评价值 y 对酿酒葡萄进行分级，葡萄质量的评价值越高，葡萄质量越好，级别数越靠前（越小）；反之，葡萄质量的评价值越低，葡萄质量越差，分级所得的级别数越靠后（越大）。于是，建立分级阶梯模型来确定酿酒葡萄的级别数 B ：

评价值 y 分段阶梯模型标准确定如下表所示：

表 8 葡萄分级阶梯模型标准设定

级别数 B	<i>super</i>	1^{st}	2^{sec}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
评价值 y	[0.8,1]	[0.7,0.8)	[0.6,0.7)	[0.5,0.6)	[0.4,0.5)	[0.3,0.4)	[0.2,0.3)	[0,0.2)

则有葡萄质量分级阶梯模型为：

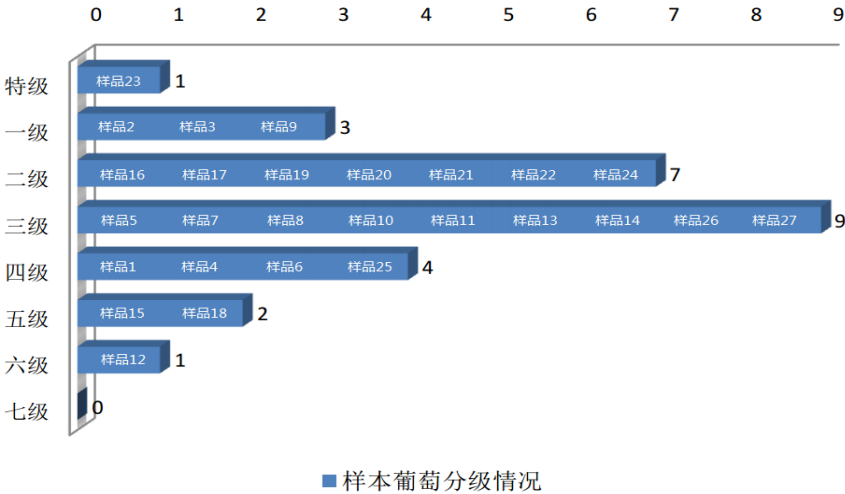
$$B = f(y) = \begin{cases} \textit{super} & 0.8 \leq y \leq 1.0 \\ 1^{st} & 0.7 \leq y < 0.8 \\ 2^{sec} & 0.6 \leq y < 0.7 \\ 3^{rd} & 0.5 \leq y < 0.6 \\ 4^{th} & 0.4 \leq y < 0.5 \\ 5^{th} & 0.3 \leq y < 0.4 \\ 6^{th} & 0.2 \leq y < 0.3 \\ 7^{th} & 0 \leq y < 0.2 \end{cases}$$

5. 2. 3 模型的求解——酿酒葡萄质量分级

利用 Excel 统计软件，求得各样本酿酒葡萄综合评价指标值与分级情况如下。

图表 9 红葡萄各样本综合评价指标值与分级

葡萄样品	1	2	3	4	5	6	7
评价值	0.40	0.71	0.71	0.49	0.59	0.46	0.52
分级	4^{th}	1^{st}	1^{st}	4^{th}	3^{rd}	4^{th}	3^{rd}
葡萄样品	8	9	10	11	12	13	14
评价值	0.59	0.75	0.55	0.50	0.23	0.57	0.58
分级	3^{rd}	1^{st}	3^{rd}	3^{rd}	6^{th}	3^{rd}	3^{rd}
葡萄样品	15	16	17	18	19	20	21
评价值	0.33	0.64	0.63	0.33	0.68	0.63	0.64
分级	5^{th}	2^{sec}	2^{sec}	5^{th}	2^{sec}	2^{sec}	2^{sec}
葡萄样品	22	23	24	25	26	27	
评价值	0.64	0.80	0.63	0.46	0.53	0.54	
分级	2^{sec}	<i>super</i>	2^{sec}	4^{th}	3^{rd}	3^{rd}	



结合以上图表可以得到：

- (1) 27 个酿酒葡萄样本中品质最优的为样本 23，品质最劣的为样本 12；
- (2) 样本集中的酿酒葡萄主要集中在二级与三级范围内，特级（最优级）与七级（最劣级）的样本个数分别为 0 和 1。越高级别的酿酒葡萄对各项指标趋于最优的要求相对较高，而越低级别的酿酒葡萄对各项指标远离最优的要求也相对较高，因此，要求越高，达到标准的样本数越少。

5.3 问题三的模型建立与求解

问题三要求分析酿酒葡萄与葡萄酒的理化指标之间的联系。这个问题对酿酒葡萄和葡萄酒的理化指标两组变量的关系分析提出了要求，对此本文从以下两个步骤进行回答：

步骤一：建立典型相关分析模型，其目的是分析酿酒葡萄与葡萄酒的理化指标之间的典型相关关系；

步骤二：根据上面的分析给出酿酒葡萄与葡萄酒的理化指标之间的联系。

5.3.1 典型相关分析模型的建立

为了研究酿酒葡萄与葡萄酒的理化指标之间的相关性，令酿酒葡萄为输入变量，葡萄酒为输出变量，采用典型相关分析法。它是利用主成分思想，分别找出输入变量与输出变量的线性组合，然后讨论线性组合之间的相关关系[10]。

典型相关分析模型的建立具体步骤如下：

Step1. 建立原始矩阵；

根据表格中原油数据，我们设酿酒葡萄的理化指标记为 $X = (X_1, X_2, \dots, X_{55})'$ ，葡萄酒的理化指标记为 $Y = (Y_1, Y_2, \dots, Y_9)'$ ， Z 为 30+9 总体的 27 次中心化观测数据阵：

$$Z = \begin{bmatrix} X_{1,1} & \cdots & X_{1,30} & Y_{1,1} & \cdots & Y_{1,9} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ X_{55,1} & \cdots & X_{55,30} & Y_{55,1} & \cdots & Y_{55,9} \end{bmatrix} = \begin{pmatrix} X_{55 \times 30} & Y_{55 \times 9} \end{pmatrix}'$$

Step2. 对原始数据进行标准化变换并计算相关系数矩阵；

我们利用问题二中公式(1)对酿酒葡萄和葡萄酒的理化指标数据进行标准化处理，然后计算两样本间的相关系数矩阵 R ，并将 R 分为

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

其中， R_{11} 、 R_{22} 分别为酿酒葡萄和葡萄酒指标内的相关系数阵， R_{12} 、 R_{21} 为酿酒葡萄指标与葡萄酒指标间的相关系数阵。

Step3. 求典型相关系数及典型变量；

首先求 $A = R_{11}^{-1} R_{12} R_{22}^{-1} R_{21}$ 的特征根 λ_i^2 ，特征向量 $S_1 \alpha_i$ ； $B = R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$ 的特征根 λ_i^2 ，特征向量 $S_2 \beta_i$ ，则有

$$\alpha_i = S_1^{-1}(S_1 \alpha_i), \beta_i = S_2^{-1}(S_2 \beta_i)$$

则随机变量酿酒葡萄的理化指标 X 和葡萄酒的理化指标 Y 的典型相关系数为 λ ，典型变量为

$$\begin{bmatrix} V_1 = \alpha_1' X \\ W_1 = \beta_1' Y \end{bmatrix}; \begin{bmatrix} V_2 = \alpha_2' X \\ W_2 = \beta_2' Y \end{bmatrix}; \dots; \begin{bmatrix} V_t = \alpha_t' X \\ W_t = \beta_t' Y \end{bmatrix} \quad (t \leq 55)$$

Step4. 检验各典型相关系数的显著性。

对典型相关系数 λ_i 进行显著性检验。在作两组变量酿酒葡萄的理化指标 X 和葡萄酒的理化指标 Y 的典型相关分析之前，首先应检验两组变量是否相关；如果不相关，即 $\text{cov}(X, Y) = 0$ ，则讨论的两组变量的典型相关就毫无意义。

5.3.2 典型相关模型的求解

附件 2 中包含 27 个红葡萄样本和 28 个白葡萄样本 2 组指标共 39 个指标的原始数据。其中，30 个是酿酒葡萄的理化指标，如下表所示。

表10 30个酿酒葡萄理化指标

x_1	x_2	x_3	x_4	x_5	x_6
氨基酸总量	蛋白质	VC 含量	花色苷	酒石酸	苹果酸
x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
柠檬酸	多酚氧化酶活力 E	褐变度	DPPH 自由基 1/IC50	总酚	单宁
x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
葡萄总黄酮	白藜芦醇	黄酮醇	总糖	还原糖	可溶性固形物
x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}
PH 值	可滴定酸	固酸比	干物质含量	果穗质量	百粒质量
x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
果梗比(%)	出汁率(%)	果皮质量	L*	a*(+红; -绿)	b* (+黄;-蓝)

除此之外，9 个是葡萄酒的理化指标： $x_{31} \rightarrow$ 花色苷(mg/L)， $x_{32} \rightarrow$ 单宁(mmol/L)， $x_{33} \rightarrow$ 总酚(mmol/L)， $x_{34} \rightarrow$ 酒总黄酮(mmol/L)， $x_{35} \rightarrow$ 白藜芦醇(mg/L)， $x_{36} \rightarrow$ DPPH 半抑制体积 (IV50) 1/IV50(uL)， $x_{37} \rightarrow$ L*(D65)， $x_{38} \rightarrow$ a*(D65)， $x_{39} \rightarrow$ b*(D65)。

利用 SPSS 软件对酿酒葡萄和葡萄酒的理化指标进行了典型相关分析，结果如下：

表11 典型相关系数

序号	1	2	3	4	5	6	7	8	9
典型相关系数	0.995	0.969	0.956	0.931	0.865	0.845	0.758	0.611	0.490
Prop Var	0.001	0.008	0.011	0.018	0.034	0.038	0.057	0.085	0.103

第一、第二、第三、第四对典型变量之间的典型相关系数都大于 0.9。由此可见这四对典型变量的解释能力比较强，并且相应典型变量之间密切相关。但要确定典型变量相关性的显著程度，需要进行典型相关系数的显著性检验。通过 SAS 检验结果如表 12 所示：

表 12 典型相关系数检验表

Test that remaining correlations are zero:			
	Wilk's	DF	Sig.
1	0.000	159.840	<0.0001
2	0.000	147.920	<0.0001
3	0.000	134.580	0.0002
4	0.001	119.780	0.0252
5	0.014	103.520	0.3337
6	0.058	85.792	0.686
7	0.202	66.608	0.9605
8	0.475	46.000	0.996
9	0.759	24.000	0.9925

显著性结果表明，在 0.01 的显著性水平下，前三对典型变量之间相关关系显著。标准化后的典型变量的系数来建立典型相关模型见表 13。

表 13 典型相关模型

序号	典型相关模型
1	$V_1 = 0.17x_{10} - 0.16x_{19} - 0.18x_{20} - 1.02x_{27}$ $W_1 = 4.13x_{36}$ $V_2 = 0.14x_3 - 0.29x_4 + 0.25x_5 - 0.58x_6$
2	$+0.37x_{11} - 0.47x_{12} + 0.54x_{13} + 0.36x_{20} - 0.35x_{28}$ $W_2 = -1.90x_{31} + 0.72x_{32} + 0.31x_{33} + 0.53x_{36} - 0.61x_{37}$ $V_3 = 0.26x_2 + 0.36x_4 + 0.25x_9 + 0.56x_{10} + 0.55x_{11}$
3	$+0.56x_{13} + 0.30x_{19} - 0.28x_{20} + 0.28x_{21} + 0.44x_{26}$ $W_3 = 0.27x_{31} + 0.24x_{32} + 0.30x_{33} + 0.58x_{34} + 0.2x_{35} + 0.43x_{36} - 0.35x_{37} - 0.19x_{38}$

根据典型变量重要程度及系数大小，从建立的典型相关模型可以看出，葡萄酒各指标受酿酒葡萄各指标变动的作用程度可用三对典型相关变量予以综合描述。

第一对，典型变量主要将 DPPH 半抑制体积从各种酿酒葡萄指标中分离出来（典型载荷为 4.13），与果皮质量呈现最大相关（相应典型载荷为-1.02）。由此可见，葡萄酒中的 DPPH 半抑制体积主要来自于葡萄的果皮，同时，酿酒葡萄中的 DPPH 自由基含量、PH 值、可滴定酸含量也对其有一定的影响。因此，增大酿酒葡萄果皮的含量对葡萄酒中 DPPH 半抑制体积含量的增加有重要影响。

第二对，典型变量将花色苷及单宁从 9 个葡萄酒指标中分离出来（典型载荷为-1.90 和 0.72），酿酒葡萄指标中与之相对应的解释变量是苹果酸、葡萄总黄酮和单宁（典型载荷为-0.58、0.54 和-0.47）。显而易见的，葡萄酒中和酿酒葡萄中的单宁具有较强的相关性，葡萄酒中的花色苷（类黄酮化合物）主要来源于酿酒葡萄中的葡萄总黄酮。值得注意的是，酿酒葡萄中的苹果酸不仅使得葡萄的发酵顺利进行，还保护着对红葡萄酒起主要呈色作用的花色苷和对花色苷起中等辅色作用的单宁物质，使得红葡萄酒呈现漂亮的宝石红色。

第三对，典型变量将酒总黄酮和 DPPH 半抑制体积从 9 个葡萄酒指标中分离出来（典型载荷为 0.58 和 0.43），酿酒葡萄指标中与之相对应的解释变量是 DPPH 自由基、葡萄总黄酮和总酚（典型载荷为 0.56、0.56 和 0.55）。在葡萄总黄酮消除自由基的抗氧化作用和总酚对清除自由基保护的共同作用下，酿酒葡萄中的 DPPH 自由基转化为葡萄酒中的 DPPH 半抑制体积。它们之间的对应关系可以用下图表示：

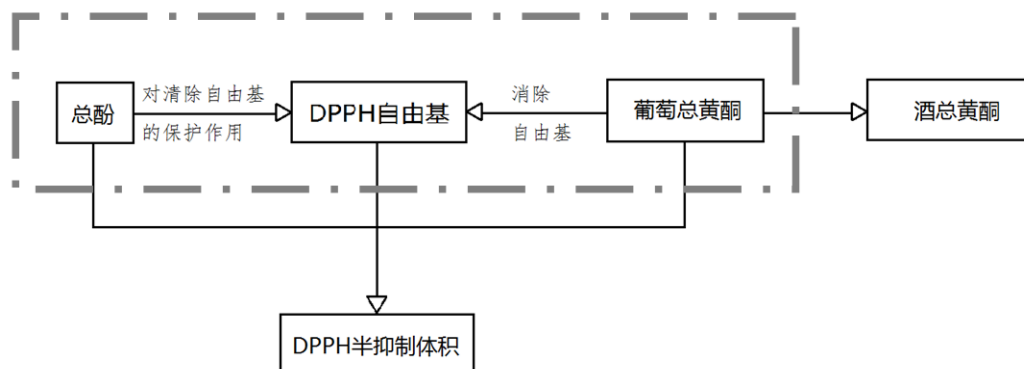


图 3 变量之间的对应关系

5.4 问题四的模型建立与求解

问题四要求分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响，并论证是否可

行。由于酿酒葡萄和葡萄酒的多个理化指标为自变量，因变量只有葡萄酒质量一个，故首先建立多元线性回归模型，后进行显著性差异检验。具体步骤如下：

Step1: 对样本进行随机筛选，选择 $n(n < N)$ 个进行分析；

Step2: 在问题三分析酿酒葡萄与葡萄酒的理化指标间联系的基础上对样本指标进行初步筛选；

Step3: 利用筛选后的指标与葡萄酒质量评价结果，建立多元线性回归模型；

Step4: 然后根据剩下的 $(N - n)$ 个样本对的酿酒葡萄和葡萄酒的理化指标，对葡萄酒质量求解得到的多元线性回归方程进行验证。

5.4.1 建立多元线性回归模型

1. 对酿酒葡萄和葡萄酒的理化指标的初级筛选

多元线性回归方程的建立要求指标之间互不相关，即无多重共线性。因此，本文在问题三分析酿酒葡萄与葡萄酒的理化指标间联系的基础上，在保留葡萄酒指标的前提下，剔除酿酒葡萄指标中某些认为可以被用于表示对应葡萄酒指标的部分。初步筛选后，留下酿酒葡萄和葡萄酒的共 23 个理化指标，结果见附件表 8.4.4。

2. 建立评价葡萄酒质量的多元线性回归模型

(1) 建立多元线性回归模型

涉及 p 个自变量的多元线性回归模型可表示为

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

为了方便，我们通过 n 组实际观察数据实际观察数据而引入矩阵记号：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

其中 X 成为模型设计矩阵，是常数矩阵， Y 与 ε 是随机向量，且

$$Y \sim N_n(X\beta, \sigma^2 I), \varepsilon \sim N_n(0, \sigma^2 I) \quad (I \text{ 为 } n \text{ 阶单位阵}),$$

ε 是不可观测的随机误差向量， β 是回归系数构成的向量，是未知待定的常数向量。

(2) 回归系数 β 的最小二乘估计

选取 β 的一个估计值，记为 $\hat{\beta}$ ，使随机误差 ε 的平方和达到最小，即

$$\begin{aligned} \min_{\beta} \varepsilon^T \cdot \varepsilon &= \min_{\beta} (Y - X\beta)^T (Y - X\beta) \\ &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \stackrel{\text{def}}{=} Q(\hat{\beta}) \end{aligned}$$

由最小二乘法的要求，由多元函数取得极值的必要条件可求解回归参数的标准方程如下

$$\begin{cases} \left. \frac{\partial Q}{\partial \beta_0} \right|_{\beta_0 = \hat{\beta}_0} = 0 \\ \left. \frac{\partial Q}{\partial \beta_i} \right|_{\beta_i = \hat{\beta}_i} = 0 \quad (j = 1, 2, \cdots, p) \end{cases}$$

可以证明任意给定的 X , Y , 正规方程组总有解, 虽然当 X 不满秩时, 其解不唯一, 但对任意一组解 $\hat{\beta}$ 都能使残差平方和最小, $Q(\hat{\beta}) = \min_{\beta} Q(\beta)$ 。

特别地, 当 X 秩时, 即 $r(X) = r(X^T X) = p$, 则正规方程组的解为 $\hat{\beta} = (X^T X)^{-1} X^T Y$, 即为回归系数的估计值。

(3) 逐步回归分析

由于建立回归模型时, 不是每一个因子对 y 的影响程度都很大。从而我们通过逐步回归的方法 (剔除优选法) 来对因子进行筛选。其具体过程如下:

Step1: 建立多元线性回归方程

Step2: 进行回归系数显著性检验, 取 t 值对应最大概率值 P_{\max} ;

Step3: 判断 P_{\max} 是否 ≤ 0.05 , 若满足则进入 step5, 若不满足, 进入 step4;

Step4: 则可接受 H_0 , 即这个指标与因变量线性关系不显著, 将指标剔除, 返回 Step1;

Step5: 则可拒绝 H_0 , 则所有指标与因变量线性关系显著, 输出方程, 结束。

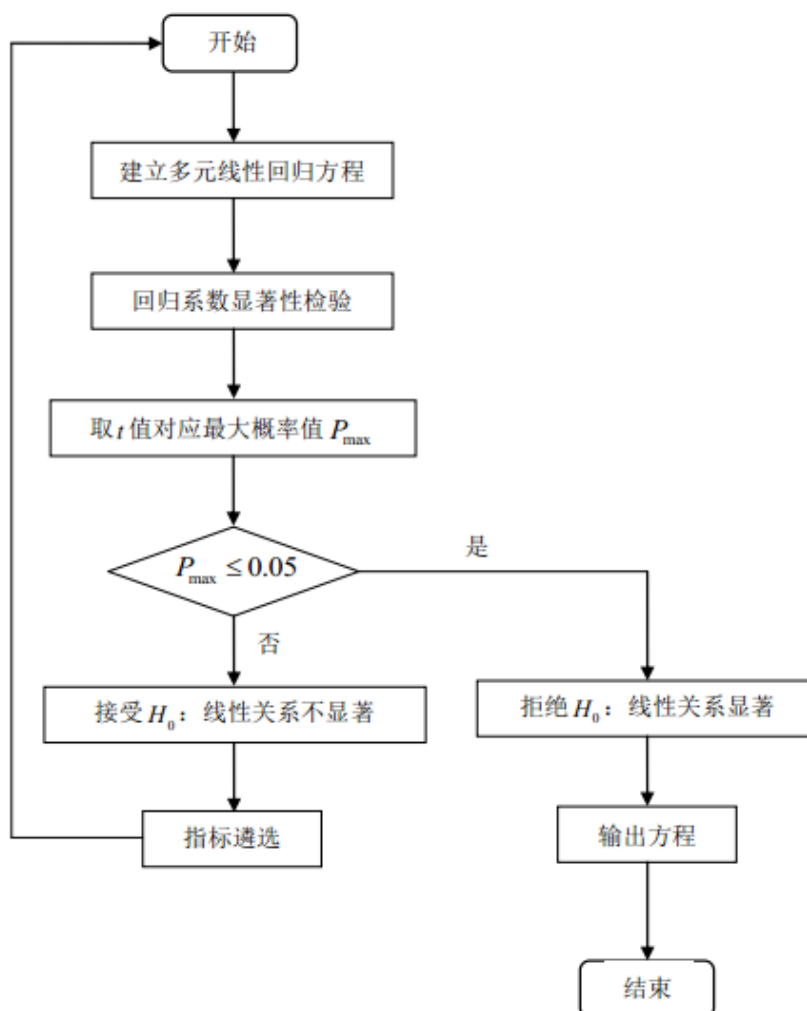


图 4 逐步回归分析流程图

(4) 多元回归模型的求解

首先, 利用 MATLAB 软件将红葡萄和对应红葡萄酒的 27 个样本随机抽取了 20 个样本 (见附录表 8.4.1)。然后利用这 20 个样本的指标值通过 SPSS 软件进行求解, 得到酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的七元线性回归方程

$$y = 797.579 + 0.862x_1 - 1.82x_2 + 0.385x_3 - 1.081x_4 - 65.783x_5 + 940.449x_6 + 7.126x_7$$

其中最后筛选和剔除后剩下评价葡萄酒质量的指标为：葡萄酒质量与葡萄理化指标黄酮醇 x_1 、可溶性固体物 x_2 、果穗质量 x_3 、百粒质量 x_4 、a*(+红；-绿) x_5 、葡萄酒理化指标 DPPH 半抑制体积 (IV50) x_6 以及 L*(D65) x_7 。

(5) 回归模型的显著性检验

为表明酿酒葡萄和葡萄酒的理化指标与葡萄酒质量有密切的关系，我们利用统计软件 SPSS 对求解得到的七元线性回归模型进行了显著性检验，结果如下：

A) 判定系数

表 14 模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.950 ^a	.902	.844	32.817

通过表 X 可以得到，复相关系数 $R=0.950$ ，多重判定系数 $R^2=0.902$ ；调整后的 $R^2=0.844$ 。调整后 R^2 的值越大，模型的拟合效果越好。

B) 回归方程的显著性检验

表 15 方差分析表

模型		平方和	df	均方	F	Sig.
1	回归	118413.803	7	16916.258	15.708	.000 ^a
	残差	12923.147	12	1076.929		
	总计	131336.950	19			

由上表可知 $F=15.708$ ， $F > F_{0.05}(7,20)=2.51$ ，回归方程显著，即自变量和因变量存在明显的函数关系。

C) 回归系数的显著性检验

表 16 回归系数表

模型	非标准化系数		标准系数	t	Sig.
	B	标准误差	试用版		
(常量)	797.579	108.653		7.341	.000
黄酮醇(mg/kg)	.862	.209	.471	4.130	.001
可溶性固形物 g/l	-1.820	.418	-.447	-4.357	.001
果穗质量/g	.385	.094	.807	4.106	.001
百粒质量/g	-1.081	.289	-.814	-3.746	.003
a*(+红；-绿)	-65.783	21.221	-.454	-3.100	.009
DPPH 半抑制体积 (IV50) 1/IV50(uL)	940.449	117.720	1.560	7.989	.000
L*(D65)	7.126	1.187	1.724	6.003	.000

表 16 给出了模型对 7 个变量的偏回归系数是否等于 0 的 T 检验结果。 t 值分别等于 7.341、4.130、-4.357、4.106、-3.746、-3.100、7.989 和 6.003，概率 p 值都小于显著性水平 0.05，因此认为偏相关系数 β 显著不等于 0，每一个指标都和因变量线性相关显著。

综上，酿酒葡萄和葡萄酒的理化指标与葡萄酒质量存在有密切的关系。

5.4.2 多元线性回归模型的论证

在 5.4.1 七元线性回归模型的基础上，将未选中的 7 个样本作为检验样本组，对模型的可行性进行验证，样本质量值的评分如下：

表 17 检验样本组 7 个样本的葡萄酒质量值

红葡萄	样品 6	样品 7	样品 11	样品 16	样品 17	样品 19	样品 27
计算值	733.96	654.25	709.45	740.12	796.04	761.78	821.37
评分值	722	715	701	749	793	786	730

套用 5.1.3 中的显著性差异模型，令 H_0 ：七元线性回归模型对品酒员的评价结果是相同的； H_1 ：两组品酒员对酒样本的评价结果是不同的。

利用统计软件 SPSS 进行 Wilcoxon 符号秩检验，结果如表 18 所示。

检验统计量 ^b	
	V2 - V1
Z	-.676 ^a
渐近显著性(双侧)	.499

a. 基于正秩。

b. Wilcoxon 带符号秩检验

由表 X 可知， $P > 0.05$ ，故接受原假设，认为酿酒葡萄和葡萄酒的理化指标与葡萄酒的质量之间的关系足够密切，通过 5.4.1 中得到的七元线性回归方程来评价葡萄酒质量的做法是可行的。

六、模型的评价、改进与推广

对于问题一，首先运用了配对样本 Wilcoxon 符号秩检验法对两组评酒员的评价结果进行判别是否有显著性差异。由于在此为主观评分结果的特殊性，类似于体育比赛中对运动员进行打分。比如高水平比赛中，由于被评价对象的水平比较接近，水平较差的评分者难以区别不出被评价对象水平的高低，而力求给保稳分。这种评分者的打分的分值波动小，具有一定的隐蔽性，用极端数据和偏差分析很难判断出来。但是，即使该某评分者总是力求给出中间分，其评分结果的序次很难与最后得分的序次具有较高的一致性。因此，采用秩相关分析具有相对的合理性及良好的可推广性。

其次，通过肯德尔和谐系数法分析评价结果的可信度的方法，比计算原始数据的矩相关系数更能反映评分者评分与最后得分之间的关系。模型的不足之处在于仅考虑了评分者评分与最后得分的一致性大小关系。若要全面衡量评分者评分的可信性以及合理性，还必须对评分者评分的相对稳定性作评价。可以考虑进行偏差分析（偏差分析可以很好地反映个评分者的稳定性好坏），即评分者的评分结果与最后得分之间距离

$$d^2 = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})',$$

通过每个评分者的一致性与稳定性建立综合评价模型

$$y = w_1(x_1) \bullet x_1 + w_2(x_2) \bullet x_2$$

对每一项赋予一定的权重值就可以计算出每个评分者的评价得分，这样才能综合的反映出每个评分者的可信度高低。

对于问题二，本文在建立层次结构时，在常规三个层次的基础上增加了两级准则层，从而有效地避免了单一层次分析法指标权重值偏离的现象。其次，运用层次分析法建立比较矩阵运用了主成分分析中的贡献度来赋值，有效地避免了主观感受对两指标间影响

程度进行赋值的人为因素。

对于问题三，能良好的反映出两组变量的指标之间多对多联系的典型相关分析，可推广性很强。例如，为了研究扩张性财政政策实施以后对宏观经济发展的影响，就需要考察有关财政政策的一系列指标如财政支出总额的增长率、财政赤字增长率、国债发行额的增长率、税率降低率等与经济发展的一系列指标如国内生产总值增长率、就业增长率、物价上涨率等两组变量之间的相关程度。

对于问题四，通过逐步回归分析将指标集中指标筛选剔除，余下能充分反映线性关系的小部分指标来得到相应结果。显然，逐步回归分析之前的指标集中的指标数越多，模型的效果会越好。因此，对于模型的改进，考虑加入附件3中葡萄和葡萄酒的芳香物质的指标。

七、参考文献

- [1] 百度百科，酿酒葡萄，<http://baike.baidu.com/view/2684347.htm>，2012年9月10日；
- [2] 百度百科，正态分布，<http://baike.baidu.com/view/45379.htm>，2012年9月9日；
- [3] 曾怀恩 黄声享，基于Kriging方法的空间数据插值研究，地球信息科学，10(7),2008；
- [4] 陈平 魏鹏超，两配对样本非参数检验在公司绩效评价中的应用探讨，《财会通讯》，2010年20期：59-60，2010年；
- [5] MBA智库百科，<http://wiki.mbalib.com/wiki/%E5%A8%81%E5%B0%94%E7%A7%91%E5%85%8B%E6%A3%AE%E7%AC%A6%E5%8F%B7%E7%A7%A9%E6%A3%80%E9%AA%8C>，2012年9月9日；
- [6] Stone H, Sidel JL, Oliver S et al. Sensory evaluation by quantitative descriptive analysis [J]. Food Technology, 1974, 28(11)24-34;
- [7] Stone H, Sidel JL, Bloomquist J. Quantitative descriptive analysis [J]. Cereal Foods World, 1980, 25; 624-634;
- [8] 司林波 黄钦，研究生招生面试评分信度模型分析，《中国高教研究》，2008年第7期：33-35,2008年；
- [9] 姜启源 谢金星 叶俊，数学模型（第三版）北京，高等教育出版社，2006.9.12；
- [10] 姜婧 张启平，典型相关分析的交叉效率模型及其在钢铁行业的应用，工业技术经济，2011年第1期（总第207期）：108-113，2011年1月。

八、附录

8.1 问题一附录

8.1.1 异常数据的修正

附件 1 第一组白葡萄酒品尝评分中品酒员 9 对样品 8 持久性的评分出现异常，经修正后分值为 6。

8.1.2 个葡萄酒样本评分数据概率分布确定的附录

1. P-P 图具体操作步骤如下：

打开数据文件，选择【分析】(Analyze)菜单，单击【描述统计】(Descriptive Statistics)命令下的【P-P 图】(P-P Plots) 命令。

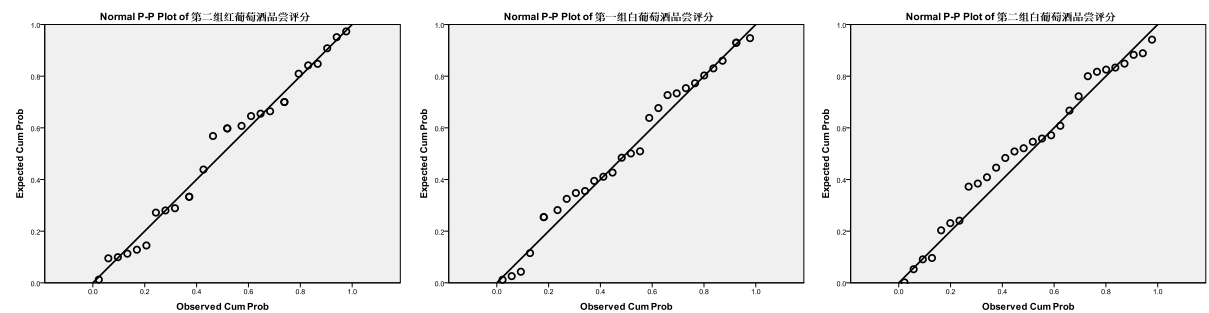


图 8.1 第二组红葡萄酒，第一组、第二组白葡萄酒评价结果的正态 P-P 图

2. 在 SPSS 中用非参数分析方法对数据资料进行正态性检验：

- 1) 读取文件第一组红葡萄酒品尝得分.xlsx。
- 2) 依次按 Aanlyze, Nonparametric Tests, 1-sample K-S 顺序逐一单击鼠标键，展开 One-sample Kolmogorov-Smirnov Test 对话框。
- 3) 选择 Varoooo2 变量进入 Test Variable 对话框。
- 4) 由于要对数据资料进行正态分布检验,故在 distribution 框中选中 Normal 复选项。

One-Sample Kolmogorov-Smirnov Test		
		第二组红葡萄酒 品尝得分
N		27
Normal Parameters ^{a,b}	Mean	70.5148
	Std. Deviation	3.97799
Most Extreme Differences	Absolute	.124
	Positive	.078
	Negative	-.124
Kolmogorov-Smirnov Z		.644
Asymp. Sig. (2-tailed)		.801

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test	
	第一组白葡萄品 尝得分

N		28
Normal Parameters ^{a,b}	Mean	74.1500
	Std. Deviation	4.84902
Most Extreme Differences	Absolute	.086
	Positive	.063
	Negative	-.086
Kolmogorov-Smirnov Z		.453
Asymp. Sig. (2-tailed)		.986

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		第二组白葡萄酒品 尝得分
N		28
Normal Parameters ^{a,b}	Mean	76.5321
	Std. Deviation	3.17094
Most Extreme Differences	Absolute	.122
	Positive	.076
	Negative	-.122
Kolmogorov-Smirnov Z		.648
Asymp. Sig. (2-tailed)		.796

a. Test distribution is Normal.

b. Calculated from data.

8.1.3 显著性差异评价附录

表 8.1 红葡萄酒显著性差异检验表

样本编号	第一组评分结果	第二组评分结果	分数差值	绝对差值	秩次	符号秩次-	符号秩次+
125	69.20	68.20	1.00	1.00	2		2
127	73.00	71.50	1.50	1.50	4		4
107	71.50	65.30	6.20	6.20	20		20
110	74.20	68.80	5.40	5.40	12.5		12.5
111	70.10	61.60	8.50	8.50	25.5		25.5
120	79.20	75.80	3.40	3.40	8		8
116	74.90	69.90	5.00	5.00	11		11
124	78.00	71.50	6.50	6.50	23		23
119	78.60	72.60	6.00	6.00	19		19
118	59.90	65.40	-5.50	5.50	14	14.00	
106	72.20	66.30	5.90	5.90	18		18
104	68.60	71.20	-2.60	2.60	6	6.00	
113	74.60	68.80	5.80	5.80	16.5		16.5
122	77.20	71.60	5.60	5.60	15		15

117	79.30	74.50	4.80	4.80	9		9
101	62.70	68.10	-5.40	5.40	12.5	12.50	
102	80.30	74.00	6.30	6.30	21.5		21.5
103	80.40	74.60	5.80	5.80	16.5		16.5
108	72.30	66.00	6.30	6.30	21.5		21.5
112	53.90	68.30	-14.40	14.40	27	27.00	
105	73.30	72.10	1.20	1.20	3		3
123	85.60	77.10	8.50	8.50	25.5		25.5
115	58.70	65.70	-7.00	7.00	24	24.00	
126	73.80	72.00	1.80	1.80	5		5
109	81.50	78.20	3.30	3.30	7		7
121	77.10	72.20	4.90	4.90	10		10
114	73.00	72.60	0.40	0.40	1		1
						83.50	294.5

表 8.2 白葡萄酒显著性差异检验表

样本编号	第一组品尝得分	第二组品尝得分	分数差值	绝对差值	秩次	符号秩次-	符号秩次+
226	81.30	74.30	7.00	7	21		21
205	71.00	81.50	-10.50	10.5	27	27	
204	79.40	76.90	2.50	2.5	10		10
223	75.90	77.40	-1.50	1.5	4.5	4.5	
220	77.80	76.60	1.20	1.2	3		3
219	77.20	76.40	0.80	0.8	1		1
228	81.30	79.60	1.70	1.7	7		7
211	72.30	71.40	0.90	0.9	2		2
215	72.40	78.40	-6.00	6	19	19	
214	72.00	77.10	-5.10	5.1	17	17	
212	63.30	72.40	-9.10	9.1	26	26	
218	73.10	76.70	-3.60	3.6	15	15	
213	65.90	73.90	-8.00	8	24	24	
221	76.40	79.20	-2.80	2.8	12.5	12.5	
210	74.30	79.80	-5.50	5.5	18	18	
207	77.50	74.20	3.30	3.3	14		14
201	82.00	77.90	4.10	4.1	16		16
203	78.20	75.60	2.60	2.6	11		11
216	74.00	67.30	6.70	6.7	20		20
202	74.20	75.80	-1.60	1.6	6	6	
206	68.40	75.50	-7.10	7.1	22	22	
222	71.00	79.40	-8.40	8.4	25	25	
208	70.40	72.30	-1.90	1.9	8	8	
217	78.80	80.30	-1.50	1.5	4.5	4.5	
209	72.90	80.40	-7.50	7.5	23	23	
225	77.10	79.50	-2.40	2.4	9	9	

224	73.30	76.10	-2.80	2.8	12.5	12.5	
227	64.80	77.00	-12.20	12.2	28	28	
						301	105

8.1.4 秩相关分析评价模型附录

表 8.3 第一组红葡萄酒秩变换结果

	1	2	3	4	5	6	7	8	9	10
1	4	13	19	8	11	8	4	6	14	2
2	11	11	5	11	17	16	14	10	23	11
3	5	8	14	12	5	22	11	2	20	26
4	8	19	20	16	13	15	16	11	7	12
5	18	3	11	10	6	10	9	9	27	7
6	25	21	15	23	19	18	18	15	25	24
7	16	16	18	21	8	11	24	13	9	9
8	12	24	25	17	23	23	10	14	10	5
9	23	22	21	14	7	24	25	20	18	25
10	6	5	2	2	2	2	3	1	3	3
11	17	7	9	6	20	9	6	8	15	27
12	2	4	6	15	4	20	19	7	19	18
13	9	23	17	5	12	17	22	16	5	19
14	19	20	12	18	25	13	17	19	11	23
15	13	15	26	19	27	21	7	22	16	13
16	1	6	3	1	15	6	12	4	4	1
17	15	18	22	24	24	19	27	21	24	8
18	26	25	24	25	9	25	15	23	21	14
19	7	10	7	13	14	14	8	25	6	15
20	3	1	1	3	3	4	1	5	1	4
21	22	9	13	9	21	7	5	24	17	6
22	27	26	23	26	26	27	26	26	22	21
23	10	2	4	4	1	1	2	3	2	16
24	20	17	10	7	16	12	13	17	12	20
25	24	14	16	27	22	26	20	27	13	22
26	21	27	27	22	10	5	23	12	26	10
27	14	12	8	20	18	3	21	18	8	17

表 8.4 第二组红葡萄酒秩变换结果

	1	2	3	4	5	6	7	8	9	10
1	6	20	23	8	11	2	5	8	4	20
2	23	24	18	27	16	24	24	6	27	25
3	27	14	19	26	12	18	13	22	21	26
4	9	5	2	14	1	10	1	16	16	6
5	10	17	20	2	3	19	11	11	14	7
6	13	6	15	17	14	13	6	18	8	11

7	17	21	8	23	26	21	22	19	25	9
8	7	7	21	4	8	1	2	17	1	3
9	24	22	22	16	22	27	23	27	15	14
10	8	11	9	5	13	14	7	9	11	17
11	2	2	1	9	2	3	4	1	6	2
12	1	1	6	3	5	11	12	4	9	12
13	14	3	4	18	20	12	25	12	17	13
14	11	9	25	12	24	16	21	24	3	22
15	18	8	24	6	15	26	20	25	22	18
16	15	18	16	13	17	20	19	26	18	21
17	3	10	10	7	4	4	9	3	7	8
18	12	12	26	10	6	5	8	13	5	5
19	20	25	7	20	7	22	16	14	2	15
20	4	13	14	24	27	15	14	10	23	10
21	21	23	13	19	18	17	26	15	19	19
22	26	27	27	25	21	25	27	23	24	23
23	25	19	11	21	9	23	3	5	20	27
24	5	15	5	22	25	7	15	20	26	16
25	22	26	12	11	19	8	17	7	12	24
26	16	16	17	1	10	9	18	2	10	1
27	19	4	3	15	23	6	10	21	13	4

表 8.5 第一组白葡萄酒秩变换结果

	1	2	3	4	5	6	7	8	9	10
1	9	18	9	27	28	25	17	20	24	21
2	22	9	2	7	19	7	6	17	13	6
3	10	25	5	12	15	22	28	23	17	26
4	7	19	6	19	22	21	11	15	19	8
5	18	22	10	23	25	20	21	2	20	14
6	11	20	12	18	9	9	14	6	4	9
7	5	24	25	20	27	24	22	21	21	27
8	16	7	18	8	10	13	25	3	22	4
9	8	11	23	24	23	6	13	8	7	18
10	14	12	17	16	20	10	12	10	15	5
11	3	2	1	2	3	8	9	1	3	2
12	27	5	13	25	7	14	23	7	8	15
13	20	3	14	1	2	11	7	4	1	1
14	23	14	19	6	13	23	18	13	23	22
15	12	4	20	9	26	16	19	16	26	3
16	24	28	15	15	11	19	15	11	11	19
17	25	26	24	11	14	28	20	27	28	12
18	26	21	26	28	16	17	1	26	25	13
19	6	15	21	13	5	26	26	5	16	10

20	15	10	22	3	21	27	24	12	5	16
21	2	6	16	14	17	3	16	9	2	23
22	4	13	27	5	8	18	8	14	14	7
23	13	8	7	4	24	4	4	22	9	20
24	19	16	28	22	18	15	27	18	27	28
25	17	23	8	10	6	1	5	28	10	24
26	28	27	11	21	12	12	10	24	12	17
27	21	17	3	26	4	5	3	25	18	25
28	1	1	4	17	1	2	2	19	6	11

表 8.6 第二组白葡萄酒秩变换结果

	1	2	3	4	5	6	7	8	9	10
1	18	3	7	20	6	16	21	4	3	16
2	2	1	8	1	2	22	1	23	1	2
3	26	4	1	26	18	7	10	2	24	8
4	5	25	2	11	3	8	11	1	4	21
5	15	27	14	2	1	2	18	8	2	10
6	20	22	16	23	24	3	12	6	27	23
7	12	14	28	27	26	27	27	5	22	17
8	27	5	9	10	28	12	5	9	9	9
9	16	8	17	25	25	26	22	3	28	22
10	24	15	3	21	4	13	25	16	17	6
11	28	11	11	19	22	23	26	10	10	27
12	17	9	12	24	13	9	13	7	14	4
13	8	12	10	8	15	14	7	21	19	19
14	21	6	4	4	10	4	14	20	18	13
15	4	19	26	12	12	17	4	13	23	14
16	9	28	23	13	8	19	6	24	12	3
17	22	13	27	14	27	18	15	14	15	5
18	25	16	24	7	19	20	19	22	13	7
19	3	17	18	17	16	1	2	12	5	18
20	13	2	19	22	20	24	28	28	20	25
21	10	26	22	18	17	21	8	25	16	26
22	19	10	25	28	11	28	16	17	6	24
23	1	23	13	9	21	6	23	15	7	20
24	6	18	5	3	7	5	9	18	11	1
25	11	7	15	6	23	11	24	19	21	12
26	23	20	20	15	14	25	20	26	26	28
27	14	21	6	5	5	15	17	11	8	15
28	7	24	21	16	9	10	3	27	25	11

8.2 问题二附录

8.2.1 指标的遴选附录

1. 红葡萄主成分分析正交解如下:

Rotated Component Matrix^a

	Component							
	1	2	3	4	5	6	7	8
葡萄总黄酮 (mmol/kg)	.893	.015	-.050	.055	-.031	.106	.095	.031
总酚(mmol/kg)	.888	.136	.057	-.111	-.111	.185	.168	.005
DPPH 自由基 1/IC50 (g/L)	.836	-.169	-.022	.037	.025	.226	.344	.050
单宁(mmol/kg)	.813	.008	.230	-.052	-.084	-.208	.238	.067
出汁率(%)	.739	.041	.040	-.038	-.130	.056	-.181	-.087
花色苷 mg/100g 鲜重	.705	.059	.565	-.131	-.113	.168	.095	.071
蛋白质 mg/100g	.568	-.260	-.129	-.011	-.087	.450	.282	.336
L*	-.553	-.362	-.142	.289	.191	.273	.189	-.238
总糖 g/L	.057	.922	-.011	-.100	.073	.053	.009	-.074
可溶性固形物 g/l	.068	.909	.081	.002	-.039	-.059	.015	-.073
干物质含量 g/100g	.114	.905	.133	-.164	.021	-.147	.038	.175
还原糖 g/L	-.148	.783	.017	-.115	.013	-.058	-.039	.084
氨基酸总量 mg/100gfw	.136	.622	-.332	-.332	.031	-.064	.409	.137
VC 含量 (mg/L)	.055	-.472	-.135	-.170	.119	-.139	.121	-.456
苹果酸 (g/L)	.168	.168	.829	-.221	.265	-.026	-.159	.025
褐变度 ΔA/g*g*min*ml	.262	-.105	.825	-.185	-.092	.113	.273	.012
多酚氧化酶活力 E (A/min·g·ml)	-.043	.169	.679	.156	-.297	.106	.288	.055
果皮质量 (g)	-.005	-.021	.042	.882	-.208	-.239	.132	-.080
果穗质量/g	-.044	-.261	-.109	.802	-.065	.172	-.181	.101
百粒质量/g	-.093	-.194	-.196	.773	-.110	-.160	-.215	-.234
a*(+红; -绿)	-.322	.009	-.084	-.101	.896	.047	-.077	.028
b*(+黄; -蓝)	-.148	.240	.094	-.048	.870	-.035	-.231	.238
白藜芦醇(mg/kg)	.011	-.216	-.064	-.235	.836	.210	.279	-.036
可滴定酸 (g/l)	-.071	.350	-.117	.124	-.092	-.849	-.116	-.145
固酸比	.116	.106	.256	-.049	.116	.806	-.043	.185
PH 值	.324	.043	-.514	-.068	-.028	.668	.042	-.121
黄酮醇(mg/kg)	.225	.169	.141	.003	.010	.031	.872	.105
果梗比(%)	.246	-.187	.231	-.380	-.117	.096	.698	-.004
酒石酸 (g/L)	.206	.105	-.285	-.178	.092	.226	.099	.793
柠檬酸 (g/L)	-.025	-.002	.409	-.094	.207	.033	.111	.769

2. 白葡萄相关性系数表见附件 8

3. 白葡萄主成分分析

表 8.7 白葡萄特征值和累计贡献率

主成分	提取平方和载入	旋转平方和载入
-----	---------	---------

	合计	方差的%	累计%	合计	方差的%	累计%
1	5.830	19.434	19.434	4.624	15.412	15.412
2	4.927	16.422	35.856	3.282	10.940	26.352
3	3.630	12.102	47.958	3.108	10.358	36.711
4	2.081	6.935	54.893	2.963	9.877	46.587
5	1.889	6.297	61.190	2.080	6.932	53.519
6	1.655	5.516	66.705	2.059	6.862	60.381
7	1.523	5.077	71.782	1.864	6.215	66.596
8	1.280	4.266	76.049	1.850	6.167	72.763
9	1.238	4.127	80.175	1.825	6.083	78.846
10	1.002	3.340	83.516	1.401	4.670	83.516

在累计方差为 83.516%的前提下分析得到十个主成分，这十个主成分提供了附件 2 酿酒白葡萄的理化指标中 83.516%的信息，满足主成分分析原则。

由以上分析利用 SPSS 统计软件计算得到主成分分析正交解如下：

Rotated Component Matrix^a

	Component									
	1	2	3	4	5	6	7	8	9	10
可溶性固形物 g/l	.896	-.014	.130	.216	.128	.086	.069	.001	-.136	.131
总糖 g/L	.833	-.070	.112	.059	-.005	.009	-.020	-.023	.004	.281
干物质含量 g/100g	.792	.071	.084	.069	-.147	.215	-.352	.005	.147	-.056
还原糖 g/L	.786	-.087	.055	-.145	-.276	.109	-.154	.088	.198	-.052
出汁率(%)	-.657	-.162	-.067	-.285	.002	.238	-.077	-.169	-.186	.419
果穗质量/g	-.630	.144	-.115	-.146	-.289	.055	.370	-.076	.301	-.019
总酚(mmol/kg)	-.217	.889	.086	-.039	-.167	.167	.093	.156	.082	.106
葡萄总黄酮 (mmol/kg)	-.233	.881	.069	-.065	-.234	.154	.162	.042	.093	-.043
单宁(mmol/kg)	.436	.781	.032	-.101	-.029	-.108	-.110	-.161	-.070	.077
a*(+红；-绿)	.004	-.015	-.886	.276	.105	-.035	-.074	.004	.127	-.023
L*	.099	.148	.854	-.107	.093	.054	-.285	.118	.054	.083
b*(+黄；-蓝)	.314	-.162	.806	-.128	.095	.133	-.273	.187	.049	.158
DPPH 自由基 1/IC50 (g/L)	.299	.403	.565	-.071	-.038	-.219	.245	-.039	.081	-.066
可滴定酸 (g/l)	.007	.099	.251	-.860	.105	.047	.165	-.089	.073	.252
固酸比	.147	-.084	-.266	.856	-.088	.063	-.091	.085	-.070	-.200
PH 值	.160	.033	-.052	.789	.088	-.082	-.066	-.160	.135	.384
果梗比(%)	-.079	-.318	-.067	-.169	.773	-.019	-.209	-.001	.093	-.084
VC 含量 (mg/L)	-.243	.006	.205	-.200	.653	-.376	.079	.081	-.243	-.010
酒石酸	.292	-.109	.019	.412	.627	.487	-.012	-.036	-.021	.027
蛋白质 mg/100g	.002	.412	.120	-.161	-.490	.014	.018	.466	.076	.273
苹果酸	-.083	.190	.110	-.283	-.071	.808	-.045	-.083	.141	.050
柠檬酸	.185	.046	-.082	.169	-.017	.582	.060	.278	-.223	-.197
氨基酸总量 mg/100g fw	.420	.124	.292	.139	-.055	.518	.140	-.199	.362	-.001

果皮质量 (g)	-.128	.188	-.159	-.031	-.188	.219	.806	-.105	.009	.088
百粒质量/g	-.203	-.041	-.071	-.352	.054	-.208	.714	.056	-.136	-.028
褐变度 Δ A/g*g*min*ml	.035	-.052	.183	-.002	-.001	-.058	-.062	.875	-.061	.144
黄酮醇(mg/kg)	.070	.585	-.119	.160	.074	.194	-.022	.655	.158	-.100
白藜芦醇(mg/kg)	-.024	-.019	-.155	-.037	-.037	.028	-.107	.024	.892	-.081
多酚氧化酶活力 E (A/min·g·ml)	-.170	-.274	-.328	-.041	.020	-.026	-.044	.004	-.650	-.303
花色苷 mg/100g 鲜 重	-.234	-.131	-.203	.141	.175	.121	-.092	-.325	-.066	-.745

正交解说明, 白葡萄理化指标当中, 主成分 1 为可溶性固形物、总糖、干物质含量和还原糖的组合, 主成分 2 为总酚、葡萄总黄酮和单宁的组合, 主成分 3 为红绿色差指标 a 值、黄蓝色差指标 b 值和亮度值 L 的组合, 主成分 4 为可滴定酸、固酸比和 PH 值的组合, 主成分 5 为 VC 含量, 主成分 6 为苹果酸, 主成分 7 为果皮质量, 主成分 8 为褐变度, 主成分 9 为白藜芦醇, 主成分 10 为花色苷。这组合说明可溶性固形物、总糖、干物质含量和还原糖, 总酚、葡萄总黄酮和单宁, 红绿色差指标 a 值、黄蓝色差指标 b 值和亮度值 L, 可滴定酸、固酸比和 PH 值可能在同一方面对酿酒葡萄分级起重要作用, 而 VC 含量、苹果酸、果皮质量、褐变度、白藜芦醇、花色苷分别在不同角度影响酿酒葡萄的分级。

4. 各元素间的比较矩阵

	葡萄酒质量	葡萄理化指标
葡萄酒质量	1	1
葡萄理化指标	1	1

	外观分析	香气分析	口感分析	平衡/整体评价
外观分析	1.00	0.50	0.34	1.36
香气分析	2.00	1.00	1.50	2.73
口感分析	2.93	0.67	1.00	4.00
平衡/整体评价	0.73	0.37	0.25	1.00

	澄清度	色调
澄清度	1	0.5
色调	2	1

	纯正度	浓度	质量
纯正度	1.00	0.75	0.38
浓度	1.33	1.00	0.50
质量	2.67	2.00	1.00

酿红葡萄酒葡萄的各因子比较矩阵

	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6	因子 7	因子 8
因子 1	1	1.165489	1.657066	1.91592	1.931088	2.025023	2.302007	2.734565

因子 2	0.858009	1	1.421778	1.643877	1.656891	1.737488	1.975143	2.346281
因子 3	0.603476	0.703345	1	1.156212	1.165366	1.222053	1.389206	1.650245
因子 4	0.521942	0.608318	0.864893	1	1.007917	1.056946	1.201515	1.427286
因子 5	0.517843	0.60354	0.8581	0.992145	1	1.048644	1.192078	1.416075
因子 6	0.493821	0.575543	0.818295	0.946122	0.953613	1	1.136781	1.350387
因子 7	0.434404	0.506292	0.719835	0.832282	0.838872	0.879677	1	1.187905
因子 8	0.365689	0.426206	0.605971	0.700631	0.706178	0.740529	0.841818	1

	因子 11	因子 12	因子 13	因子 14
因子 11	1	1.005541	1.068311	1.098709
因子 12	0.99449	1	1.062424	1.092655
因子 13	0.936057	0.941244	1	1.028455
因子 14	0.910159	0.915202	0.972333	1

	因子 31	因子 32
因子 31	1	1.005724
因子 32	0.994308	1

	因子 21	因子 22	因子 23
因子 21	1	1.015368	1.018881
因子 22	0.984865	1	1.00346
因子 23	0.981469	0.996552	1
	因子 41	因子 42	
因子 41	1	1.098962	
因子 42	0.90995	1	

	因子 51	因子 52
因子 51	1	1.030308
因子 52	0.970584	1

	因子 61	因子 62
因子 61	1	1.053778
因子 62	0.948966	1

5、酿白葡萄酒葡萄的因子比较矩阵

	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6	因子 7	因子 8	因子 9	因子 10
因子 1	1	1.4087 25	1.4878 59	1.5604 3	2.2233 26	2.2460 34	2.4798 42	2.4991 91	2.5335 42	3.3004 59
因子 2	0.7098 62	1	1.0561 74	1.1076 89	1.5782 54	1.5943 73	1.7603 45	1.7740 8	1.7984 65	2.3428 69
因子 3	0.6721 07	0.9468 14	1	1.0487 76	1.4943 12	1.5095 75	1.6667 19	1.6797 24	1.7028 11	2.2182 61

因子 4	0.6408 49	0.9027 8	0.9534 93	1 16	1.4248 16	1.4393 69	1.5892 04	1.6016 04	1.6236 18	2.1150 96
因子 5	0.4497 77	0.6336 12	0.6692 04	0.7018 45	1 14	1.0102 14	1.1153 75	1.1240 78	1.1395 28	1.4844 69
因子 6	0.4452 29	0.6272 06	0.6624 38	0.6947 49	0.9898 9	1 1	1.1040 98	1.1127 13	1.1280 07	1.4694 61
因子 7	0.4032 51	0.5680 7	0.5999 81	0.6292 46	0.8965 59	0.9057 16	1 1	1.0078 03	1.0216 55	1.3309 15
因子 8	0.4001 29	0.5636 72	0.5953 36	0.6243 74	0.8896 18	0.8987 04	0.9922 58	1 1	1.0137 45	1.3206 11
因子 9	0.3947 04	0.5560 3	0.5872 64	0.6159 08	0.8775 56	0.8865 19	0.9788 04	0.9864 42	1 1	1.3027 05
因子 10	0.3029 88	0.4268 27	0.4508 04	0.4727 92	0.6736 41	0.6805 22	0.7513 63	0.7572 25	0.7676 33	1 1

	因子 11	因子 12	因子 13	因子 14
因子 11	1	1.07624	1.13187	1.140677
因子 12	0.929161	1	1.051689	1.059872
因子 13	0.883494	0.950851	1	1.007781
因子 14	0.876673	0.94351	0.992279	1

	因子 21	因子 22	因子 23
因子 21	1	1.009408	1.138373
因子 22	0.990679	1	1.127762
因子 23	0.878447	0.886712	1

	因子 31	因子 32	因子 33
因子 31	1	1.037754	1.099283
因子 32	0.96362	1	1.059291
因子 33	0.909684	0.944028	1

	因子 41	因子 42	因子 43
因子 41	1	1.004695	1.089718
因子 42	0.995327	1	1.084626
因子 43	0.917669	0.921977	1

6. 红酿酒葡萄的权重

葡萄质量	葡萄酒质量	0.5	外观分析	0.15203	0.076015	澄清度	0.333333	0.025338	
						色调	0.666667	0.050677	
			香气分析	0.37031	0.185155	纯正度	0.2	0.037031	
						浓度	0.266667	0.049375	

						质量	0.533333	0.098749	
			口感分析	0.366172	0.183086	纯正度	0.136364	0.024966	
						浓度	0.181818	0.033288	
						持久性	0.181818	0.033288	
						质量	0.5	0.091543	
			平衡/整体评价	0.111489	0.055744				
	葡萄理化指标	0.5	因子 1	0.208543	0.104271	因子 11	0.260369	0.027149	葡萄总黄酮
						因子 12	0.258934	0.026999	总酚
						因子 13	0.24372	0.025413	DPPH 自由基
						因子 14	0.236977	0.02471	单宁
			因子 2	0.178931	0.089466	因子 21	0.337117	0.03016	总糖
						因子 22	0.332014	0.029704	可溶性固形物
						因子 23	0.330869	0.029601	干物质含量
			因子 3	0.12585	0.062925	因子 31	0.501427	0.031552	苹果酸
						因子 32	0.498573	0.031373	褐变度
			因子 4	0.108847	0.054424	因子 41	0.523574	0.028495	果皮质量
						因子 42	0.476426	0.025929	果穗质量
			因子 5	0.107992	0.053996	因子 51	0.507464	0.027401	a*(+红; -绿)
						因子 52	0.492536	0.026595	b* (+黄;-蓝)
			因子 6	0.102983	0.051491	因子 61	0.513092	0.02642	可滴定酸
						因子 62	0.486908	0.025072	固酸比
			因子 7	0.090592	0.045296	因子 71		0.045296	黄酮醇
			因子 8	0.076262	0.038131	因子 81		0.038131	酒石酸 (g/L)

7. 白酿酒葡萄的权重

葡萄质量	葡萄酒质量	0.5	外观分析	0.15203	0.076015	澄清度	0.333333	0.025338
						色调	0.666667	0.050677
			香气分析	0.37031	0.185155	纯正度	0.2	0.037031
						浓度	0.266667	0.049375
						质量	0.533333	0.098749
			口感分析	0.366172	0.183086	纯正度	0.136364	0.024966
						浓度	0.181818	0.033288
						持久性	0.181818	0.033288
			平衡/整体评价	0.111489	0.055744			0.055744
	葡萄理化指标	0.5	因子 1	0.208543	0.104271	因子 11	0.260369	0.027149
						因子 12	0.258934	0.026999
						因子 13	0.24372	0.025413
						因子 14	0.236977	0.02471
			因子 2	0.178931	0.089466	因子 21	0.337117	0.03016
						因子 22	0.332014	0.029704
						因子 23	0.330869	0.029601

			因子 3	0.12585	0.0629 25	因子 31	0.501427	0.031552
						因子 32	0.498573	0.031373
			因子 4	0.10884 7	0.0544 24	因子 41	0.523574	0.028495
						因子 42	0.476426	0.025929
			因子 5	0.10799 2	0.0539 96	因子 51	0.507464	0.027401
						因子 52	0.492536	0.026595
			因子 6	0.10298 3	0.0514 91	因子 61	0.513092	0.02642
						因子 62	0.486908	0.025072
			因子 7	0.09059 2	0.0452 96	因子 71	1	0.045296
			因子 8	0.07626 2	0.0381 31	因子 81	1	0.038131

8、白葡萄各影响分级程度指标的权重汇总如下：

澄清度	色调	纯正度	浓度	质量	纯正度
0.025338	0.050677	0.037031	0.049375	0.098749	0.024966
浓度	持久性	质量	整体评价	可溶性固形物	总糖
0.033288	0.033288	0.091543	0.055744	0.02501	0.023238
干物质含量	还原糖 g/L	总酚(mmol/kg)	葡萄总黄酮	单宁(mmol/kg)	a*(+红；-绿)
0.022096	0.021925	0.022829	0.022616	0.020054	0.021583
L*	b*（+黄;-蓝）	可滴定酸（g/l）	固酸比	PH 值	VC 含量（mg/L）
0.020798	0.019634	0.020299	0.020204	0.018628	0.041501
苹果酸	果皮质量（g）	褐变度 Δ	白藜芦醇(mg/kg)	花色苷	
0.041081	0.037208	0.03692	0.036419	0.027957	

9、随机一致性指标

<i>n</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>RI</i>	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.54	1.56	1.58	1.59

8.2.2 分级评价模型的建立附录

1. 红葡萄标准化处理结果和白葡萄标准化处理结果见附件 9
2. 白葡萄酒中各个指标对目标层酿酒葡萄的权重为下表所示：

澄清度	色调	纯正度	浓度	质量	纯正度	浓度	持久性
0.0253	0.05068	0.037031	0.049375	0.098749	0.024966	0.033288	0.033288351
质量	整体评价	可溶性固形物	总糖 g/L	干物质含量	还原糖	总酚	葡萄总黄酮
0.0915	0.05574	0.02501	0.023238	0.022096	0.021925	0.022829	0.022616025
单宁	a*(+红；-绿)	L*	b*（+黄;-蓝）	可滴定酸	固酸比	PH 值	VC 含量（mg/L）
0.0201	0.02158	0.020798	0.019634	0.020299	0.020204	0.018628	0.04150076
苹果酸	果皮质量	褐变度	白藜芦醇	花色苷鲜重			
0.0411	0.03721	0.03692	0.036419	0.027957			

3. 白葡萄各样本综合评价指标值

表 8.8 白葡萄各样本综合评价指标值

葡萄样品	1	2	3	4	5	6	7
评价值	0.70	0.54	0.59	0.66	0.41	0.41	0.62
分级	1 st	3 rd	3 rd	2 ^{se c}	4 th	4 th	2 ^{se c}
葡萄样品	8	9	10	11	12	13	14
评价值	0.41	0.48	0.56	0.43	0.25	0.36	0.44
分级	4 th	4 th	3 rd	4 th	6 th	5 th	4 th
葡萄样品	15	16	17	18	19	20	21
评价值	0.56	0.47	0.69	0.53	0.46	0.55	0.56
分级	3 rd	4 th	2 ^{se c}	3 rd	4 th	3 rd	3 rd
葡萄样品	22	23	24	25	26	27	28
评价值	0.50	0.51	0.48	0.56	0.64	0.31	0.69
分级	3 rd	3 rd	4 th	3 rd	2 ^{se c}	5 th	2 ^{se c}

8.3 问题三附录

8.3.1 典型相关分析

```
proc cancorr data=_proj_.fkk;
var _COL1 _COL2 _COL3 _COL4
_COL5 _COL6 _COL7 _COL8 _COL9
_COL10 _COL11
_COL12 _COL13 _COL14 _COL15
_COL16 _COL17 _COL18 _COL19
_COL20 _COL21
_COL22 _COL23 _COL24 _COL25
_COL26 _COL27 L_ _COL29 _COL30;
with _COL31 _COL32 _COL33
_COL34 _COL35 _COL36 L_ D65_
A_ D65_ B_ D65_;
run; quit;
```

Canonical Correlation Analysis

Raw Canonical Coefficients for the VAR Variables

V1	V2	V3
V4	V5	

_COL1	氨基酸总量mg/100gfw
2.8375621E-6	0.0000563499
0.0002130867	-0.000208389
0.0001522384	
_COL2	蛋白质mg/100g
0.0009555603	-0.000514716
-0.001446732	0.0013079563
-0.012042568	
_COL3	VC含量（mg/L）
0.0231640381	0.1009254687
0.0693214075	-0.027772247
-0.250579582	
_COL4	花色苷mg/100g鲜重
0.0057075879	-0.003552114
0.011282816	-0.000470638
-0.000802931	
_COL5	酒石酸（g/L）
0.003447964	0.0893729646
0.0147415535	0.1231684506
0.0094209531	
_COL6	苹果酸（g/L）
0.034841638	-0.184531226
-0.103207119	-0.01186258
-0.052464748	
_COL7	柠檬酸（g/L）
0.026021002	-0.04650814
-0.045045272	-0.074610922
-0.134186735	
_COL8	多酚氧化酶活力E（A/min g ml）
-0.005421559	-0.003098431
0.0210185967	0.0069646102
	0.0164462499
_COL9	褐变度ΔA/g*g*min*ml
-0.000035288	-0.000515777
0.0001822099	-0.000641781
0.0015541754	
_COL10	DPPH自由基1/IC50（g/L）
0.1680045671	-0.623117175
-0.902937743	-2.730587132
-0.010176605	
_COL11	总酚（mmol/kg）
-0.008718836	0.059861779
0.1136482501	0.20898738
0.0621219707	
_COL12	单宁（mmol/kg）
0.0154013992	-0.067340073
-0.055177023	-0.073956548

8.3.2 Sas输出结果

The CANCERR		
Procedure	Canonical Correlation	
Analysis	Adjusted Square Canonical	
Approximate Canonical	Correlation	
Standard Error	Canonical Correlation	Adjusted Square Canonical Correlation
1	0.995211	0.991933
0.001300	0.990446	
2	0.969058	0.945198
0.008291	0.939074	
3	0.956550	0.930115
0.011569	0.914988	
4	0.931041	0.892306
0.018121	0.866837	
5	0.865289	0.762136
0.034194	0.748725	
6	0.845662	.
0.038764	0.715145	
7	0.758226	0.646146
0.057848	0.574906	
8	0.611388	0.404807
0.085216	0.373795	
9	0.490927	0.245674
0.103286	0.241010	
Test of H0: The canonical correlations in the		
Eigenvalues of Inv(E)*H		
current row and all that follow are zero		
=		
CanRsqr/(1-CanRsqr)		
Likelihood	Approximate Eigenvalue	Difference Ratio
Proportion Cumulative		
F Value Num DF Den DF		Pr > F

Multivariate Statistics and F Approximations

S=9	M=10	N=7
Value DF	F Value Pr > F	Statistic Num DF Den
		Wilks' Lambda
0.00000010	3.25	270
159.84	<.0001	
		Pillai's Trace
6.36492538	1.93	270
216	<.0001	
		Hotelling-Lawley
Trace	144.10948086	7.72
270	62.846	<.0001
		Roy's Greatest
Root	103.66636934	82.93
30	24	<.0001

NOTE:

F Statistic for Roy's Greatest Root is an upper bound.

The CANCERR Procedure

0.0172989176
_COL13 葡萄糖黄酮 (mmol/kg)
-0.039628792 0.1251551601
0.1005154541 -0.25828439
0.0358632638
_COL14 白藜芦醇(mg/kg)
-0.020160726 0.0433284089
0.0441190486 0.0262898964
0.1181781911
_COL15 黄酮醇(mg/kg)
0.0052897574 0.0048025136
-0.004089473 0.024204594
-0.02359164
_COL16 总糖/g/L
-0.003107495 -0.002103545
-0.009665785 -0.002546521
0.0299471534
_COL17 还原糖g/L
-0.005296203 0.0059293532
0.0095201495 0.0162300726
0.001971925
_COL18 可溶性固形物g/l
0.0130294586 0.0115345282
0.0059397209 0.0052844102
-0.051652208
_COL19 PH值
-0.157297953 -0.753897735
-0.975190464 -0.005747072
-0.973984275
_COL20 可滴定酸 (g/l)
-0.178743314 0.2328972728
0.0199142208 0.0419597797
0.2005696429
_COL21 固酸比
-0.029468148 0.0148145964
0.0196704111 0.0158851081
0.0387872046
_COL22 干物质含量g/100g
0.0204140059 -0.06719046
-0.09956079 0.0000795522
0.2026549679
_COL23 果穗质量/g
0.0009226424 -0.000328171
-0.000664587 0.0008421493
0.0021893876
_COL24 百粒质量/g
-0.00033491 -0.000296952
0.0003315596 -0.000071839
0.0002491003
_COL25 果梗比(%)
0.0074536305 -0.177580672
-0.271073317 -0.126459027
0.347040685
_COL26 出汁率(%)
-0.00778194 0.0075592687
0.0210159944 0.0197597583
-0.073640792
_COL27 果皮质量 (g)
-1.01754365 -2.556461176
0.8284266413 -1.35887675
4.5100103523
L_ L*
-0.04688782 -0.052072948
0.1328248589 -0.023565972
-0.058355115
_COL29 a*(+红; -绿)
-0.03449403 0.0037690212
0.1158270314 -0.070639233
0.0043001704
_COL30 b* (+黄; 蓝)
-0.020426919 0.0090690615
0.0512087037 0.017139359
0.0765115107

Raw Canonical Coefficients for the VAR
Variables

V6 V7
V8 V9
_COL1 氨基酸总量mg/100gfw
-0.000239481 0.0004071198
-0.000032928 -0.000718546
_COL2 蛋白质mg/100g
0.0116838068 0.0030974237
-0.01035539 -0.008476989
_COL3 VC含量 (mg/L)
0.0269959218 -0.076061212
-0.055386882 -0.110264252
_COL4 花色苷mg/100g鲜重
-0.008124946 0.0002527806
-0.003231182 0.0024635893
_COL5 酒石酸 (g/L)
-0.029010956 -0.143943668
-0.021212028 0.191191934
_COL6 苹果酸 (g/L)
0.1739926597 0.0588939728
-0.035835738 -0.048721729
_COL7 柠檬酸 (g/L)
-0.047018075 0.2230302713
-0.094338768 -0.326667871
_COL8 多酚氧化酶活力E
(A/min g ml)
0.0074824195 0.0095731728
-0.056301071 0.0070552578
_COL9 褐变度ΔA/g*g*min*ml
0.0003145178 0.0011649921
0.0018407112 -0.000587342
_COL10 DPPH自由基1/IC50(g/L)
-0.856834245 4.8816683478
-4.409256605 -1.317573453
_COL11 总酚(mmol/kg)
-0.049881234 -0.257250724
0.0206896625 0.1675036993
_COL12 单宁(mmol/kg)
-0.052265987 -0.019874045
0.0206494826 0.0300965612
_COL13 葡萄糖黄酮 (mmol/kg)
0.1304343144 0.1369834158
0.0852500047 -0.233168995
_COL14 白藜芦醇(mg/kg)
0.0363066514 0.036216309
0.0094319299 0.0341342917
1
3
The CANCELL Procedure
Canonical Correlation Analysis
Raw Canonical Coefficients for the VAR
Variables
V6 V7
V8 V9
_COL15 黄酮醇(mg/kg)
-0.00615021 -0.00804198
0.0022308587 0.0291044266
_COL16 总糖g/L
-0.02913582 0.00352834
-0.001697994 0.0001883679
_COL17 还原糖g/L
0.0027112116 -0.018585038
0.0096594032 -0.014527302
_COL18 可溶性固形物g/l
0.0238487147 -0.012187349
0.0081087171 0.0527114293
_COL19 PH值
0.4804059706 3.7396620079
1.4635188998 -2.643304062

_COL20 可滴定酸 (g/l)
0.2842953893 0.9434600585
0.2339691226 -0.624661207
_COL21 固酸比
0.0483526951 0.0850373153
0.0020691772 -0.083561914
_COL22 干物质含量g/100g
0.2557711822 0.2907024422
-0.126134843 -0.003465728
_COL23 果穗质量/g
-0.003803701 -0.003806235
0.0007300695 -0.002155653
_COL24 百粒质量/g
-0.000482539 0.0006457787
-0.001877701 0.0033874612
_COL25 果梗比(%)
-0.110381389 0.4862888168
-0.05361565 -0.514294117
_COL26 出汁率(%)
-0.014473306 -0.000773022
0.0299549282 0.0478888581
_COL27 果皮质量 (g)
3.5419965111 6.1494361329
-2.479875502 0.0506881944
L_ L*
0.2275474901 0.2590003056
0.3431969493 0.034153551
_COL29 a*(+红; -绿)
-0.027155534 0.0324278158
-0.137761495 0.0121034463
_COL30 b* (+黄; 蓝)
-0.269817817 -0.290750218
-0.348593628 0.0769164999

Raw Canonical Coefficients for the WITH
Variables

W1 W2 W3
W4 W5
_COL31 花色苷(mg/L)
0.0017549354 -0.009123219
-0.005088945 0.0012125284
0.0134027279
_COL32 单宁(mmol/L)
0.1460435408 0.2106151022
-0.149197431 0.5654849423
-0.574577783
_COL33 总酚(mmol/L)
-0.006846347 0.1033113659
0.0755118998 -1.304030331
0.7189990192
_COL34 酒总黄酮(mmol/L)
-0.004683938 -0.05503808
0.2622669098 0.1301276586
-0.258560653
_COL35 白藜芦醇(mg/L)
-0.009991485 0.0426643011
0.0859272135 0.0889189214
0.1277809546
_COL36 DPPH半抑制体积 (IV50)
1/IV50(uL) -4.124557919
4.2156905978 0.7395143935
-1.308537606 5.9586875668
L_D65_ L*(D65)
-0.020467281 -0.018046699
-0.064735343 -0.050745365
0.1481041122
a_D65_ a*(D65)
0.0042293785 -0.005797949
-0.065343123 -0.071288023
0.025090639
b_D65_ b*(D65)
-0.013867944 0.0008243472
-0.029729763 0.1499803385
0.1867651286

-0.2154 0.0111	0.0744	-0.3040	-0.4095 0.0450	0.2718	0.1192	积 (IV50)	1/IV50(uL)	-0.0003 0.1866
0.2974 0.0336	_COL6 0.0399	苹果酸 (g/L) -0.1644	-0.2648 -0.0338	_COL26 -0.0645	出汁率(%) 0.2757	0.1149 0.0136 -0.0070	L__D65_ 0.1506 0.1230	L*(D65) -0.0849
0.0379 -0.1834	_COL7 0.1416	柠檬酸 (g/L) -0.2834	0.1283 0.0994	_COL27 0.0684	果皮质量 (g) 0.1466	0.0878 -0.0877	a__D65_ -0.1928	a*(D65) 0.1763
力E (A/min g ml) -0.0360	_COL8 0.1046	多酚氧化酶活 -0.0755 -0.1766	-0.1106 0.0971	L_ 0.1276	L* 0.0433	0.1273 -0.0360	b__D65_ -0.2492	b*(D65) 0.0291
ΔA/g*g*min*ml 0.1802 0.1231	_COL9 -0.0021	褐变度 0.1577	0.1878 -0.1217	_COL29 -0.0046	a*(+红; -绿) -0.3243	1 8		
	_COL10	DPPH自由基 -0.1888 0.2167	-0.1560 0.1241	_COL30 0.1171	b* (+黄;-蓝) -0.0947	The CANCERR Procedure		
1/IC50 (g/L) 0.1141	-0.1343	0.2167				Canonical Structure		
0.1122 0.0586	_COL11 -0.0115	总酚(mmol/kg) 0.1120	Between the WITH Variables and Their Canonical Variables			Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables		
0.0351 0.1015	_COL12 0.0925	单宁(mmol/kg) 0.0651	W1 W4	W2 W5	W3	W1 W4	W2 W5	W3
	_COL13	葡萄糖黄酮 0.1036 0.0032	_COL31 0.9191 -0.0209	花色苷(mg/L) -0.2119 0.1368	0.2687	_COL1 0.2005 0.3164	氨基酸总量mg/100gfw 0.3837 0.0640	0.1904
(mmol/kg) -0.0506	0.1031	0.0032	_COL32 0.8872 0.0123	单宁(mmol/L) 0.3147 0.0601	0.2361	_COL2 0.4858 -0.0084	蛋白质mg/100g 0.2180 -0.3082	0.2454
(mg/kg) 0.0301 1 7	-0.3256	0.0541	_COL33 0.8761 -0.1096	总酚(mmol/L) 0.2997 0.1712	0.3045	_COL3 0.0640 -0.3074	VC含量 (mg/L) 0.0247 -0.3068	-0.1330
The CANCERR Procedure			_COL34 0.6778 -0.0967	酒总黄酮(mmol/L) 0.2339 -0.0201	0.5799	_COL4 0.8982 -0.0127	花色苷mg/100g鲜重 -0.1277 0.0650	0.3423
Canonical Structure			_COL35 0.5419 -0.2040	白藜芦醇(mg/L) 0.5095 0.3084	0.1979	_COL5 -0.1156 0.2916	酒石酸 (g/L) 0.1819 0.0962	0.1747
Correlations Between the VAR Variables and Their Canonical Variables			_COL36 0.4324 -0.9588 -0.0126	DPPH半抑制体积 (IV50) 0.7695 -0.0638 L*(D65) -0.2250 -0.0665	0.3485	_COL6 0.4223 0.0266	苹果酸 (g/L) -0.4925 0.2030	0.2245
V6 V9	V7	V8	L__D65_ -0.9588 -0.0126	L*(D65) -0.2250 -0.0665	-0.0544	_COL7 -0.0709 0.1766	柠檬酸 (g/L) -0.2212 -0.0232	0.0742
-0.0423 0.0902	_COL15 0.2550	黄酮醇(mg/kg) 0.0175	a__D65_ 0.7721 -0.1041	a*(D65) 0.4303 0.0023	-0.3524	_COL8 0.3928 0.0824	多酚氧化酶活力E (A/min g ml) 0.0307 -0.0027	0.2656
0.0904 0.2205	_COL16 0.3641	总糖g/L -0.0837	b__D65_ 0.6689 0.2464	b*(D65) 0.5684 0.2305	-0.1875	_COL9 0.4599 0.0943	褐变度ΔA/g*g*min*ml -0.3707 0.0648	0.2407
0.1822 -0.0260	_COL17 0.0711	还原糖g/L -0.0261	Between the WITH Variables and Their Canonical Variables			_COL10 0.1714 -0.0970	DPPH自由基/IC50 (g/L) 0.3524 -0.1675	
g/l 0.4230	-0.0804	0.2906	W6 W9	W7	W8	_COL11 0.7045 -0.0762	总酚(mmol/kg) 0.2951 -0.0441	0.5244
-0.2724 -0.0788	0.1671	-0.0398	_COL31 -0.0227 0.0896	花色苷(mg/L) 0.0879 0.0525		_COL12 0.8400 -0.1818	单宁(mmol/kg) 0.1964 -0.0032	0.2119
-0.0140 0.1446	0.1910	0.1419	_COL32 0.0543 0.1098	单宁(mmol/L) 0.1978 -0.0146		_COL13 0.5733 -0.2480	葡萄糖总黄酮 (mmol/kg) 0.3808 0.0014	0.5402
0.1201 -0.1280	-0.0207	-0.1749	_COL33 0.0502 0.0643	总酚(mmol/L) 0.0405 -0.0167		_COL14 0.2913 -0.0653	白藜芦醇(mg/kg) 0.2337 0.1693	0.0604
g/100g 0.3196	-0.1033	0.0387	(mmol/L) 0.3453 0.0249	酒总黄酮 0.0235 0.1397		_COL15 0.6126 0.3721	黄酮醇(mg/kg) 0.1739 -0.1027	0.1404
-0.1666 -0.2965	-0.2962	0.0974	_COL34 -0.1929 -0.3730	白藜芦醇(mg/L) 0.0148 0.3072		_COL16 0.1910 0.2829	总糖g/L 0.2858 0.2719	0.0032
-0.0862 0.0824	0.0122	-0.0542	_COL25	果梗比(%)				

_COL17	还原糖g/L		ΔA/g* ^g *min*ml			0.1588	-0.0035	-0.1983
-0.0153	0.1654	0.0340	0.1524	-0.0016	0.0964	-0.0598		
0.5473	0.3128		0.0604				_COL30	b* (+黄;-蓝)
_COL18	可溶性固形物g/l			_COL10	DPPH自由基	-0.1319	0.0888	-0.0579
0.2904	0.1994	-0.0054	1/IC50 (g/L)		-0.1597	0.0609		
0.2765	0.1982		0.0865	-0.0821	0.1064			
_COL19	PH值			_COL11	总酚(mmol/kg)		Correlations Between the	
-0.2648	0.1014	0.2917	0.0949	-0.0087	0.0685	WITH Variables and the Canonical Variables		
0.1702	-0.2928		0.0288			of the VAR Variables		
_COL20	可滴定酸 (g/l)			_COL12	单宁(mmol/kg)			
0.2010	0.2535	-0.2660	0.0297	0.0701	0.0398			
-0.1478	0.1117		0.0498			V1	V2	V3
_COL21	固酸比			_COL13	葡萄总黄酮	V4	V5	
-0.2172	-0.2088	0.2674	(mmol/kg)		0.0876			
0.2156	-0.0207		-0.0384	0.0631	0.0016	_COL31	花色苷(mg/L)	
_COL22	干物质含量g/100g			_COL14	白藜芦醇	0.9147	-0.2053	0.2570
0.2831	0.1639	0.0898	(mg/kg)		0.0242	-0.0194	0.1183	
0.3991	0.3123		0.0228	-0.1991	0.0265	_COL32	单宁(mmol/L)	
_COL23	果穗质量/g		1			0.8830	0.3050	0.2258
0.1191	0.0134	-0.1397	9			0.0114	0.0520	
-0.1532	-0.0721					_COL33	总酚(mmol/L)	
_COL24	百粒质量/g					0.8719	0.2904	0.2912
-0.1229	-0.0137	-0.0959				-0.1020	0.1481	
-0.1816	-0.0321		The CANCORR Procedure			_COL34	酒总黄酮(mmol/L)	
_COL25	果梗比(%)					0.6745	0.2266	0.5547
0.2570	-0.1644	0.0679	Canonical Structure			-0.0900	-0.0174	
0.1192	0.0689					_COL35	白藜芦醇(mg/L)	
_COL26	出汁率(%)					0.5393	0.4937	0.1893
-0.1349	-0.1507	0.4194				-0.1899	0.2669	
-0.0455	-0.2732		Correlations Between the			_COL36	DPPH半抑制体积 (IV50)	
			VAR Variables and the Canonical Variables			1/IV50(uL)	0.7658	0.3377
_COL27	果皮质量 (g)					0.4136	-0.0594	0.1359
-0.0713	-0.0064	-0.0632	W6	W7	W8	L__D65_	L*(D65)	
-0.0945	-0.1039		W9			-0.9542	-0.2180	-0.0520
L_	L*					-0.0117	-0.0575	
-0.8209	-0.3244	0.1971				a__D65_	a*(D65)	
0.0470	-0.0860		-0.0358	0.1934	黄酮醇(mg/kg)	0.7684	0.4170	-0.3371
_COL29	a*(+红; -绿)		0.0443			-0.0969	0.0020	
0.2468	0.1999	-0.1468				b__D65_	b*(D65)	
-0.0005	0.1415		0.0764	0.2761	总糖g/L	0.6657	0.5508	-0.1794
_COL30	b* (+黄;-蓝)		0.1083			0.2294	0.1994	
-0.7893	-0.3290	0.2100						
0.0921	-0.0407		0.1541	0.0539	还原糖g/L			
			-0.0127		-0.0160			
							</	

8.4 问题四附录

1. Matlab 随机出 27 中 20 个不重复的随机数

```
a=randperm(27);  
for i=1:1:20  
    s(i)=a(i);  
end  
s
```

2. 通过 SPSS 进行线性回归分析

- 1、读取 excel 文件。
- 2、依次按分析，回归，线性顺序逐一单击鼠标键，展开线性回归对话框。
- 3、将葡萄酒质量作为因变量，葡萄酒和葡萄的理化指标作为自变量进行线性回归操作。

3. 通过 SPSS 进行显著性检验

- 1、数据导入；
- 2、依次点击分析，非参数检验，旧对话框，2 个相关样本；
- 3、将“V1”、“V2”依次选入右边“检验对”框中；
- 4、选择“Wilcoxon 匹配样本对符号秩”复选框；
- 5、点击进入“选项”对话框，选择检验方法，勾选“描述性”、“四分位数”；
- 6、点击确定进行两配对样本显著性检验。

4. 筛选后的 23 个指标

1	2	3	4	5	6	7	
氨基酸总量	柠檬酸	多酚氧化酶活力	白藜芦醇	黄酮醇	总糖	还原糖	
8	9	10	11	12	13	14	
可溶性固形物	干物质含量	果穗质量	百粒质量	果梗比(%)	a*(+红; -绿)	b* (+黄;-蓝)	
15	16	17	18	19	20	21	22
单宁	总酚	酒总黄酮(白藜芦醇	DPPH 半抑制体积	L*(D65)	a*(D65)	b*(D65)

