

## 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 \_\_\_\_\_

我们的参赛报名号为（如果赛区设置报名号的话）：\_\_\_\_\_ 20121523 \_\_\_\_\_

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

参赛队员（打印并签名）：1. \_\_\_\_\_ 文铖 \_\_\_\_\_

2. \_\_\_\_\_ 赵凯 \_\_\_\_\_

3. \_\_\_\_\_ 徐欢 \_\_\_\_\_

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

日期：\_\_\_\_\_ 2012 年 9 月 9 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）：

# 葡萄酒相关指标分析

## 摘要

本题讨论葡萄酒的评价问题，可将红白两色葡萄、葡萄酒分开处理，建立不同的模型，从主观评判标准和客观评判标准分别入手。

对于第一问，建立两配对独立样本 t 检验模型和方差求和比较模型。检验得出标准化的两组数据服从正态分布，将两组数据全部相减，检验它的期望均值不为 0，得出两组人员对酒的评价有显著性差异。将两组标准化处理后的评分进行方差求和，比较大小，值越小可信度越高，判断第二组结果更具有可信性。

对于第二问，建立主成分分析模型，得到酿酒葡萄的理化指标的综合得分，再结合评酒员对相应葡萄酒质量的评价得分，依据实际经验对两种得分加权处理综合打分。最终将红白两色葡萄分别分为 4 类：①红葡萄 2,3,9 为特等品，红葡萄 1,14,16,17,19,21,22,23 为一等品，红葡萄 4,5,6,7,8,11,12,13,18,20,24,26,27 为二等品，红葡萄 10,15,25 为三等品。②白葡萄 1,17,21 为特等品，白葡萄 2,4,8,13,14,15,19,22 为一等品，白葡萄 5,6,9,10,11,12,16,18,20,23,25,26 为二等品，白葡萄 3,7,24,27,28 为三等品。

对于第三问，建立逐步回归分析模型。利用 SPSS 软件，得出 10 组成线性相关关系的指标间的直线回归方程，对于成曲线相关关系的指标，通过变量转换，转化为线性相关关系，从而得到 2 组曲线回归方程。还剩 3 个葡萄酒的指标未能找到相关的酿酒葡萄的指标。

对于第四问，同样建立主成分分析模型。对于理化指标标准化后的数据，当不考虑芳香物质时，观察相关系数矩阵，结合第三问的回归方程，找出相关性大的理化指标，判断单宁和总酸对葡萄酒质量影响很大。考虑芳香物质后，两者对比，得出含芳香物质时理化指标与葡萄酒的质量相关性好。说明除单宁、总酸外，芳香物质对葡萄酒的质量评价非常重要。故可认为可以利用葡萄和葡萄酒的理化指标来评价葡萄酒的质量，但应考虑到芳香物质这一理化指标。

## 关键字：

相关性 两配对独立样本 t 检验 主成分分析法 逐步回归分析

## 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。附录 1 给出了某一年份一些葡萄酒的评价结果，附录 2 和附录 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. 分析附录 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

本题主要涉及酿酒葡萄和葡萄酒之间各个指标的的一系列关系，我们主要通过通过对各种数据应用多元统计分析里的一系列方法得出之间的关联。实际问题是葡萄酒存在了很长时间，业内已经有一系列判别标准，虽然有些是经验方法，但是在某些主要因素上这些标准时准确的。这也给了我们检验模型是否正确的方法：通过模型答案和实际结果的相符度判断模型的可靠性。

如在解答之前我们在网上找到葡萄酒的总酚，单宁和总酸等理化参数与葡萄酒等级划分有极显著相关，其中总酚和总酸可判别葡萄酒的等级。这对于第四问的某些主成分选择有意义。

由于红葡萄和白葡萄所给的数据都一一对应，在涉及酒和葡萄的打分里我们都采用分开考虑的方法，主要可以对比形成相互的检验。

第一问对数据标准化消除量纲差异，得到两组正态分布的数据。两者差值越小表示相关度越高，对于差值小的标准我们就采用独立样本 t 检验的方法。而对于结果的可信性的判别标准，我们采用比较每组数据方差大小的方法，方差和越小表示十个人的判断越一致，十个人一致度越高表示那组数据更可信。

第二问需要葡萄的理化指标和酒的质量对葡萄分类，显然这两个主要标准既有独立的影响又有相关的影响。对于这两个指标首先分开考虑影响对葡萄打分，接着对两组打分加权得到总分来对葡萄的质量分组。对于葡萄的理化指标由于每种葡萄有几十组关系，我们采用主成分分析的方法，求出标准化矩阵的特征根和特征向量再通过矩阵的一系列变换以及两次线性变换的综合评价得出葡萄的排序。而酒的质量就采用第一问可信度高的一组的人员的最后打分的平均值作为打分依据。然后对两组打分加权。

第三问考虑葡萄酒与酿酒葡萄之间的联系，通过联系实际“葡萄酒由酿酒葡萄制成”，再观察发现题目给出的所有酿酒葡萄的指标集合包含了葡萄酒的指标集合，可判断出酿酒葡萄的指标为自变量，葡萄酒的指标为因变量。通过 MATLAB 观察散点图，猜想自变量与因变量之间可能在线性相关关系。用 SPSS 软件对葡萄酒与酿酒葡萄的指标进行逐步回归分析，得到回归方程。若自变量与因变量之间是曲线相关关系，则通过变量转换转化为线性相关关系。同时注意检验相关系数，判断拟合的合理性。

第四问中，主要是借某种测度来度量某个解靠近“正理想解”和“负理想解”的程度，而要刻画理化指标与葡萄酒质量的联系程度，最好的方法就是在定义某种测度的基础上，计算理化指标与葡萄酒质量的贴近度，一般贴近度越大，则表明理化指标对葡萄酒的质量影响越大。

### 三、模型假设

- (1) 假设所有数据准确，误差小；
- (2) 假设评酒员对各指标的打分近似服从正态分布；
- (3) 假设评酒员的打分中有明显错误和缺少的数据用平均值代替；
- (4) 假设不考虑评酒员在评酒时心态变化等偶然因素的影响；
- (5) 假设依据经验采取的赋值具有较高的科学性；
- (6) 假设可将每组对应的数据相比较来近似代替标准化；
- (7) 假设包含于一级指标里的二级指标对解题影响不大；

### 四、符号系统

- (1)  $X_1, X_2$  表示评酒员对酒评价的所有数据；
- (2)  $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$  表示两组数据的正态分布；
- (3)  $\mu$  表示总体平均数；
- (4)  $Y$  表示两组评酒员对应评价数据的差值；
- (5)  $y_i$  表示  $Y$  数组里对应的所有数据；
- (6)  $D(X_1), D(X_2)$  表示评酒员对葡萄酒所以打分的方差；
- (7)  $a, e$  表示二十七种不同的红葡萄  $a=1, 2, \dots, 27$ ；
- (8)  $b$  表示人员对葡萄酒的十项打分小项  $b=1, 2, \dots, 10$ ；
- (9)  $d$  表示芳香物质的类别总和  $d=1, 2, \dots, 73$ ；
- (10)  $c_{ad}$  表示一种红葡萄某一芳香物质的含量；
- (11)  $Mef$  某一葡萄某一指标的具体数值；
- (12)  $f$  表示葡萄中二十九种不同指标；
- (13) 右上角有\*表示数据经过初始标准化得到的对应数据；
- (14)  $\hat{y}_{\text{红}i}$  表示红葡萄酒的第  $i$  个指标的估计值；
- (15)  $\hat{y}_{\text{白}i}$  表示白葡萄酒的第  $i$  个指标的估计值；
- (16)  $x_i$  表示（红或白）葡萄的第  $i$  个指标；
- (17)  $\beta_i$  表示  $x_i$  的系数。

## 五、模型建立

### 5.1 模型一（两配对独立样本 t 检验模型）

对于第一问，要判断出两组评酒员的评价结果是否有显著性差异。首先的区分是两组人员对所有相同酒的评价，其次是红白两种葡萄酒，最后是同一种酒的十个人的打分。

由于只是找出两组评酒员之间是否有显著性差异，所以要考虑把其它数据按照一个标准只划为一类。所以暂时就不考虑第三个差异，即只看不同组的评酒员对同一种酒的各个指标的显著性差异。由上述分析以及两独立样本t检验是根据样本数据对两个配对样本来自的两配对总体的均值是否有显著差异进行推断。考虑采用两配对样本t检验的方法建立模型。同时考虑到这种方法的限制条件为两组数据的观察值顺序相同且数目相同。从合理性来说，建立的模型是有一定的理论基础的。

对于独立样本t检验，还要考虑到使用的样本必须满足正态分布。然后在正态分布的情况下看两组样本的均值是否具有显著性差异。

首先需要考虑到两种葡萄酒是不同的类别，所以适合采用纵向对比，即同一种颜色的酒不同规格的比较。这需要对两组红葡萄酒进行正态分布的检验。即检测总体 $X_1$ 是否符合正态分布 $N(\mu_1, \sigma_1^2)$ ，总体 $X_2$ 是否符合正态分布 $N(\mu_2, \sigma_2^2)$ ，由于这是初始的数据，我们可以画出两组酒的频数分布直方图来初步判断。但是由于评酒员对不同项目的打分的满分标准是不一样的，所以对于这些不一样的数据，我们需要把它们纳入一个统一的标准里，而数据标准化就可以消除这种差异，转化在0~1之间求解，如此转化后我们画出图形可见数据基本上是满足正态分布的。

由于满足独立样本t检验的基本条件得到了验证且两样本相互配对，所以我们需要检验 $\mu_1, \mu_2$ 是否有显著性差异。

考虑到实际情况是如果两组数据百分百没有显著性差异的话，则他们对应的每一个数据都是相同的，他们之间一一对应之差全为零。因此他们之间的差可以作为新的判断标准。对应的就是我们引进一个新的变量 $Y=X_1-X_2$ 。其对应的样本值为 $(y_1, y_2, \dots, y_n)$ 。 $y_i = x_{1i} - x_{2i} (i=1, 2, \dots, 2700)$ 。这样所求的问题就转化为单样本t值检验问题，即转化为检验Y的均值是否与0有显著差异。

接下来就是假设的t分布检验：

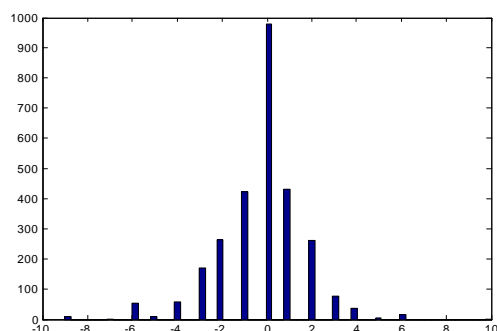
建立零假设 $H_0: \mu_Y = 0$ ；

$$\text{构造t统计量: } t = \frac{\bar{y}}{s_y / \sqrt{n-1}} \sim t(n-1)$$

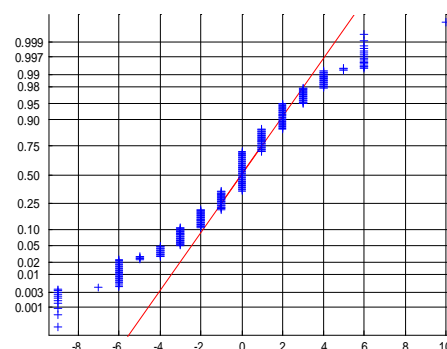
由于之前我们假设了两个对应的数相减就可以消除由于分数在比重之间的差异，这就相当于无量纲化，所以所得到的所有的数据现在就可以看成是一个类型的，就可以对所有的数据进行正态分布的检验。此时的所有的数是指27种不同

红葡萄酒在第一组人的评价下所得到的所有数据与第二组十个人所得的数据之差。由上面的分析，我们可以利用matlab自带的函数进行求解。利用hist函数作频数分布直方图（图一），再利用normplot函数进行分布的正态性检验（图二）。

注：具体程序见附录一



图一



图二

对图一分析我们可以知道图形中间高两边低而且近似对称，数据大体服从正态分布。对于图形二，我们可以看出，中间数据比较符合，两边偏差比较大，但是考虑到我们并没有严格的无量纲化，让数据在  $0 \sim 1$  之间，所以也可以近似认为数据符合正态分布。

在得出数据  $Y$  符合正态分布之后，我们就要对数据的参数进行估计。均值为 0.257，标准差为 1.7656，均值的 95% 的置信区间  $[0.1904, 0.3237]$ ，标准差的 95% 的置信区间  $[1.7197, 1.8140]$ 。

另外我们还要在方差未知的情况下，检验其均值是不是等于 0.257，通过 `ttest` 函数的调用，我们得出：

(1) 布尔变量  $h=1$ ，表示零假设被拒绝， $H_0$  不成立，说明提出的假设两组评酒员的评价结果无显著性差异是不合理的；

(2) 95% 的置信区间为  $[0.1904, 0.3237]$ ；

(3) Sig 的值为  $5.2958e-14$ ，远小于 0.5，不能接受假设“两组评酒员的评价结果无显著性差异”。

对于白葡萄酒来说，由于只是类别不同，这两个因素是独立不相关的，而其它的相对应的条件都是相同的，所以我们思考的步骤仍然同红葡萄酒的分析方法。建立  $Y'$  处理可知两组数据相减得到的结果仍然服从正态分布。对  $Y'$  的数据参数进行估计，得出均值为 0.257，标准差为，均值的 95% 的置信区间  $[0.1851, 0.329]$ ，标准差的 95% 的置信区间  $[1.8577, 1.9595]$ 。

同样还要在方差未知的情况下，调用 `ttest` 函数，检验其均值是不是等于 0.257，我们得出：

(1) 布尔变量  $h=1$ ，表示零假设被拒绝， $H_0$  不成立，说明提出的假设两组评酒员的评价结果无显著性差异是不合理的。

(2) 95% 的置信区间为  $[-0.3242, -0.1808]$

(3) Sig 的值为  $6.2534e-14$ ，远小于 0.5，不能接受假设“两组评酒员的评价结果无显著性差异”。

由此我们得出，两组评酒员的评价结果有显著性差异。

## 5.2 模型二（方差求和比较模型）

这主要考虑两组数据与标准的差别，差别越小表示可信性越大。考虑到稳定性用平均数、方差、标准差来检验，采用此类方法时对于平均数来说，每一行的数据都是基于不同标准的，平均数就没有实际的意义了。而方差和标准差只是同一标准的不同表现形式，所以出于简便性的理解，我们用所有数据的方差来作为主要的评判标准。

另外，类似于第一小问，考虑到两种葡萄酒是一种主要的差异，仍然把葡萄酒分开考虑方差，这样最后就会得到一一对应的两组数据，通过对应的方差  $D(X_1)$ ,  $D(X_2)$  的比较，可以得到两组人员两组的方差的可信性，如果两组都显示的是某一组人员可信性高，那么结果更加具有可信性。

对于方差的设定，需要找出一个统一的标准，直接对最后的数据进行比较。考虑到一组人员有十个人，十人对同一种酒的同一个指标都有评价，如果这一组人的可信性高就表示十个人对同一个指标评价都相同。即每一个对应的指标的十个数据求方差，方差越小表示越可靠。同一种葡萄（红或白）对应的一种酒的一个指标都有十个人的两个指标，所以也就是每一组的每一个值都有一个性质与之一样的对应值，所以只需要直接把所得到的所有的方差相加，即：

$$D(X_1) = \sum_{a=1}^{27} \sum_{b=1}^{10} d_{1ab} \quad (a \text{ 表示 } 27 \text{ 种葡萄酒, } b \text{ 表示每一个酒的十个理化指标})$$

$$D(X_2) = \sum_{a=1}^{27} \sum_{b=1}^{10} d_{2ab}$$

最后得到的值虽然是不同指标下的方差，但是每一个方差都是有对应的性质一样的方差。虽然没有对应的标准差去量化精确，但是所求的值还是有一定可信性的。

由方差的性质可知，对于所求得最终值越小越好。

具体的做法示例如下：

红葡萄酒第二十五种外观澄色度

	1	2	3	4	5	6	7	8	9	10
第一组	4	4	4	5	4	4	4	4	3	4
第二组	3	3	4	4	3	4	5	4	4	3

表一

通过对例子中的数据求方差，得出第一组为 2，第二组方差为 4.7，通过对 27 种葡萄每一个的十项指标总共 270 个都类似的求和。（具体数据见附录二。）

最后得到对于红葡萄酒来说，第一组人员得到的方差为 456.6333，第二组人员得到的方差为 325.7444。由此可以看出第二组结果更可信。

同理对于白葡萄酒来说，所有对应的数据的性质也都是相同的，我们可以用



完全一样的方法求出答案。

最后得到对于白葡萄酒来说，第一组人员得到的方差为 690.0667，第二组人员得到的方差为 389.5222。由此可以看出第二组结果更可信。

综上所述对于两组人员来说，第二组数据更加具有可信性。

### 5.3 模型三（主成分分析模型）

根据题意要由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。即根据葡萄的各种指标和酒的质量之间的关系与取舍求出葡萄的等级。

题中涉及了的葡萄理化指标和葡萄酒的等级初步看来是两个联系不大的条件，所以首先先分开考虑它们对实验葡萄的等级的影响，然后再考虑采用一定的方式加权，求出两个指标对于葡萄等级的总的影响，得出答案。

对于葡萄的理化指标来说，题目给出了两个表格的许多组数据，很显然不同类别的数据对葡萄等级好坏的贡献比例是不相同的，不可以把它们平均看待，要知道主要因素的影响程度。很显然，这是多元统计分析处理的多元变量问题，可以采用主成分分析的方法。

同样，由于是不同的葡萄酒，我们还是会两种葡萄酒分开考虑。

对于红葡萄来说，首先我们先对其中的理化指标数据进行初步的分类，把显而易见的相关性的数据只用关联性比较大的数据代替。对于葡萄的第一种理化指标（如氨基酸，蛋白质等），其中有影响比较大的一级因素和其次的二级因素。在氨基酸一类里，由于既有总的氨基酸又有各个细类的氨基酸，我们考虑只用总量氨基酸这一指标表示所有的氨基酸。类似的用白藜芦醇这一个指标代替第二级的四个醇和苷，用还原性糖代替果糖和葡萄糖，用黄酮醇的含量表示杨梅黄酮，槲皮素，山奈酚，异鼠李素。另外对于颜色指标来说，L，A，B 有相关的计算公式：

$$\Delta = \sqrt{L^2 + A^2 + B^2}$$

而颜色的二级指标 H，C 就不考虑。所以对于颜色的五个指标只考虑 L，A，B 三个。所以对于表格二的所有指标最后只取出关联性小的 28 组考虑。

对于表三里的芳香物质可以看出不同的葡萄同一种物质含量差别比较大，有的葡萄还缺少某些芳香物质，所以每一类方向物质都比较的话，对于类别差别比较大的葡萄显然没有可比性，所以如同表二，我们对一种葡萄的所有芳香物质求和，即

$$\sum_{d=1}^{73} c_{ad}, \quad d \text{ 表示芳香物质的类别总和, } c_{ad} \text{ 表示一种红葡萄的某一芳香物质的}$$

含量

用总和代表每一个芳香物质。所以表三的数据最后简化为一组葡萄的理化指标，即总芳香物质的含量，这样就可以与第二组简化数据结合。总共用 29 组数据代表红葡萄的理化指标。

建立横坐标表示各项指标，纵坐标表示不同的葡萄的指标，建立一个 27\*29 的矩阵。

表示为  $M_{ef}$ ,  $e \sim (0,27); f \sim (0,29)$

具体操作过程如下

(1) 我们要将原始数据标准化处理。

$$M_{ef}^* = \frac{M_{ef} - E(M_{ef})}{\sqrt{D(M_{ef})}}$$

(2) 建立变量的相关矩阵， $M_{ef}^*$  的协方差矩阵就是  $M_{ef}$  的相关系数矩阵 R。

(3) 求出 R 的特征根为  $\lambda_1^* \geq \lambda_2^* \geq \dots \lambda_{29}^* \geq 0$ , 相应的特征向量为  $T_{e1}^*, T_{e2}^*, \dots, T_{e29}^*$ ;

(4) 由累计方差的贡献率确定主成分的个数 g 以及所有影响因素的 h, 主成分的表达式:  $P_{ef} = (T_{ef}^*)' M_{ef}^*$ .

在得到所有数据之后, 需要对主成分进行加权综合, 将所有成分的权数根据它们的方差贡献率来确定, 已知  $P_{e1}, P_{e2}, \dots, P_{e29}$  是所求得所有成分, 它们的特征根分别是  $\lambda_1^*, \lambda_2^*, \dots, \lambda_{29}^*$ , 由于特征根的数值差别比较大, 单独利用数值没有可比性, 我们将其归一化, 即:

$\omega_f = \frac{\lambda_f^*}{\sum_{f=1}^{29} \lambda_f^*}$ 。记作  $W = (\omega_1, \omega_2, \dots, \omega_{29})'$ , 由  $P_{ef} = (T_{ef}^*)' M_{ef}^*$  构造综合评价函数为:

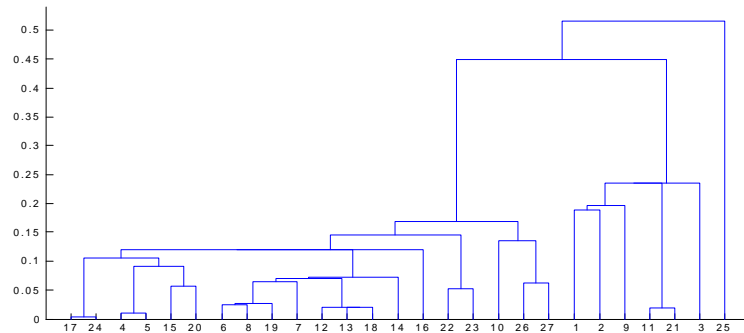
$$Q_e = \omega_1 * P_{e1} + \omega_2 * P_{e2} + \dots + \omega_{29} * P_{e29} = W' P_{ef} = W' (T_{ef}^*)' M_{ef}^* = (T_{ef}^* W)' M_{ef}^*;$$

在式中, 令  $T_{ef}^* W = W_{k*1}^*$ , 带入上式得到  $Q_e = (W_{k*1}^*)' M_{ef}^*$ 。 [1]

经过上面的数据处理所得到的综合函数就是对就是对原始指标的线性综合, 从计算主成分到对之加权, 经过两次线性运算后得到的综合评价函数。

在题目的求解中, 利用 matlab 软件进行具体的操作求解。(具体数据见附录三。)

在所得到的 27\*29 的矩阵, 把其中的数据都按行 (即葡萄的理化指标) 标准化, 然后再调用 matlab 里的主成分分析的函数, 对标准化的矩阵计算指标数据的主成分, 图形如图三所示, 数据如下表 (程序见附录四)。



图三

主成分	特征值	贡献率%	累计贡献率%
1	7.0209	24.2099	24.2099
2	4.7628	16.4235	40.6334
3	3.1530	10.8724	51.5058
4	2.8967	9.9888	61.4946
5	1.7909	6.1754	67.6699
6	1.5461	5.3315	73.0015
7	1.2903	4.4493	77.4508
8	1.1520	3.9723	81.4231
9	0.9087	3.1333	84.5564

表二

由于篇幅的限制，我们只列出了影响因素大于 3% 的九个主因素的影响，其他详见附录。

通过特征向量与所得的标准化数据的矩阵相乘就得到另外一个  $P_{ef}$  的矩阵。然后以 29 个影响着的因素加权综合得出最后 27 个葡萄的加权得分。得出的加权得分和排名见表三：

通过表三我们知道 27 种葡萄的最终排名，由题意要得到对葡萄的最终排名还需要知道葡萄酒的质量，我们需要找一个指标对葡萄酒进行打分排序。通过第一问的分析可知第二组人员对葡萄酒的打分比较合理，我们就采用第二组的打分作为对酒的评价标准，把十个人的每一项打分求和再求平均得出的 27 组数值作为对酒的评价。得出的数据如表四，分析知标号的葡萄酒和标号的葡萄是一一对应的。

这样我们对每种葡萄就有两种评价的值，由这两组的值通过加权得到对葡萄的最后评价。但是对于这两组的数据来说，取值的范围相差很大，很显然这两组的数值即使是加权之后我们也不可以直接相加，就要考虑对数据归一化把数据划到 0~1，然后再利用这两种数据。

对于权值的选择，我们考虑到实际中我们都是依据经验选择，在本题中我们考虑到葡萄的各种理化指标显然是对葡萄的重要评价的，但是有的有意的指标含量越高并不是对于酿酒越好。综合考虑，对于葡萄的最终评价，葡萄的理化指标占 0.7 的权重，葡萄酒的理化指标占 0.3。另外，对于葡萄的最终分类来说我们分为四类，得分 1~0.8 为特级，0.6~0.8 为一级，0.3~0.6 为二级，0~0.3 为三级，具体得分见下表五。

通过的数据处理我们得出了葡萄酒的最后排名，结果显示如表五，通过表五知，红葡萄 2, 3, 9 为特等品；红葡萄 1, 14, 16, 17, 19, 21, 22, 23 为一等品；红葡萄 4, 5, 6, 7, 8, 11, 12, 13, 18, 20, 24, 26, 27 为二等品，红葡萄 10, 15, 25 为三等品。

对于白葡萄来说，每一个对应的性质都和红葡萄一样，我们完全可以采用一样的方法对白葡萄进行评价，由于考虑到白葡萄的得分区间，我们划分为  $0 \sim 0.6$ ,  $0.6 \sim 0.7$ ,  $0.7 \sim 0.8$ ,  $0.8 \sim 1$  我们得出白葡萄的最后综合得分和等级如表六：

通过表六知，白葡萄 1, 17, 21 为特等品；白葡萄 2, 4, 8, 13, 14, 15, 19, 22 为一等品；白葡萄 5, 6, 9, 10, 11, 12, 16, 18, 20, 23, 25, 26 为二等品，白葡萄 3, 7, 24, 27, 28 为三等品。（排序程序见附录五）

红葡样品	综合得分	最后排名
红葡1	1.016594	4
红葡2	1.33146	2
红葡3	1.918301	1
红葡4	-0.67869	23
红葡5	-0.24902	17
红葡6	0.01817	14
红葡7	-0.14559	16
红葡8	0.094139	10
红葡9	1.082295	3
红葡10	-1.11636	26
红葡11	0.271645	7
红葡12	0.07163	12
红葡13	-0.05182	15
红葡14	0.074754	11
红葡15	-0.61212	22
红葡16	0.140184	9
红葡17	-0.29197	18
红葡18	-0.29718	19
红葡19	0.02727	13
红葡20	-0.58314	21
红葡21	0.971442	5
红葡22	0.455231	6
红葡23	0.258645	8
红葡24	-0.306	20
红葡25	-1.76127	27
红葡26	-0.92626	25
红葡27	-0.71234	24

表三

红葡萄酒样品	综合得分	最后排名
红葡萄酒1	68.1	21
红葡萄酒2	74	6
红葡萄酒3	74.6	4
红葡萄酒4	71.2	15
红葡萄酒5	72.1	10
红葡萄酒6	66.3	22
红葡萄酒7	65.3	26
红葡萄酒8	66	23
红葡萄酒9	78.2	1
红葡萄酒10	68.8	17
红葡萄酒11	61.6	27
红葡萄酒12	68.3	19
红葡萄酒13	68.8	17
红葡萄酒14	72.6	7
红葡萄酒15	65.7	24
红葡萄酒16	69.9	16
红葡萄酒17	74.5	5
红葡萄酒18	65.4	25
红葡萄酒19	72.6	7
红葡萄酒20	75.8	3
红葡萄酒21	72.2	9
红葡萄酒22	71.6	12
红葡萄酒23	77.1	2
红葡萄酒24	71.5	13
红葡萄酒25	68.2	20
红葡萄酒26	72	11
红葡萄酒27	71.5	13

表四

	得分	等级
葡萄1	0.645929	2
葡萄2	0.812456	1
葡萄3	0.93494	1
葡萄4	0.379443	3
葡萄5	0.477448	3
葡萄6	0.423459	3
葡萄7	0.374233	3
葡萄8	0.43249	3
葡萄9	0.840958	1
葡萄10	0.252807	4
葡萄11	0.386741	3
葡萄12	0.469774	3
葡萄13	0.455325	3
葡萄14	0.54808	2
葡萄15	0.292711	4
葡萄16	0.511732	2
葡萄17	0.512651	2
葡萄18	0.347203	3
葡萄19	0.539046	2
葡萄20	0.480754	3
葡萄21	0.711436	2
葡萄22	0.602389	2
葡萄23	0.664388	2
葡萄24	0.455766	3
葡萄25	0.119277	4
葡萄26	0.346803	3
葡萄27	0.378463	3

表五

	得分	分类
白葡萄1	0.923944	1
白葡萄2	0.739824	2
白葡萄3	0.507773	4
白葡萄4	0.726423	2
白葡萄5	0.69757	3
白葡萄6	0.645657	3
白葡萄7	0.54106	4
白葡萄8	0.75281	2
白葡萄9	0.649147	3
白葡萄10	0.687463	3
白葡萄11	0.618182	3
白葡萄12	0.623666	3
白葡萄13	0.752651	2
白葡萄14	0.786644	2
白葡萄15	0.767582	2
白葡萄16	0.652768	3
白葡萄17	0.943134	1
白葡萄18	0.65855	3
白葡萄19	0.751687	2
白葡萄20	0.657112	3
白葡萄21	0.815039	1
白葡萄22	0.752844	2
白葡萄23	0.693669	3
白葡萄24	0.402132	4
白葡萄25	0.651147	3
白葡萄26	0.55856	3
白葡萄27	0.20493	4
白葡萄28	0.593809	4

表六

#### 5.4 模型四（逐步回归分析模型）

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，考虑到酿酒葡萄与葡萄酒之间存在因果关系，即葡萄酒由酿酒葡萄制成，酿酒葡萄的指标可能会影响到葡萄酒的各项指标，又可通过观察发现题目给出的葡萄酒的各项指标，都能在酿酒葡萄的指标中找到对应的，即酿酒葡萄的指标集合包含了葡萄酒的指标集合，故可将酿酒葡萄的各项指标看为自变量，葡萄酒的各项指标看为因变量，两者之间可能存在某种函数关系。下面进行模型的建立，在该题的解决过程中，所用的酿酒葡萄和葡萄酒的指标数据均为前面已处理好的。

分别选取 27 个红葡萄酒样品的花色苷作为因变量，相应的红葡萄的氨基酸总量、蛋白质、VC 含量等作为自变量，通过 MATLAB 作出红葡萄酒的花色苷对应红葡萄的 29 个指标（氨基酸总量、蛋白质、VC 含量……果皮质量、果皮颜色、芳香物质）形成的散点图。一共有 29 张散点图，每张散点图中有 27 个散点（程序见附录六）。观察这些散点图。以下图图 4、图 5 为例，可以明显看出，某些红葡萄的指标与红葡萄酒的花色苷之间很可能存在线性关系（或者该种指标与其他指标一起作为自变量，与红葡萄酒的花色苷一起存在线性关系），而某些红葡萄的指标则与红葡萄酒的花色苷之间不存在线性关系（或者该种指标对红葡萄酒的花色苷贡献太小，被其他影响较大的指标“遮盖”了）。由此，初步假设葡萄酒的各项指标可能与葡萄的各项指标之间存在线性关系，其他可能存在非线性关系。

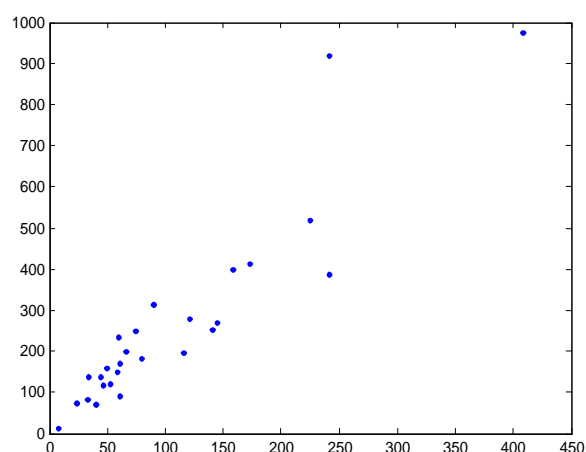


图 4

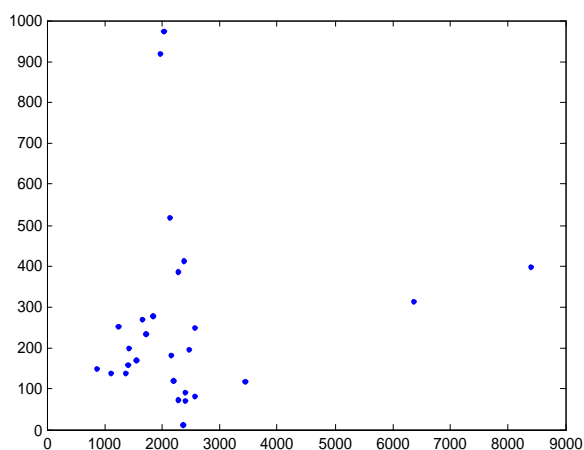


图 5

由于不清楚酒酿葡萄的各项指标间是否存在多重共线性、相互依赖性，为了得到一个可靠的回归模型，需要有效地从众多自变量中挑选出对因变量贡献大的变量，故选取逐步回归分析法，求出因变量（葡萄酒的指标）与自变量（酿酒葡萄的指标）存在的线性关系，并进行检验，对于不符合线性关系要求的因变量与自变量，即非线性回归（曲线回归），通过转变变量，转化为线性关系。

逐步回归方法是从一个自变量开始，是自变量  $x$  作用的显著程度，从大到小地依次逐个引入回归方程。但当引入的自变量由于后面变量的引入而变得不显著是，要将其剔除掉。引入一个自变量或从回归方程中剔除一个自变量，为逐步回归的一步。对于每一步都要进行  $y$  值检验，以确保每次引入新的显著性变量前回归方程中只包含对  $y$  作用显著的变量。这个过程反复进行，直至既无不显著的变量从回归方程中剔除，又无显著变量可引入回归方程时为止。[2]

设回归方程为： $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ ，逐步回归方法的目的可理解为：对于因变量  $y_i$ ，找出对  $y_i$  贡献最大的  $x_i$ ，及其相对应的  $\beta_i$ 。

利用 SPSS 软件依次求出葡萄酒指标与酿酒葡萄指标的线性关系模型，并依据各个模型的相关系数  $R$ 、 $R^2$ ，判断线性回归的拟合程度，依据  $F$  值、显著性水平  $\text{sig}$  判断方程的显著性。若  $R^2 > 0.8$ ，说明方程对样本点的拟合效果很好，

0.5~0.8 之间也可以接受，若  $R^2 < 0.5$ ，则说明方程的拟合效果很差。Sig 值是  $t$  统计量对应的概率值，所以  $t$  和 Sig 两者是等效的，看 Sig 就够了。Sig 要求小于给定的显著性水平，一般是 0.05、0.01 等，Sig 越接近于 0 越好。

对于符合线性关系要求，即拟合效果较好，线性关系较显著的模型，在每个案例中，选取误差最小的一个模型，即最优模型作为该案例的最终模型，依据“非标准化回归系数”栏目中的“B”列系数和回归方程  $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ ，写出每个模型的预报方程，即两类指标之间的线性关系，再根据剩余均方估计写

出预测值  $\hat{y}$  的标准差。

对于非线性关系，将  $y$  转化为  $\ln y$ 、 $\sqrt{y}$ 、 $y^2$ 、 $\lg y$ 、 $e^y$  等基本函数，再以新得到的  $Y_{\text{新}}$  代替原来的因变量  $Y_{\text{原}}$ ，重新计算因变量  $Y_{\text{新}}$  与自变量  $X$  之间的线性关系。若符合线性关系要求，则保留该模型，并由将预报方程中的  $Y_{\text{新}}$  转化为  $Y_{\text{原}}$ ，得到葡萄酒与酿酒葡萄指标间的函数关系。

对于不能求出具体的函数关系的案例，可暂且认为没有明显的具体函数关系，或由于目前知识水平有限、葡萄酒酿制过程本身的复杂性，该种葡萄酒与酿酒葡萄指标之间的联系不明。

目前暂且可以得到的函数关系如下：

葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒花色苷 y1	0.884	很好	91.777	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{红1}} = 437.517 + 2.655x_4 - 6.747x_{26}$				$\sqrt{159090.029}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒单宁 y2	0.939	很好	65.164	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{红2}} = -12524 + 0.06x_{18} + 17.34x_{10} - 0.059x_8 + 0.003x_9 + 0.000x_1$				$\sqrt{13.28}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒总酚 y3	0.806	很好	49.706	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{红3}} = 1.684 + 0.253x_{11} + 0.008x_4$				$\sqrt{32.246}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒总黄酮 y4	0.780	较好	88.543	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{红4}} = -0.950 + 0.398x_{11}$				$\sqrt{51.010}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒白藜芦醇 y5	0.735	较好	13.040	0.000	极显著
预报方程				$S_{\hat{y}}$	



$\hat{y}_{\text{总5}} = 38.003 + 0.343x_{13} - 0.071x_2 - 0.77x_6 + 0.031x_4 + 0.313x_5 + 0.15x_{14}$				$\sqrt{44.335}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒 DPPH 半抑制体积 y6	0.765	较好	81.223	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{总6}} = -0.024 + 0.17x_{11}$				$\sqrt{0.1}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒色泽 y7	0.904	很好	234.293	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{总7}} = 89.538 - 0.17x_4$				$\sqrt{643.473}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
红葡萄酒芳香物质 y8	0.629	较好	9.339	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\ln \hat{y}_{\text{总8}} = 5.1 - 0.013x_{14} - 0.004x_{15} + 0.016x_{12} + 0.004x_{18}$				$\sqrt{0.56}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
白葡萄酒单宁 y1	0.514	可以接受	13.22	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{白1}}^2 = -6.209 + 1.336 * x_{13} + 25.234 * x_{28}$				$\sqrt{191.394}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
白葡萄酒总酚 y2	0.606	可以接受	19.252	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{白2}} = -2.503 + 0.173x_{14} + 0.016x_{19}$				$\sqrt{2.943}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性
白葡萄酒总黄酮 y3	0.803	很好	17.920	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{白3}} = 2.924 + 0.359x_{12} + 0.062x_{22} - 2.948x_{20} + 0.007x_3 + 0.000x_2$				$\sqrt{14.490}$	
葡萄酒指标	$R^2$	拟合效果	F	Sig	显著性

白葡萄酒色泽 y6	0.637	较好	21.946	0.000	极显著
预报方程				$S_{\hat{y}}$	
$\hat{y}_{\text{白6}} = 102.182 - 0.076x_{23} + 0.02x_{27}$				$\sqrt{0.745}$	

其中,  $\hat{y}_{\text{红}i}$  表示红葡萄酒的第  $i$  个指标的估计值,  $\hat{y}_{\text{白}i}$  表示白葡萄酒的第  $i$  个指标的估计值。葡萄酒的指标排序为: (花色苷)、单宁、总酚、酒总黄酮、白藜芦醇、DPPH 半抑制体积、色泽。 $x_i$  表示 (红或白) 葡萄的第  $i$  个指标。葡萄的指标排序为: 氨基酸总量、蛋白质、VC 含量……果皮质量、果皮颜色、芳香物质。

共有 10 组直线回归, 2 组曲线回归, 还有 3 个案例尚未找到合适的函数关系。从上述函数关系可以直观看出葡萄酒指标与酿酒葡萄指标之间的联系。结合红白两色葡萄酒指标的回归方程, 可以认为酿酒葡萄中的花色苷、可溶性固形物、总酚、白藜芦醇、葡萄总黄酮、单宁、蛋白质这几种指标对葡萄酒的指标影响较大。

该模型也可进行回带检验, 进行预估, 评判模型误差, 使用方便简单, 适用范围广。

## 5.5 模型五 (主成分分析模型)

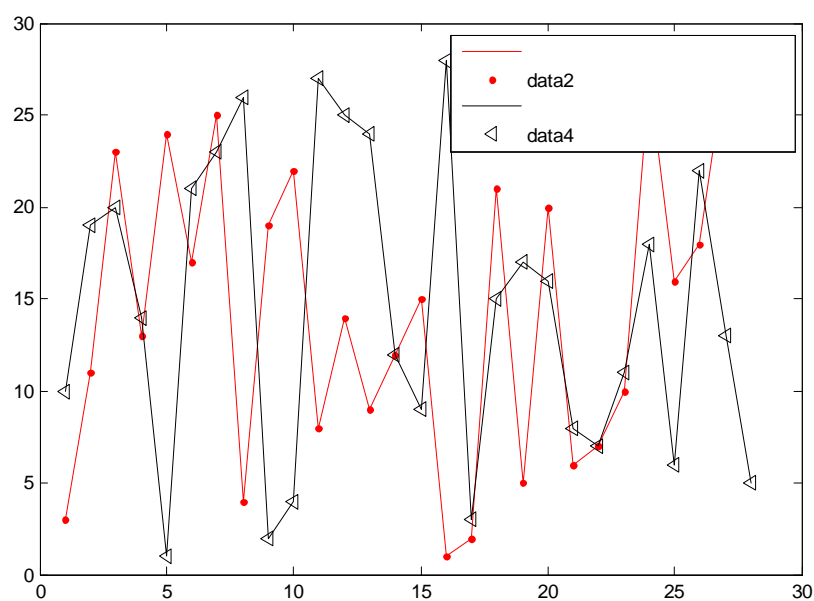
对于第四问, 我们可以采用第二问的主成分分析法。题目要求分析酿酒葡萄和葡萄酒的指标对葡萄酒质量的影响, 我们分析时把酿酒葡萄和葡萄酒的理化指标看成一个整体。用两种处理方法, 第一种不考虑附录 3 中的芳香物质对葡萄酒质量的影响; 第二种把附录 3 中的芳香物质放入理化指标中。

我们把评酒员的评价看成葡萄酒的质量指标。即看用酿酒葡萄和葡萄酒的所有理化指标得出的综合指标得分进行排序, 是否与评酒员的评价排序是否大致相符。如果评价大致符合, 说明可以用葡萄和葡萄酒的理化指标来评价葡萄酒的质量。反之, 这说明不能用来评价。

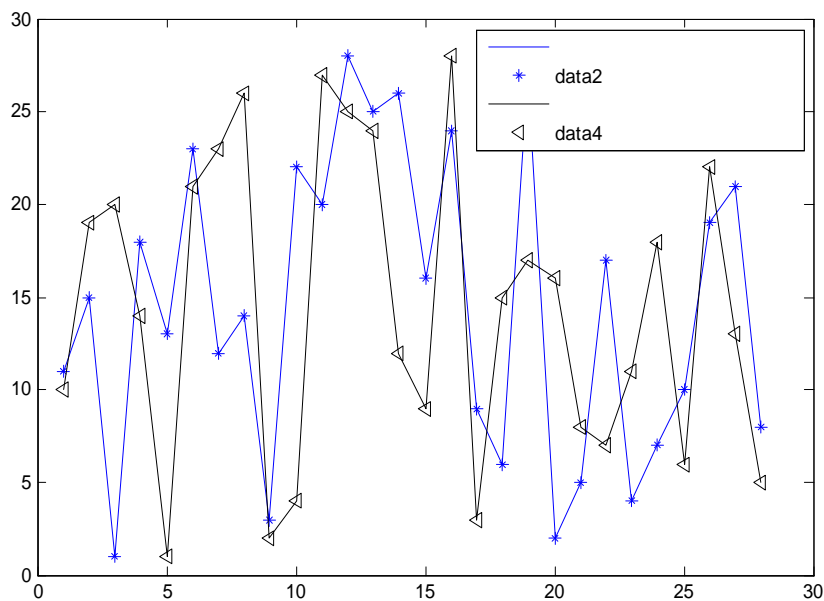
	不含芳香物质得分	排名	含芳香物质得分	排名	评酒员的评价得分	排名
白葡萄1	0.926027	3	0.30304	11	77.9	10
白葡萄2	0.243718	11	0.215246	15	75.8	19
白葡萄3	-0.4476	23	1.118175	1	75.6	20
白葡萄4	0.130463	13	-0.09474	18	76.9	14
白葡萄5	-0.65466	24	0.267507	13	81.5	1
白葡萄6	-0.20029	17	-0.5753	23	75.5	21
白葡萄7	-0.67088	25	0.292528	12	74.2	23
白葡萄8	0.890967	4	0.223833	14	72.3	26
白葡萄9	-0.27958	19	1.035078	3	80.4	2

白葡萄10	-0.34167	22	-0.42822	22	79.8	4
白葡萄11	0.319572	8	-0.33358	20	71.4	27
白葡萄12	-0.0689	14	-3.94217	28	72.4	25
白葡萄13	0.315369	9	-0.69913	25	73.9	24
白葡萄14	0.231655	12	-0.77719	26	77.1	12
白葡萄15	-0.13481	15	0.139279	16	78.4	9
白葡萄16	1.129983	1	-0.59638	24	67.3	28
白葡萄17	1.109086	2	0.592118	9	80.3	3
白葡萄18	-0.33333	21	0.758079	6	76.7	15
白葡萄19	0.630763	5	-1.73827	27	76.4	17
白葡萄20	-0.29434	20	1.089915	2	76.6	16
白葡萄21	0.617433	6	0.899321	5	79.2	8
白葡萄22	0.443652	7	0.122702	17	79.4	7
白葡萄23	0.260425	10	0.996972	4	77.4	11
白葡萄24	-1.25889	27	0.626168	7	76.1	18
白葡萄25	-0.19794	16	0.514206	10	79.5	6
白葡萄26	-0.23212	18	-0.27626	19	74.3	22
白葡萄27	-1.4595	28	-0.3484	21	77	13
白葡萄28	-0.67458	26	0.615471	8	79.6	5

通过 matlab 可以直观的看出相关性



不含芳香物质的排序图



含芳香物质的排序图

通过以上两表对比，我们可以发现在不含芳香物质时理化指标与评酒员的评价相关性不好，在加入芳香物质后，经理化指标打分后的排序与评酒员的相关性较好，说明了芳香物质对葡萄酒的质量评价非常重要。红葡萄酒与白葡萄酒的分析类似，见附录七。

通过以上分析，我们可以发现酿酒葡萄和葡萄酒的理化指标（含芳香物质）可以在一定程度上评价葡萄酒质量的好坏。但在有些葡萄酒的评价中误差过大，像第 3 组和第 24 组就相差过大。所以说按照理化指标可以进行一定的分析，但存在一定的不确定性。

## 六、模型分析

本题虽然只是考虑了葡萄酒，葡萄不同类别之间的关系以及等级之类的问题，可以说是一个大问题之间的细节的关系，但是由于每一问题所要考虑的方面不同，实际我们需要采取数理统计中的不同方法。

对于第一问的独立数据 t 检验，主要需要的是检测两个相减的数据是否符合正态分布，其次是两组数据类型要相同，所以对数据归一化然后检验是否符合正态分布，由于我们没有对数据很好的归一化，只是得到大概而非严格的正态分布，所以理论上不能做 t 检验。但是我们模型的优点就是十分的浅显明了，对于有些数据不太多的问题，直接转化为单样本的 t 值检验，简化问题。

对于主成分分析的方法，我们在论文里为了减少数据处理并不是对所有的每一类影响因素都采用，而是事先根据所查得的资料人为地分类。虽然都是次要的因素，但是仍然会对结果产生一定的影响，这些就是误差的来源。但是我们的主成分分析的方法在实际中仍有广泛地应用。这里我们不只是考虑主成分而是考虑所有影响因素的权重，但是权重不是一次性的采用，利用经过两次线性运算得到

的综合评价函数对原始指标线性综合，在这个层次上可以最大的减小误差。

对于第三问，建立逐步回归分析模型，配合使用 SPSS 软件，可以方便快捷地得到符合要求的模型，解出对因变量贡献大的自变量及其系数，并可通过  $R^2$ 、F、Sig 有效检验拟合的准确性、合理性，同时可以进行回带检验、预估、确定置信度等。非线性相关关系通过变量转换也可以转化为线性相关关系，解出相关函数。该模型使用方便快捷，通用面广，所得数据全面详尽，抓住主要矛盾，忽略了次要矛盾，具有很高的模拟准确性和实用价值。

## 七、模型推广

由于本论文考虑的是一个大数据的几个细节方面，采取了三个主要的方法，虽然很难处理到一个类似的大问题，但是三个主要的方法仍有其各自适用的范围。

配对样本 t 检验主要是面对两组服从正态分布配对样本是否有显著性差异，这在医学上检验不同药物对病人效果的关系上很有利用价值。

主成分分析的方法可以从选定的指标体系中归纳出大部分信息，根据指标间的相对重要性客观加权，可以避免评价者的主观影响，在工业部门各项检验指标里有实际应用，尤其在越来越注重环保的社会。

逐步回归的方法可以从所有的指标里选择重要的指标，通过显著性水平一次的引入和移除找到最主要的因素。如在气象问题里的温度和雨量、雨量和雨日之间的关系中可以采用。

## 八、结论

通过模型的建立和对数据的处理，我们得到了相关的答案，并了解了对于葡萄酒的评估，主观评判标准与客观评判标准相结合、相影响。

另外，通过与实际数据比较以及参考其它的方法，我们思考了所选模型的优缺点，并进行了修改与完善。

在与实际评价指标对比时会有一些细微差距，分析可能是在假设的时候有一些差距，如不是严格的正态分布，忽略二级指标等，但是误差率都不太大。在我们的模型里，每一个问题我们都是采用典型的方法，没有方法的嵌套等，这些就会保证方法的简洁性，所以在处理一些细小的问题时就可以很好的满足。

在第三问，第四问的求解里，由于题目只是叙述了关系，影响等粗略目标，我们其实还可以采用灰色关联度，模糊分析的方法，通过这几种方法的对比，更加详细地找出数据的关系以及最好的方案。

## 九、参考文献

- [1]朱建平，应用多元统计分析，P102-102，北京：科学出版社，2007。
- [2]赵静等，数学建模与数学实验，P267-267，北京：高等教育出版社，2009。
- [3]曾建军，MATLAB 语言与数学建模，安徽：安徽大学出版社，2005。

## 附录

### 附录一：(MATLAB) 两组白的进行配对检验

```
clear,clc
a=[3    4    4    3    3    5    3    4    5    4
8    8    8    2    8    8    8    6    10    8
5    5    5    5    5    5    4    5    6    5
7    7    8    6    6    7    7    7    8    7
14   14   14   12   12   14   12   14   14   14
5    4    5    3    5    6    5    4    6    4
7    7    8    4    6    7    8    7    7    6
7    6    7    5    6    8    7    7    6    6
19   16   19   13   16   22   19   16   22   16
10   9    10   8    9    11   10   10   11   9

3    1    4    3    3    5    3    3    5    3
8    4    8    4    8    10   10   6    10   8
5    2    5    5    5    6    6    3    6    5
7    4    8    4    7    7    8    7    6    7
14   8    14   10   14   14   14   12   12   14
4    2    5    2    4    6    4    3    4    4
6    4    7    4    6    7    7    6    4    7
6    4    6    5    6    7    7    6    5    7
16   10   19   10   16   19   16   13   13   16
9    8    10   7    10   10   10   9    8    10

3    2    4    3    3    5    4    4    5    4
8    8    8    6    8    10   1    8    10   8
5    5    5    5    5    2    0    5    6    4
7    7    8    7    7    6    5    6    7    4
14   14   14   14   12   10   7    14   14   12
5    3    5    4    5    4    4    4    6    5
7    2    8    6    7    6    6    7    7    6
7    5    7    5    6    7    6    5    6    7
19   13   19   16   16   16   22   16   19   19
10   8    11   9    9    9    10   10   10   10

3    3    4    3    4    5    5    4    5    4
6    6    8    6    8    8    8    8    10   8
5    4    5    4    5    5    5    4    6    4
6    7    7    4    6    6    7    7    7    6
14   14   14   10   12   12   14   14   14   12
3    4    4    4    5    5    6    4    6    5
```

6	7	6	6	6	6	7	6	6	7
7	7	6	7	6	7	7	5	6	8
16	16	16	13	16	19	19	16	16	22
9	9	10	8	9	10	10	10	9	10

4	1	2	3	2	1	2	2	1	2
6	4	6	6	6	4	6	6	6	4
6	3	5	4	5	3	6	5	5	5
7	4	7	6	7	6	7	7	6	7
14	10	14	12	14	12	12	14	12	14
4	2	4	2	5	3	4	4	6	4
7	2	7	4	6	6	7	6	8	6
7	4	6	5	6	6	6	5	6	6
19	10	16	10	19	13	16	16	19	16
10	7	10	8	9	8	8	9	10	10

3	1	4	3	3	4	3	3	5	3
8	2	6	4	8	6	8	6	8	8
3	2	5	3	5	2	5	5	3	5
4	4	7	6	7	4	7	6	4	6
10	8	14	10	14	8	14	12	10	14
3	3	5	3	4	2	4	3	3	5
4	4	7	6	6	4	7	6	6	7
5	4	6	6	6	5	6	5	5	7
13	10	19	16	16	13	16	13	13	19
8	7	10	8	9	8	10	8	8	10

2	3	4	3	4	4	4	4	5	3
8	8	6	2	6	6	8	4	8	6
5	5	5	4	4	6	5	5	5	5
7	7	7	6	6	7	7	6	6	6
14	14	14	12	12	14	14	12	12	14
5	5	5	3	5	5	4	3	5	5
7	7	6	6	6	6	7	6	6	7
7	6	7	6	6	6	6	6	5	7
19	16	19	16	16	16	16	13	16	19
10	10	10	8	9	10	9	9	9	10

3	1	4	3	3	3	2	3	5	3
8	4	8	4	6	4	8	6	10	8
3	2	5	4	5	3	4	5	4	5
4	6	7	6	6	4	6	7	6	6
10	8	14	10	14	10	12	14	12	14
5	2	5	2	5	3	4	4	4	5

7	2	6	4	7	6	6	6	6	7
7	4	6	4	6	5	6	7	16	6
19	10	16	10	19	13	16	16	13	19
9	7	10	7	10	8	9	9	9	10

4	2	4	4	5	4	5	4	5	4
8	8	8	8	8	6	8	8	10	8
5	4	5	3	3	3	4	5	6	5
7	7	6	7	6	4	7	6	8	7
14	12	14	10	12	10	12	14	14	14
4	4	5	2	3	2	3	4	3	5
6	4	6	4	6	4	6	7	4	7
6	6	7	5	5	5	7	7	5	6
16	13	16	10	13	10	13	16	13	19
9	9	10	7	9	7	8	10	8	10

2	1	3	2	2	3	3	2	3	1
8	2	8	2	6	4	8	4	8	4
6	2	5	5	6	5	6	5	6	5
7	4	8	7	8	7	8	7	8	7
16	8	14	14	14	14	14	14	16	12
3	2	5	2	6	4	4	5	6	4
6	2	7	4	8	7	7	6	8	6
5	4	7	6	8	6	7	5	7	6
13	10	19	10	19	16	16	16	19	16
9	7	10	8	10	9	10	9	10	10

3	2	3	3	4	4	4	4	5	4
8	4	8	2	6	6	8	4	8	6
5	3	5	4	5	4	5	5	5	4
7	3	7	6	6	6	7	4	8	6
14	8	14	12	12	12	12	12	14	12
5	3	5	2	5	5	5	3	6	4
6	2	7	4	6	4	7	4	7	6
6	4	7	6	5	6	7	5	6	5
16	10	19	13	16	16	22	13	19	16
9	7	10	8	9	8	9	8	10	9

2	1	2	2	3	2	3	2	3	1
8	2	6	2	6	4	8	4	6	4
5	2	5	4	5	4	5	2	5	5
4	4	6	4	6	4	7	4	7	6
12	8	14	10	12	12	12	10	14	12
3	2	4	3	5	3	3	3	3	4



4	2	6	4	4	6	7	4	4	7
5	4	6	6	5	6	7	6	5	6
13	10	16	10	13	13	16	13	13	16
8	7	10	7	8	8	9	8	8	9

2	1	3	2	2	2	3	2	3	2
8	4	6	2	6	4	8	4	6	4
5	2	5	4	4	3	4	3	3	4
7	2	7	4	4	6	7	7	4	6
14	8	14	10	10	10	14	10	10	12
5	2	5	2	5	3	5	4	4	4
6	2	7	4	6	6	7	6	6	6
6	4	7	4	5	6	6	5	5	6
19	10	19	10	16	16	13	13	16	16
10	7	10	7	8	9	9	8	8	9

3	1	4	3	4	3	4	2	4	2
8	2	6	2	8	6	6	6	6	6
4	3	5	5	5	4	4	4	5	5
7	6	7	8	6	4	7	6	6	6
14	8	14	16	12	12	14	14	12	12
5	3	5	2	5	4	5	3	5	4
6	4	7	4	7	4	6	6	7	7
6	4	7	6	7	6	6	6	7	6
16	10	19	13	16	13	16	13	19	16
9	7	10	8	9	8	10	8	10	9

2	1	4	3	4	3	3	2	5	3
8	2	8	2	8	6	8	4	8	6
4	3	5	5	4	3	5	5	6	5
7	6	8	7	7	4	6	7	7	7
12	8	14	14	12	10	14	14	14	14
4	2	5	4	5	3	5	3	4	5
6	4	7	6	6	6	7	6	4	6
6	5	7	5	7	5	6	5	5	7
16	10	19	16	19	13	16	13	13	19
9	7	10	9	9	8	9	8	8	10

2	1	3	2	2	3	3	2	5	2
8	4	8	6	6	8	10	4	8	6
4	2	5	4	5	6	5	4	6	4
4	4	8	7	6	7	7	6	7	6
10	10	14	12	12	14	14	10	14	12
4	2	5	3	5	5	5	3	6	5

6	4	7	4	6	7	7	6	7	7
6	4	7	6	6	8	7	5	6	6
16	10	19	13	13	22	19	13	16	19
9	8	10	8	9	11	10	9	9	10

3	3	4	3	4	4	5	4	5	4
8	4	10	6	6	8	8	8	10	8
5	4	5	2	5	6	6	5	6	5
7	6	8	7	7	7	7	7	8	7
14	10	14	12	14	14	14	14	16	14
4	3	6	3	4	3	4	3	6	6
7	2	7	6	6	4	7	6	7	7
7	4	7	6	6	5	7	5	5	7
16	10	19	16	16	13	19	13	19	22
10	8	10	9	10	7	10	9	10	11

4	2	4	3	4	4	5	4	5	4
8	2	8	2	8	6	6	2	8	8
5	3	5	5	5	5	5	4	6	4
7	4	7	4	6	7	7	6	7	6
14	8	16	10	12	14	14	12	12	12
5	2	5	5	4	4	5	3	4	5
7	2	6	6	6	4	7	6	6	7
7	4	6	7	6	6	7	6	5	6
19	10	16	19	13	13	19	13	13	19
10	7	10	10	8	8	10	8	8	10

2	2	3	4	4	3	2	2	5	2
8	6	8	4	8	6	6	4	8	6
4	4	5	4	4	3	5	4	4	4
6	7	7	7	6	6	7	4	6	7
12	10	16	12	10	12	14	12	12	12
5	4	5	3	5	3	4	4	5	4
6	6	7	6	6	4	7	6	6	7
7	6	6	6	6	6	7	6	6	7
16	13	16	13	16	13	19	13	13	19
9	8	10	9	8	8	9	8	8	9

2	2	4	4	4	5	5	3	5	4
8	8	8	8	8	6	8	4	8	6
4	5	5	5	5	5	5	3	5	5
6	7	7	7	6	7	8	6	6	6
12	14	14	14	14	14	14	10	12	14
5	3	5	2	5	5	5	4	6	5

7	4	7	4	7	6	7	6	7	6
7	5	6	6	6	7	7	5	6	6
19	13	16	13	19	16	16	13	22	19
10	7	10	8	9	10	9	8	10	9

2	1	3	2	2	4	4	2	4	2
8	4	6	2	6	6	8	4	8	4
5	3	6	5	5	5	5	5	6	5
7	6	7	7	7	8	6	7	8	7
14	10	14	14	14	14	14	14	14	14
5	2	5	2	4	5	4	4	6	5
7	2	8	4	7	7	7	6	7	7
7	4	7	5	6	8	7	7	6	7
19	10	19	10	16	19	19	13	19	22
10	7	10	8	9	10	9	8	10	11

4	2	4	3	3	4	3	2	5	4
6	6	8	4	8	10	10	6	10	8
4	2	5	5	5	6	6	5	6	5
6	4	8	6	7	8	7	7	8	6
12	8	16	14	14	16	12	14	16	14
3	2	5	2	4	3	3	3	3	4
4	2	8	2	4	4	7	6	6	6
5	4	7	5	5	6	6	5	5	6
13	10	19	10	13	13	13	13	13	13
8	8	10	7	9	7	9	9	8	8

2	2	4	3	5	3	3	2	5	2
8	6	8	6	8	6	8	4	8	4
5	4	4	4	4	5	5	4	5	5
6	6	6	7	6	8	6	7	7	7
14	10	12	12	12	12	12	12	14	14
4	4	5	3	5	5	4	4	6	5
6	7	6	6	7	7	7	7	7	6
5	6	6	7	7	7	7	6	6	7
13	13	19	13	16	19	16	16	19	16
8	8	10	8	10	10	10	9	10	9

4	3	4	4	3	4	2	4	5	3
8	6	8	8	6	8	8	6	10	8
5	3	5	5	3	2	3	5	3	5
7	2	7	7	6	4	6	6	6	7
12	10	14	14	12	10	12	14	10	14
4	2	4	2	3	2	3	4	6	5

7	4	6	6	6	4	7	7	8	7
6	5	6	6	6	5	6	6	6	7
19	13	16	13	13	13	13	16	22	19
10	8	9	8	9	7	8	10	10	10

4	4	5	4	4	5	5	4	5	4
8	6	8	8	6	8	8	8	10	8
5	5	5	4	5	4	6	5	5	5
8	7	7	7	7	7	7	7	7	7
14	14	14	12	14	12	14	14	14	14
4	4	4	2	4	2	3	4	4	4
7	7	7	4	4	4	7	6	6	7
7	7	7	7	5	5	6	5	5	6
19	16	16	13	16	13	13	16	13	16
10	10	9	8	9	7	8	9	8	10

4	3	4	4	4	5	4	4	5	4
6	8	8	8	8	8	8	10	10	8
5	5	5	5	5	6	5	5	5	5
7	6	8	8	8	8	7	6	8	7
14	10	14	12	16	14	14	12	14	14
4	4	5	4	6	5	4	3	5	4
7	6	6	6	8	8	7	6	8	6
7	6	6	6	8	8	7	5	6	7
13	10	16	13	19	19	16	16	19	19
8	8	10	9	11	10	9	9	10	10

3	1	3	3	2	2	2	2	3	2
8	2	8	8	6	4	6	6	8	6
2	2	5	3	4	3	3	5	5	5
4	2	7	4	4	4	6	7	7	6
8	8	14	10	10	10	12	14	14	12
3	2	4	3	3	2	3	4	3	5
4	2	6	6	4	4	6	6	7	6
5	4	6	6	5	6	7	5	5	6
13	10	16	16	13	13	13	16	13	19
8	7	10	8	8	7	8	9	8	10

4	4	4	4	4	5	5	4	5	4
6	6	10	8	8	10	10	8	10	8
3	5	5	5	5	5	5	4	6	5
4	7	7	7	7	7	8	7	7	7
10	12	14	14	14	14	14	14	14	14
3	5	5	2	5	4	4	4	5	5

```

6  7  8  2  8  7  7  6  6  7
6  7  7  6  8  7  6  7  6  7
16 13 19 13 19 19 16 13 19 22
8  9  10 8  10 9  10 9  10 11
];

```

```

b=[4  4  4  4  3  4  3  2  4  3
8  8  6  6  6  6  8  6  6  6
5  5  5  4  5  5  4  4  5  4
6  7  7  6  7  7  7  4  6  4
14 12 12 14 14 12 14 12 12 12
5  5  5  4  5  6  4  4  5  5
6  6  7  6  6  8  6  6  6  6
7  6  7  6  7  8  7  6  6  6
19 16 19 16 16 19 19 16 16 16
10 9  10 9  10 9  9  9  9  10

```

```

3  4  4  4  3  4  3  3  4  3
8  8  8  8  6  8  8  6  8  6
5  4  4  6  5  3  5  3  4  4
7  7  7  7  7  4  7  6  6  4
14 12 12 14 14 12 12 10 12 10
5  4  5  5  5  5  5  3  5  5
6  6  6  6  6  7  6  2  6  6
6  6  6  6  6  7  6  5  6  6
16 16 16 19 16 19 19 13 16 16
9  9  9  10 9  10 9  8  9  10

```

```

3  3  4  4  2  4  3  2  3  3
8  8  8  8  6  8  6  2  6  6
5  4  4  5  5  5  5  2  5  4
7  7  6  7  7  6  7  2  7  4
14 14 10 14 14 12 14 8  14 12
5  4  4  5  4  5  4  3  5  5
7  4  6  7  6  7  6  2  7  6
7  5  7  8  6  7  7  4  7  7
19 16 13 19 19 16 19 13 19 16
10 9  9  10 10 9  9  7  10 10

```

```

4  4  3  4  3  4  3  2  4  3
8  8  6  8  6  6  8  6  8  6
5  5  4  5  5  5  5  3  5  3
7  7  6  6  4  6  6  4  6  4

```

14	12	12	12	12	12	14	10	12	10
5	5	4	5	4	5	5	4	5	5
7	6	7	7	4	7	7	6	6	6
6	6	7	7	6	8	7	6	6	7
19	16	16	19	16	19	19	16	16	19
9	9	9	10	9	10	10	9	9	9

3	4	3	4	3	5	3	2	4	3
6	8	6	8	6	10	8	6	8	6
5	5	5	5	5	5	5	5	5	6
7	7	7	7	7	8	6	7	7	7
14	12	14	14	14	12	14	12	14	16
5	4	4	4	4	5	5	4	5	6
7	7	7	7	6	6	7	6	6	7
7	7	7	6	6	7	7	6	6	7
19	16	16	16	16	19	19	16	19	22
10	9	10	9	10	10	8	9	10	11

4	3	3	4	3	5	3	2	4	3
8	8	4	6	2	6	6	4	6	6
5	4	5	4	5	4	5	4	5	5
7	7	7	4	7	6	7	6	6	6
14	12	12	12	14	12	14	12	14	12
5	4	4	4	5	5	4	4	5	5
7	6	7	6	7	7	6	4	6	6
7	6	7	5	6	7	7	6	6	6
16	16	16	16	16	16	19	16	16	19
10	9	9	8	10	9	9	9	9	10

3	4	3	3	3	4	3	3	4	3
8	8	6	6	6	8	8	4	6	6
5	5	4	5	4	6	4	4	5	4
7	7	6	6	6	8	7	4	6	6
14	12	12	12	12	14	14	10	12	12
4	5	5	4	4	5	5	3	4	5
6	7	6	6	6	6	6	6	6	7
6	6	7	7	6	6	6	6	5	6
16	16	16	13	13	16	19	13	16	19
9	9	9	7	9	9	8	8	8	10

4	4	3	4	3	4	3	2	4	3
6	8	6	6	6	8	8	6	6	6
4	5	5	4	5	5	4	4	4	4
6	6	6	4	6	6	6	6	6	4

12	12	12	12	12	12	14	12	12	10
5	5	5	4	4	5	5	4	4	4
6	6	6	4	6	6	7	4	6	4
6	6	6	5	6	6	6	6	6	5
16	16	16	16	16	16	16	13	16	13
9	10	9	8	9	9	10	9	9	9

3	4	4	4	3	4	4	3	4	3
8	8	8	8	6	10	8	4	8	6
5	4	5	5	5	6	6	4	5	5
8	7	8	8	7	8	7	4	8	6
16	12	16	14	14	16	14	12	14	14
4	5	5	5	6	5	5	3	4	5
6	7	7	8	7	7	6	2	6	7
5	6	7	7	7	7	6	5	5	6
13	16	19	19	19	16	19	10	16	19
9	9	10	10	10	10	10	7	9	10

3	4	4	3	3	4	3	2	4	3
6	8	6	4	4	6	8	4	4	8
6	4	5	5	5	6	5	4	4	5
8	7	7	8	7	8	7	6	7	7
14	12	12	16	14	16	14	10	12	14
5	4	5	5	5	4	4	4	5	6
8	7	7	8	7	7	6	4	6	7
7	6	6	5	7	7	8	6	6	7
19	16	16	19	19	19	19	13	16	22
10	9	9	9	10	10	10	8	9	11

3	4	4	3	3	3	3	2	4	4
8	8	4	4	4	4	8	6	6	6
5	4	4	4	4	5	4	4	3	4
7	7	7	6	4	6	6	4	4	6
14	14	14	10	10	12	12	10	10	12
4	4	5	4	4	5	4	4	3	5
7	7	7	4	4	6	7	4	4	6
6	6	7	6	6	6	8	5	5	7
16	19	16	13	13	16	19	13	13	16
9	10	10	9	8	10	10	9	8	10

3	4	4	3	2	3	3	3	3	4
8	8	6	8	6	8	8	4	10	6
5	4	4	5	5	5	4	2	5	5
7	7	6	7	7	7	6	2	7	7

12	12	12	14	12	13	14	8	14	14
3	5	4	4	4	4	4	2	2	5
6	6	6	7	4	7	6	2	2	7
5	6	6	7	6	7	6	4	4	7
16	19	16	16	13	16	19	10	10	19
8	10	9	8	8	9	10	7	7	10

4	4	4	4	3	4	3	2	4	3
6	8	6	6	6	6	6	4	4	6
4	4	5	5	5	3	4	3	4	5
6	7	6	7	7	6	6	4	6	7
12	12	12	14	14	10	14	10	12	14
3	5	5	4	5	5	4	4	4	5
4	6	6	7	6	6	7	4	4	6
5	7	7	6	7	7	7	6	6	6
16	16	19	19	16	16	16	16	13	19
8	9	9	9	9	9	8	9	8	10

4	4	4	4	3	4	3	3	4	3
8	8	6	8	6	6	8	4	6	6
4	4	5	4	5	5	4	4	5	5
6	7	6	6	7	7	6	6	7	7
12	12	12	12	14	12	12	12	12	14
4	5	5	4	5	5	4	4	5	5
6	6	7	6	6	7	6	4	7	7
6	6	6	7	7	7	7	6	6	6
16	16	16	16	16	19	19	16	16	19
9	9	9	9	9	10	10	9	10	10

2	3	4	4	3	4	3	3	4	3
8	8	6	8	6	6	8	6	6	6
6	4	6	5	5	6	4	4	5	4
7	7	8	7	8	8	7	6	7	4
14	12	16	14	14	14	12	10	12	12
4	5	5	5	5	5	5	4	5	4
7	6	7	6	7	6	6	4	6	6
6	7	7	6	7	8	7	5	6	6
19	16	19	16	19	16	19	13	16	16
10	9	10	9	10	10	9	8	9	9

3	4	4	1	2	3	3	3	4	3
6	4	6	4	6	4	6	4	6	4
3	4	5	3	4	5	2	4	2	4
6	4	6	4	4	7	4	6	4	4



10	10	10	10	10	14	12	12	8	10
3	5	5	4	4	6	4	5	2	4
6	4	7	6	6	7	6	6	4	6
6	6	7	6	6	8	6	6	5	5
16	13	16	13	16	22	16	16	10	16
9	9	9	9	9	10	8	9	7	8

4	4	4	4	3	4	3	2	4	3
8	8	6	6	6	8	8	6	6	6
5	4	5	5	5	5	5	4	3	5
6	4	7	7	7	6	7	6	7	7
12	12	14	14	14	14	14	12	12	14
5	4	4	5	5	6	5	4	4	6
6	6	7	7	6	7	7	6	7	7
6	5	7	6	7	8	7	6	7	7
16	13	16	19	16	19	22	19	19	22
9	9	9	10	10	10	10	10	9	11

4	4	4	4	3	4	3	3	4	3
8	8	6	6	6	6	8	6	6	6
4	4	4	5	4	6	4	4	4	4
7	7	7	7	4	6	6	6	6	4
12	14	12	14	10	12	14	12	12	10
4	5	5	5	5	6	4	4	5	4
6	6	8	6	7	8	6	6	6	6
5	6	8	6	7	8	5	5	6	5
16	19	19	16	19	19	19	16	16	16
9	10	9	10	9	9	9	9	9	9

3	4	4	3	3	4	3	2	4	3
6	8	6	6	6	8	8	4	6	6
5	5	5	4	6	5	5	4	5	4
7	4	6	6	6	7	7	6	6	6
14	12	14	12	12	12	14	12	12	12
4	5	4	4	5	5	4	3	5	5
6	6	7	7	7	7	7	4	6	6
6	6	7	7	7	7	7	6	6	6
16	16	16	13	19	16	19	16	19	19
9	9	9	8	10	9	9	9	9	10

3	4	3	3	3	5	4	3	4	3
8	8	6	8	6	6	8	8	6	6
6	4	4	5	5	5	4	3	4	4
7	4	7	6	7	8	6	4	6	6

14	12	12	14	16	14	14	10	12	12
5	5	4	4	5	5	4	3	4	5
7	6	7	7	7	7	6	4	6	7
7	6	7	7	7	6	7	5	6	6
19	16	16	16	19	16	16	13	16	16
10	9	9	8	10	9	9	8	9	10

4	3	3	3	3	4	3	3	4	4
8	8	6	8	6	8	8	4	8	8
5	4	5	5	5	4	4	3	5	4
7	7	7	8	7	6	6	2	7	7
14	12	14	14	14	10	14	10	14	14
5	5	4	5	5	5	5	3	5	5
6	7	7	7	7	8	6	6	7	7
6	6	7	6	7	6	6	6	6	7
16	19	16	19	19	16	19	13	19	19
10	9	10	10	10	9	9	8	10	10

4	4	4	4	3	4	3	2	4	3
8	8	6	8	6	10	8	6	8	6
5	5	5	6	5	6	4	4	5	6
7	7	7	8	8	8	6	6	7	7
14	12	14	14	14	14	14	12	12	14
5	5	4	5	4	5	4	4	4	6
6	4	7	7	4	7	7	4	4	7
6	6	7	7	6	7	7	6	6	7
16	16	19	19	16	19	19	13	13	19
9	9	9	10	9	9	8	9	9	11

4	4	3	4	3	5	3	2	4	3
6	8	8	8	6	10	8	6	8	6
5	5	5	5	5	4	4	4	5	4
6	7	7	7	7	7	6	6	7	4
12	12	14	12	12	10	12	12	14	12
4	5	4	5	4	5	4	4	5	5
6	7	7	7	6	6	6	7	7	7
6	6	6	6	6	7	6	7	7	6
16	16	16	16	16	16	16	16	16	19
9	10	10	10	9	9	10	9	10	10

3	4	3	4	3	4	3	2	4	3
6	8	8	8	6	8	8	6	8	6
4	4	4	5	4	4	4	3	3	5
4	7	7	7	7	6	6	4	6	7

10	12	12	12	12	12	14	10	10	14
4	4	4	5	5	5	5	4	4	5
6	7	7	6	6	7	7	6	6	7
5	6	7	6	7	7	7	5	6	7
16	19	16	16	19	16	19	16	16	19
9	9	9	8	10	9	10	9	9	10

3	4	4	4	3	4	3	2	4	3
8	8	8	8	6	8	8	4	8	6
5	4	5	4	5	6	5	3	5	5
7	6	7	7	7	8	7	4	7	7
14	12	12	12	14	14	14	8	14	14
4	5	5	5	5	5	4	3	5	5
6	6	7	7	7	7	6	2	7	7
7	6	6	7	7	7	7	5	6	7
16	16	16	22	19	19	19	13	19	19
9	9	9	10	10	10	10	8	10	11

4	4	4	2	2	5	4	3	3	4
8	8	8	8	6	10	8	4	8	8
5	5	5	4	5	4	5	3	3	5
6	4	6	7	7	4	7	4	4	7
14	10	12	14	14	10	12	10	10	14
4	4	3	5	3	5	5	3	3	5
7	6	6	7	6	7	6	4	4	6
6	5	6	7	6	7	7	5	6	6
16	16	16	19	13	22	19	10	13	16
10	10	9	10	9	9	10	7	8	10

3	4	4	4	3	5	3	2	4	3
8	8	8	8	6	10	6	6	8	6
4	4	5	5	5	5	4	4	5	4
6	7	8	7	7	7	6	6	7	6
12	12	14	12	12	12	14	12	14	12
4	5	5	4	4	5	4	3	4	5
4	6	7	7	7	7	7	4	6	6
6	7	7	7	7	7	7	5	6	7
16	16	16	16	16	16	16	13	16	19
9	10	10	9	9	9	10	8	9	10

4	4	4	4	3	4	3	3	4	3
8	8	8	8	6	8	8	6	8	6
4	4	4	5	5	5	4	4	4	5
6	7	7	7	7	7	6	7	6	7

```

12 12 12 14 14 12 14 14 12 14
4 5 4 5 5 5 5 4 5 6
6 7 7 6 6 7 7 6 6 7
6 6 7 6 7 7 7 5 6 8
16 19 19 16 16 19 16 13 16 22
9 10 9 10 9 10 9 9 9 11

```

```
];
```

```

x=a-b;
x=x(:);
x=x';
hist(x,70)
figure
normplot(x)
[muhat,sigmahat,muci,sigmaci]=normfit(x,0.05)

```

## 附录二：（MATLAB）

求各组方差

```

a1=var(a,0,2);
sum(a1)
b1=var(b,0,2);
sum(b1)

```

## 附录三：（MATLAB）

求特征向量、特征值、贡献率

```

clear,clc
x=[];
s=size(x);
n=s(1);
m=s(2);
stdr=std(x);
ra_m=mean(x);
sr=(x-repmat(ra_m,n,1))./repmat(stdr,n,1);
[pc,score,variance,t2]=princomp(sr)
percent_v=variance/sum(variance)
n=length(percent_v);
s0=0;
s1=zeros(n,1);
for i=1:n
    s0=s0+percent_v(i);

```

```

        s1(i)=s0;
    if s0<0.85
        j=i;
    end
end
fprintf('主成分数: %g\n',j+1)
fprintf('主成分      特征值      贡献率%%      累计贡献率%%\n')
for i=1:n

fprintf('%4g%13.4f%13.4f%13.4f\n',i,variance(i),100*percent_v(i),
100*s1(i))
end

xx=sr;
wa=pc;
r=variance;
r=r';
wb=xx*wa;
R=sum(r);
n=size(xx,1);
m=size(xx,2);
Y=zeros(m,1);
for i=1:n
    for j=1:m
        Y(i)=Y(i)+(r(j)*wb(i,j))/R;
    end
end
end
wm=sort(Y, 'descend');

```

#### 附录四：（MATLAB）

##### 模糊聚类

```

clear,clc
X=[68.1 74 74.6 71.2 72.1 66.3 65.3 66 78.2 68.8
61.6 68.3 68.8 72.6 65.7 69.9 74.5 65.4 72.6
75.8 72.2 71.6 77.1 71.5 68.2 72 71.5
1.266728651 1.077968953 1.699930555 -0.48039938 -0.47084584
-0.07938563 -0.014909719 -0.104269163 1.464050086 -0.796313328
0.823957857 -0.243280291 -0.222114575 0.056889293 -0.57142384
0.177354365 -0.365799603 -0.201549679 -0.131390039
-0.627199646 0.843125935 0.374644604 0.322499801 -0.362894364
-1.509246064 -0.993928264 -0.932200674
];

```

```

for i=1:2
    Hmax=max(X(i,:));
    Hmin=min(X(i,:));
    for j=1:10
        X(i,j)=(X(i,j)-Hmin)/(Hmax-Hmin);
    end
    x=X(i,:);
    X(i,:)=zscore(x);
    for g=1:2
        for h=1:10

            if abs(X(g,h))<1e-6
                X(g,h)=0;
            end
        end
    end
end
X(2,:)=X(2,:)*2;
X=X'
BX=zscore(X);
Y=pdist(X);
D=squareform(Y)
Z=linkage(Y)
figure
T=cluster(Z,3)
find(T==2)
[H,T]=dendrogram(Z)

```

#### 附录五：（MATLAB）

找出每个指标对应的排序

```

clear,clc
a=[77.9 75.8 75.6 76.9 81.5 75.5 74.2 72.3 80.4 79.8 71.4 72.4 73.9
77.1 78.4 67.3 80.3 76.7 76.4 76.6 79.2 79.4 77.4 76.1 79.5 74.3 77
79.6];
b=sort(a,'descend');
n=size(a,1);
for i=1:n
    w(i)=find(a==b(i));
    e(w(i))=i;
end
e=e';

```

## 附录六：（MATLAB）

### 画散点图

```
clear,clc
x=[11 15 1 18 13 23 12 14 3 22 20 28 25 26 16 24 9 6 27 2 5 17 4 7
10 19 218];
y=[3 11 23
13
24
17
25
4
19
22
8
14
9
12
15
1
2
21
5
20
6
7
10
27
16
18
28
26
];

z=[10
19
20
14
1
21
```

```

23
26
2
4
27
25
24
12
9
28
3
15
17
16
8
7
11
18
6
22
13
5
];
a=1:28;
plot(a,x)
hold on
plot(a,x,'*')

plot(a,z,'k')

plot(a,z,'k<')

figure
plot(a,y,'r')
hold on
plot(a,y,'r.')

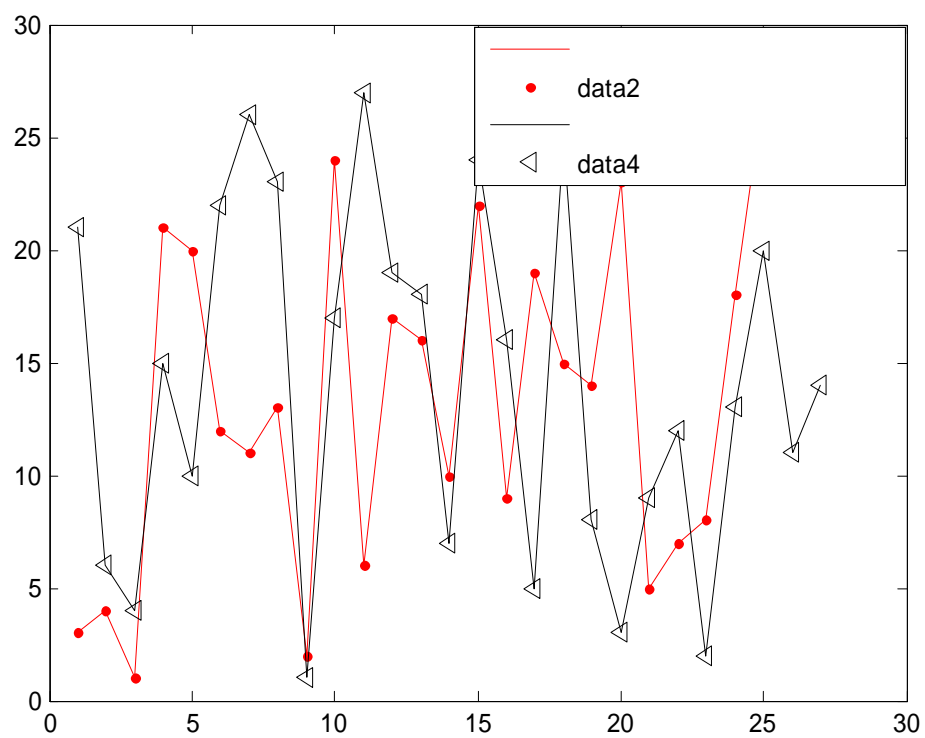
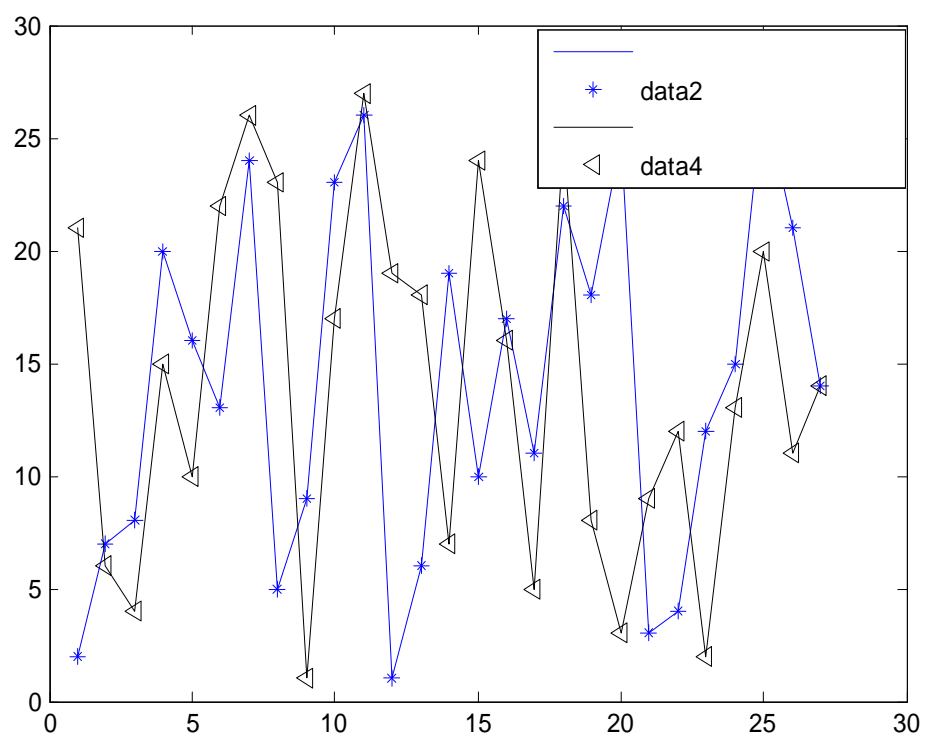
plot(a,z,'k')

plot(a,z,'k<')

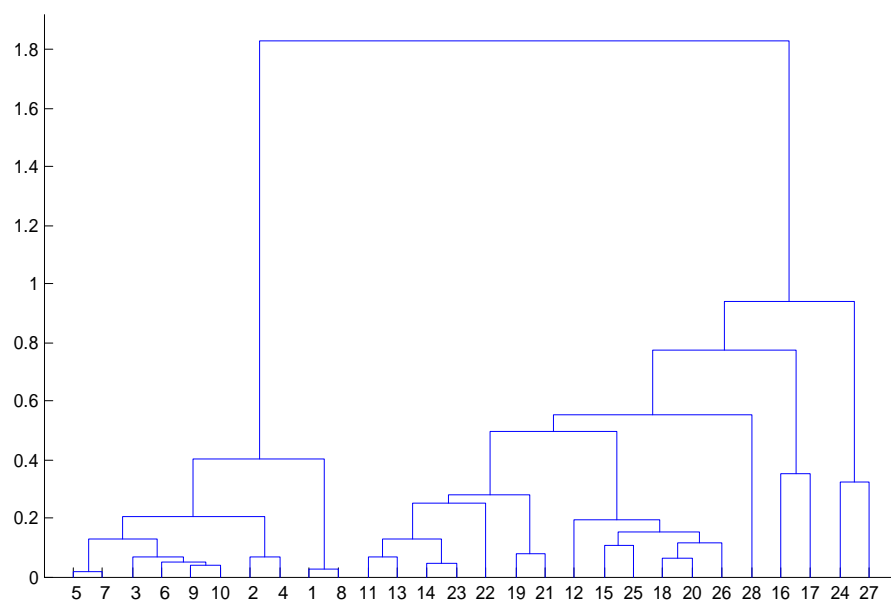
```

附录七: (MATLAB)  
红色葡萄相关性分析

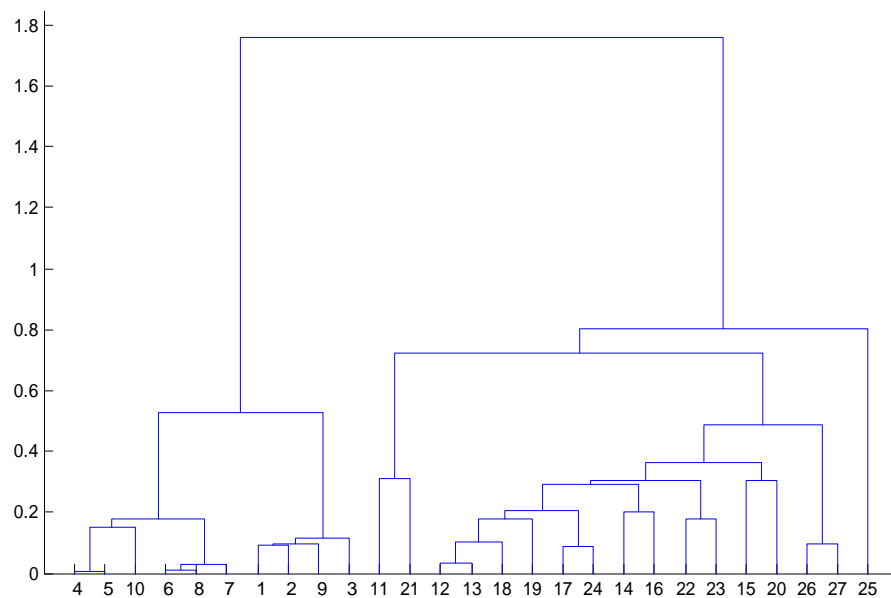




附录八：(MATLAB)



白的（葡萄+葡萄酒）



红的（葡萄+葡萄酒）