

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）： _____

所属学校（请填写完整的全名）： _____ 重庆大学 _____

参赛队员（打印并签名）： 1. _____ 邵伟华 _____

2. _____ 杨余鸿 _____

3. _____ 肖春明 _____

指导教师或指导教师组负责人（打印并签名）： _____ 肖剑 _____

日期： _____ 2012 _____ 年 _____ 9 _____ 月 _____ 10 _____ 日

赛区评阅编号（由赛区组委会评阅前进行编号）：

2012 高教社杯全国大学生数学建模竞赛

编号专用页

赛区评阅编号（由赛区组委会评阅前进行编号）：

赛区评阅记录（可供赛区评阅时使用）：

评阅人										
评分										
备注										

全国统一编号（由赛区组委会送交全国前编号）：

全国评阅编号（由全国组委会评阅前进行编号）：

基于数理分析的葡萄及葡萄酒评价体系

摘要

葡萄酒的质量评价是研究葡萄酒的一个重要领域,目前葡萄酒的质量主要由评酒师感官评定。但感官评定存在人为因素,业界一直在尝试用葡萄的理化指标或者葡萄酒的理化指标定量评价葡萄酒的质量。本题要求我们根据葡萄以及葡萄酒的相关数据建模,并研究基于理化指标的葡萄酒评价体系的建立。

对于问题一,我们首先用**配对样品 t 检验**方法研究两组评酒员评价差异的显著性,将红葡萄酒与白葡萄酒进行分类处理,用 SPSS 软件对两组评酒员的评分的各个指标以及总评分进行了配对样本 t 检验。得到的部分结果显示:红葡萄酒**外观色调、香气质量**的评价存在显著性差异,其他单指标的评价不存在显著差异,白葡萄、红葡萄以及整体的评价**存在显著性差异**。

接着我们建立了数据可信度评价模型比较两组数据的可信性,将数据的可信度评价转化成对两组评酒员评分的稳定性评价。首先我们对单个评酒员评分与该组所有评酒员评分的均值的**偏差**进行了分析,偏差不稳定的点就成为**噪声点**,表明此次评分不稳定。然后我们用两组评酒员评分的偏差的方差衡量评酒员的稳定性。得到第 2 组的方差 10.6 明显小于第 1 组的 33.3,从而得出了第 2 组评价数据的可信度更高的结论。

对于问题二,我们根据酿酒葡萄的理化指标和葡萄酒质量对葡萄进行了分级。一方面,我们对酿酒葡萄的一级理化指标的数据进行标准化,**基于主成分分析法**对其进行了因子分析,并且得到了 27 种葡萄理化指标的综合得分及其排序(见正文表 5)。另一方面,我们又对附录给出的各单指标百分制评分的权重进行评价,并用**信息熵法重新确定了权重**,用新的权重计算出 27 种葡萄酒质量的综合得分并排序(见正文表 6)。最后我们对两个排名次序用基于**模糊数学评价方法**将葡萄的等级划分为 1-5 级(见正文表 8)。

对于问题三,首先我们将众多的葡萄理化指标用主成分分析法综合成**6 个主因子**,并将葡萄等级也列为主因子之一。对葡萄的 6 个主因子,以及葡萄酒的 10 个指标用 SPSS 软件进行**偏相关分析**。之后对相关性较强的主因子和指标作**多元线性回归**。得到了葡萄酒 10 个单指标与主因子之间的多元回归方程,该回归方程就表示两者之间的联系。

对于问题四,我们首先将葡萄酒的理化指标**标准化处理**,对葡萄酒的质量与葡萄的 6 个主因子和葡萄酒的 10 个单指标作偏相关分析,并求出多元线性回归方程。该方程就表示了葡萄和葡萄酒理化指标对葡萄酒质量的影响。之后,我们通过**通径分析方法**中的逐步回归分析得到葡萄与葡萄酒的理化指标只确定了葡萄酒质量信息的 47%。从而得出了**不能用葡萄和葡萄酒的理化指标评价葡萄酒的质量**的结论。接着我们还采用通径分析中的**间接通径系数分析**求出各自变量之间通过传递作用对应变量的影响。

最后,我们对模型的改进方向以及优缺点进行了讨论。

关键词: 配对样本 t 检验 数据可信度评价 主成分分析 模糊数学评价
综合评分 信息熵 偏相关分析 多元线性回归

1 问题重述

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

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

2 模型的假设与符号的约定

2.1 模型的假设与说明

- (1) 评酒员的打分是按照加分制（不采用扣分制）；
- (2) 假设 20 名评酒员的评价尺度在同一区间（数据合理，不需要标准化）；
- (3) 每位评酒员的系统误差较小，在本问题中可以忽略不计；
- (4) 假设附件中给出的葡萄和葡萄酒理化指标都准确可靠。

2.2 符号的约定与说明

符号	符号的意义
H_0	原假设
P	显著性概率
\bar{x}_{1n}	第 1 组评酒员对第 n 号品种葡萄酒评分的平均值， $n=1,2,\dots,27$
\bar{x}_{2n}	第 2 组评酒员对第 n 号品种葡萄酒评分的平均值， $n=1,2,\dots,27$
s_{ij}^2	第一组评酒员 i 对指标 j 评分的偏差的方差， $i=1,2,\dots,10$ ， $j=1,2,\dots,27$
y_{ij}^n	第二组评酒员 i 对指标 j 评分的偏差的方差， $i=1,2,\dots,10$ ， $j=1,2,\dots,27$
\bar{x}_{ij}^n	第 1 组 10 位评酒员对 n 号酒样品第 j 项指标评分的平均分
δ	第 1 组第 i 号评酒员对 n 号酒样品第 j 项指标评分与平均值的偏差
$\bar{\delta}$	第 1 组第 i 号评酒员对其 j 项指标评分与平均值的偏差的平均值
s_i^2	第 2 组第 i 个评酒员的总体指标偏差的方差

ω_j	重新确立的第 j 项指标的权重
s^2	第 2 组 10 个评酒员的总体指标偏差的方差
y_j^n	评酒员指标 j 的平均评分, $j=1,2,\dots,10$
x_i	葡萄的第 i 项指标, $i=1,2,\dots,27$
F_i	葡萄的第 i 项因子, $i=1,2,\dots,10$
M_j	葡萄酒的第 j 项理化指标, $j=1,2,\dots,10$

3 问题一的分析与求解

3.1 问题一的分析

题目要求我们根据两组评酒员对 27 种红葡萄酒和 28 种白葡萄酒的 10 个指标相应的打分情况进行分析, 并确定两组评酒员对葡萄酒的评价结果是否有显著性差异, 然后判断哪组评酒员的评价结果更可信。

初步分析可知: 由于评酒员对颜色、气味等感官指标的衡量尺度不同, 因此两组评酒员评价结果是否具有显著性差异应该与评价指标的类型有关, 不同的评价指标的显著性差异可能会不同。同时, 由于红葡萄酒和白葡萄酒的外观、口味等指标差异性较大, 处理时需要将白葡萄酒和红葡萄酒的评价结果的显著性差异分开讨论。

基于以上分析, 我们可以分别两组品尝同一种类酒样品的评酒员的评价结果进行两两配对, 分析配对的数据是否满足配对样品 t 检验的前提条件, 而且根据常识可知评酒员对同一种酒的同一指标的评价在实际中是符合 t 检验的条件。接着我们就可以对数据进行多组**配对样品的 t 检验**, 从而对两组评酒员评价结果的显著性差异进行检验。

由于对同一酒样品的评价数据只有两组, 我们只能通过评价结果的稳定性来判定结果的可靠性。而每组结果的可靠性又最终决定于每个评酒员的稳定性, 因此将问题转化为对评酒员稳定性的评价。

3.2 配对样品的 t 检验简介

统计知识指出: **配对样本**是指对同一样本进行两次测试所获得的两组数据, 或对两个完全相同的样本在不同条件下进行测试所得的两组数据。在本问中我们可以把配对样品理解为有 27 组两个完全相同的酒样品在两组不同评酒员的检测下得到的两组数据, 两组中各个指标的数据为每组评酒员对该指标打分的平均值。

配对样品的 t 检验可检测配对双方的结果是否具有显著性差异, 因此就可以检验出配对的双方 (第一组与第二组) 对葡萄酒的评价结果是否有差异性。

配对样品 t 检验具有的前提条件为:

- (1) 两样品必须配对
- (2) 两样品来源的总体应该满足正态性分布。

配对样品 t 检验基本原理是: 求出每对的差值如果两种处理实际上没有差异, 则差值的总体均数应当为 0, 从该总体中抽出的样本其均数也应当在 0 附近波动; 反之, 如果两种处理有差异, 差值的总体均数就应当远离 0, 其样本均数也应当远离 0。这样, 通过检验该差值总体均数是否为 0, 就可以得知两种处理有无差异。该检验相应的假设为:

$$H_0: \mu_d = 0, \text{ 两种处理没有差别, } H_1: \mu_d \neq 0 \text{ 两种处理存在差别。}$$

3.3 葡萄酒配对样品的 t 检验

问题一中配对样品为 27 组两个完全相同的酒样品在两组不同评酒员的检测下得到的两组数据, 其中两组中各个指标的数据为各组 10 个评酒员对该指标打分的平均值。该问题中的 10 个指标分别为: 外观澄清度、外观色调、香气纯正度、香气浓度、香气质量、口感纯正度、口感浓度、口感持久性、口感质量、平衡/总体评价。

根据 t 检验的原理, 对葡萄酒配对样品进行 t 检验之前我们要对样品进行正态性检验。首先我们根据附件一并处理表格中的数据, 得到配对样品的两组数据, 绘制红葡萄酒配对样品表格部分数据如表 1:

表 1 红葡萄酒配对样品数据表

	澄清度(1组均值)	澄清度(2组均值)	...	平衡/整体评价(1组均值)	平衡/整体评价(2组均值)
红 1	2.3	3.1	...	7.7	8.4
红 2	2.9	3.1	...	9.6	9.1
⋮	⋮	⋮	...	⋮	⋮
红 26	3.6	3.7	...	8.9	8.8
红 27	3.7	3.7	...	9	8.8

白葡萄酒配对样品表格部分数据如表 2:

表 2 白葡萄酒配对样品数据表

	澄清度(1组均值)	澄清度(2组均值)	...	平衡/整体评价(1组均值)	平衡/整体评价(2组均值)
白 1	2.3	3.1	...	7.7	8.4

白 2	2.9	3.1	...	9.6	9.1
⋮	⋮	⋮	...	⋮	⋮
白 26	3.6	3.7	...	8.9	8.8
白 27	3.7	3.7	...	9	8.8

从上表中我们能看出，将白葡萄酒和红葡萄酒中的每个指标分别进行样品的配对后，每一个指标的配对结果有 27 对，每一对的双方分别是 1 组和 2 组的评酒员对该指标的评分的平均值。

3.3.1 样本总体的 K-S 正态性检验

配对样品的 t 检验要求两对应样品的总体满足正态分布，则总体中的样品应该满足正态性或者近似正态性，样本的正态性检验如下：

以红葡萄酒的澄清度的 27 组数据为例分析：利用 SPSS 软件绘制两样品的直方图和趋势图如图 1 所示：

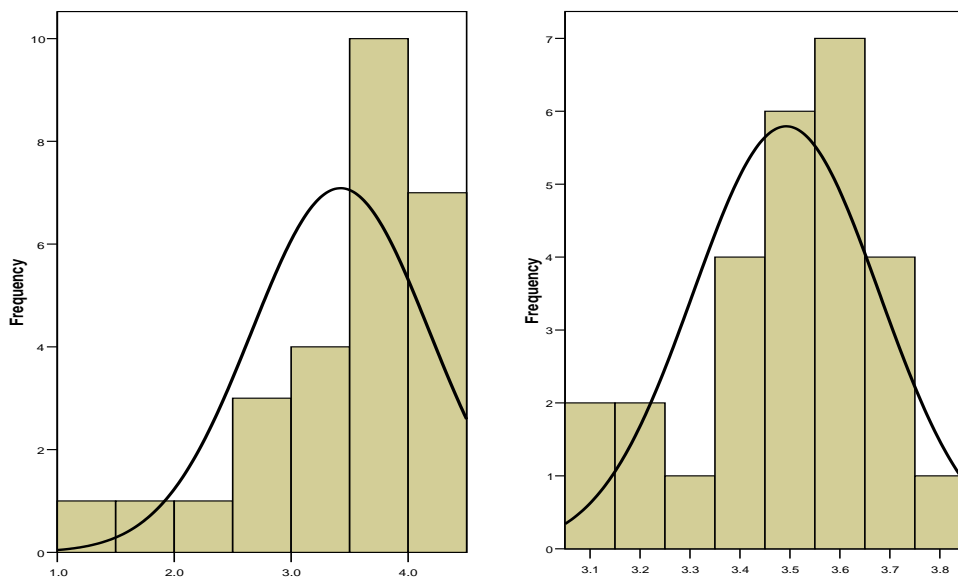


图 1 红葡萄酒澄清度两组数据直方图

我们假设两组总体数据都服从正态分布，利用 SPSS 软件进行 K-S 正态性检验的具体结果见附录 2.3。两组数据的近似相伴概率值 P 分别为 0.239 和 0.329，大于我们一般的显著水平 0.05 则接受原来假设，即两组红葡萄酒的澄清度数据符合近似正态分布。

同理可用 SPSS 软件对其他指标的正态性进行检验，得到结果符合实际猜想，

都服从近似正态分布。

3.3.2 葡萄酒配对样品 t 检验步骤

两种葡萄酒的处理过程类似,这里我们以对红葡萄酒评价结果的差异的显著性分析为例。

Step1: 我们以第一组对葡萄酒的评价结果总体 X_1 服从正态分布 $N(\mu_1, \sigma^2)$, 以第二组对葡萄酒的评价结果总体 X_2 服从正态分布 $N(\mu_2, \sigma^2)$ 。我们已分别从两总体中获得了抽样样本 $(\bar{x}_{11}, \bar{x}_{12} \cdots \bar{x}_{127})$ 和 $(\bar{x}_{21}, \bar{x}_{22} \cdots \bar{x}_{227})$, 并分别进行两样品相互配对。(具体数据见附录 2.1)

Step2: 引进一个新的随机变量 $Y = X_1 - X_2$, 对应的样本为 $(y_1, y_2 \cdots y_{27})$, 将配对样本的 t 检验转化为单样本 t 检验。

Step3: 建立零假设 $H_0: \mu = 0$, 构造 t 统计量;

Step4: 利用 SPSS 进行配对样品 t 检验分析, 并对结果做出推断。

3.4 显著性差异结果分析

3.3.1 红葡萄酒各指标差异显著性分析

由 SPSS 软件对红葡萄酒各指标的配对样品 t 检验后, 得到各指标的显著性概率 P 分布表。(结果如表 3 所示)

表 3 红葡萄酒酒各指标显著性概率 P

指标	外观澄清度	外观色调	香气纯正度	香气浓度	香气质量
P	0.614	0.002	0.151	0.100	0.010
指标	口感纯正度	口感浓度	口感持久性	口感质量	平衡/整体
P	0.437	0.158	0.251	0.055	0.674

由统计学知识, 如果显著性概率 $P < \text{显著水平 } \alpha$, ($\alpha = 0.05$), 则拒绝零假设, 即认为两总体样本的均值存在显著差异。若 $P > \text{显著水平 } \alpha$, 则不能拒绝零假设, 即认为两总体样本的均值不存在显著差异。

则根据表 3 可得: 两组评酒员对红葡萄酒各项指标的评价中除外观色调、香气质量存在显著性差异以外, 其他 8 项指标都无显著性差异。

3.3.2 白葡萄酒各指标差异显著性分析

代入白葡萄酒的评价数据，重复以上步骤，得到白葡萄酒各指标的显著性概率 P 分布表。（结果如表 4 所示）

表 4 白葡萄酒各指标显著性概率 P 分布表

指标	外观澄清度	外观色调	香气纯正度	香气浓度	香气质量
P	0.299	0.089	0.937	0.238	0.714
指标	口感纯正度	口感浓度	口感持久性	口感质量	平衡/整体
P	0.000	0.005	0.863	0.000	0.001

分析表 4 可得：两组评酒员对白葡萄酒各项指标的评价中只有口感纯正度、口感浓度、口感质量、平衡/整体评价存在显著性差异，其他 6 项指标都无显著性差异。

3.3.3 葡萄酒总体差异显著性分析

(1) 红葡萄酒总体差异显著性分析

该问题的附件中已经给出了 10 项指标的权重，因此将 10 项指标利用加权重合并成总体评价。对于红葡萄酒两组评价结果构造两组配对 t 检验。得到显著性概率 $P=0.030 < 0.05$ 。即红葡萄酒整体评价结果**有显著性差异**。

(2) 白葡萄酒总体差异显著性分析

同理对于白葡萄酒两组评价结果构造两组配对 t 检验。得到显著性概率 $P=0.02 < 0.05$ 。即白葡萄酒整体评价结果**有显著性差异**。

(3) 葡萄酒总体差异显著性分析

对于白葡萄酒和红葡萄酒总体评价结果配对 t 检验。得到显著性概率 $P=0.002 < 0.05$ 。即两组对整葡萄酒的评价**有显著性差异**。

3.5 评分数据可信度评价

3.5.1 数据可信度评价分析

前面我们已经对两组评酒员评价结果的差异显著性进行了分析，部分指标存在显著性差异，但两组评酒员对葡萄酒总体评价并无显著性差异。也即我们不能通过显著性差异指标明显地看出哪一组评酒员的数据可信。因此比较两组评酒员所评数据的可信度要建立更贴切的数据可信度指标。

3.5.2 数据可信度评价指标建立

由于整体评价数据无显著性差异，我们可以认为 20 名评酒员的水平在一个区间内。因此评酒员的评价结果的稳定性将决定该评酒员评价的数据的可信度。若某一评酒员的评价数据不稳定，则其所评数据可信度较低，其所在组别的数据评价可信度也将相应降低。

因此，我们将**数据的可信度比较转化为两组评酒员评论水平的稳定性比较**。查阅相关资料获知，评酒员的评价尺度是有一定的系统误差的。如不同评酒员对色调的敏感度或许是不同的，如果某一评酒员评价的色调稍高于标准色调，但他每次评价的色调都稍高，而且一直很稳定。虽然与均值间始终存在误差，由于其稳定性，这样的评酒员的评价数据仍然是可信的。

所以，我们建立的数据可信度评价指标为评酒员评价的稳定性。评酒员的评价数据越稳定，数据越可信。

3.5.2 数据可信度评价模型的建立与求解

我们已分析将数据可信度的评价转化为对评酒员评价稳定性评价。通过对数据的初步观察处理，发现每位评酒员的系统偏差都较小，20 位评酒员的评价尺度近似处在同一区间，因此我们不对附件中的数据进行标准化处理，认为附件中的数据的系统偏差可以忽略。

(1) 噪声点分析

首先作出观察评酒员稳定性的偏差图，其中偏差为评酒员对同一个单指标的评分值与该组评论员评分的平均值之差。下面利用 matlab 软件作出第 2 组中 1 号和 2 号评酒员对 27 种红葡萄酒的澄清度评分与组内平均值的偏差如下（程序见附录 1.1）：

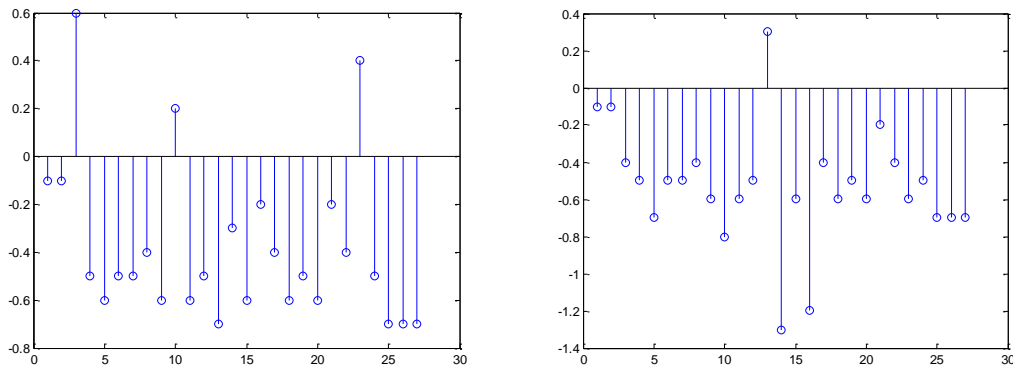


图 2 第 2 组中 1（左）、2 号评酒员对澄清度评分与组内平均值偏差图

分析上图可以看出，1 号评酒员在对 27 种酒的澄清度评分时，出现了 3 个噪声点，（即偏离自己的平均水平较大的点）。2 号评酒员在评分的时候只出现了

1 个噪声点。因而可以初步判定 2 号评酒员的稳定性比 1 号评酒员的稳定性好。

(2) 各指标偏差的方差计算

基于以上分析：要评价一个评酒员评价的稳定性，我们可以观察该评论员在评价时具有的噪声点的个数。噪声点的个数也可用评酒员的评酒数据与该组所评数据平均值的偏差的方差 s^2 进行计算衡量。

在此问中我们仍然选择两组红葡萄酒的评分求解偏差的方差。评酒员评价数据中包含 10 个评价指标，分别为外观澄清度、外观色调平衡……整体评价等。我们给它们分别标号为从 1-10。其中符号的含义为：

① i 号评论员对 j 个单指标评分的偏差的方差 s_{ij}^2 。

$$i=1,2,\dots,10 \quad j=1,2,\dots,10$$

② x_{ij}^n 表示第 1 组中 i 号评酒员对 n 号样品酒 j 号单指标的评分，其中 $i=1,2,\dots,10$ ， $j=1,2,\dots,10$ ， $n=1,2,\dots,27$ 。

③ y_{ij}^n 表示第 2 组中 i 号评酒员对 n 号样品酒 j 号单指标的评分其中 $i=1,2,\dots,10$ ， $j=1,2,\dots,10$ 。 $n=1,2,\dots,27$

在第 1 组中：

① 10 位评酒员对 n 号酒样品的 j 项指标评分的平均分为：

$$\bar{x}_{ij}^n = \frac{\sum_{i=1}^{10} x_{ij}^n}{10} \quad (1)$$

② 第 i 号评酒员对 n 号酒样品第 j 项指标评分与平均值的偏差为：

$$\delta = \frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \quad (2)$$

③ 第 i 号评酒员对酒样品的 j 项指标评分与平均值的偏差的平均值为：

$$\bar{\delta} = \frac{1}{27} \sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \right) \quad (3)$$

④ 第 i 号评酒员对酒样品的 j 项指标评分与平均值的偏差的方差为：

$$s_{ij}^2 = \frac{1}{27} \left(\frac{\sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \right)}{27} - \left(\frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \right) \right)^{27} \quad (4)$$

同理，第 2 组中第 i 号评酒员对酒样品 j 项指标评分与平均值的偏差的方差为：

$$s_{ij}^2 = \frac{1}{27} \left(\frac{\sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right)}{27} - \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right) \right)^{27} \quad (5)$$

(3) 总体的偏差的方差计算

问题 1 的附件中应该给出了 10 项单指标的权重 ω_j （每项指标的满分值），利用该权重可得到第 2 组总体指标偏差的方差为：

$$s_i^2 = \omega_j \cdot \sum_{j=1}^{10} \left(\frac{1}{27} \left(\frac{\sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right)}{27} - \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right) \right)^{27} \right) \quad (6)$$

第 2 组 10 名评酒员的 27 个酒样品的 10 项单指标的总体的偏差的方差为：

$$s^2 = \sum_{i=1}^{10} \sum_{j=1}^{10} \left(\frac{\omega_{2j}}{27} \left(\frac{\sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right)}{27} - \left(\frac{\sum_{i=1}^{10} y_{ij}^n}{10} - y_{ij}^n \right) \right)^{27} \right) \quad (7)$$

第 1 组 10 名评酒员的 27 个酒样品的 10 项单指标的总体的偏差的方差为：

$$s^2 = \sum_{i=1}^{10} \sum_{j=1}^{10} \left(\frac{\omega_{1j}}{27} \left(\frac{\sum_{n=1}^{27} \left(\frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \right)}{27} - \left(\frac{\sum_{i=1}^{10} x_{ij}^n}{10} - x_{ij}^n \right) \right)^{27} \right) \quad (8)$$

3.5.3 数据可信度评价结果分析

由附件中的数据求得：1 组的 10 名评酒员的 27 个酒品的 10 项单指标的总体的偏差的方差 $s^2 = 33.34329492$ ；2 组的 10 名评酒员的 27 个酒品的 10 项单指标的总体的偏差的方差 $s^2 = 10.6398025$ ；

因此，我们认定 2 组的评酒员的评价的稳定性较高，第 2 组的数据更可信。

3.6 问题 1 的结果分析

在本问中,我们通过对两组评酒员的品酒打分情况统计数据按照指标进行配对 t 检验,发现有部分指标存在显著性差异。接着,我们又对样本总体做了一次 t 检验,发现两组评酒员之间的评分已经不存在显著性差异。随后,我们把对每组数据可靠性的评价转化为对每组各个评酒员稳定性的评价,最后得出了第二组数据更加可靠的结论。

4 问题二模型的建立与求解

4.1 问题二的分析

题目要求我们根据酿酒葡萄的理化指标和葡萄酒的质量对酿酒葡萄进行分级。经验告诉我们,葡萄的理化指标越合理、葡萄酒的质量越好该酿酒葡萄的质量也就越好。这就要求我们分析葡萄的具体理化对葡萄的综合得分的贡献,并结合所酿葡萄酒的得分去评价葡萄的等级。

在葡萄品质的评价过程中,如果将葡萄所具备的每个理化指标不分主次进行评判不仅会增加工作量,也极有可能对评判结果产生比较大的影响。因此,必须对所考虑的众多变量用数学统计方法,经过正交化处理,变成一些相互独立、为数较少的综合指标(即主导因子)。利用主成分分析法确定出附件 2 给出的各个一级指标的主成分,在贡献率达到统计要求的情况下进行必要的因子剔除以后,保留产生主导因素的因子,把原来较多的评价指标用较少的几个综合指标来代替,综合指标既保留了原有指标的绝大多数信息,又把复杂的问题简单化。

此外,由于原有的葡萄酒评分体系的建立并不一定准确,我们考虑用熵值法重新确立在葡萄酒得分中各个指标的权重系数(即百分制的重新划分),最后和问题 1 中确定的评判标准比较,采用更准确一组的打分情况重新得到各品种葡萄酒的评价总分。

最后,根据理化指标的综合得分和葡萄酒质量的综合得分确立一个等级划分表,以这个等级划分表为依据划分葡萄的等级。

4.2 基于主成分分析的酿酒葡萄理化指标的综合评分

在问题二的分析中我们已经探讨出利用主成分分析将众多葡萄理化指标归纳到几个主成分中,并且利用主成分分析去求葡萄酒理化指标的综合得分。考虑到问题的复杂性和指标的实际意义,在此我们只选取葡萄的一级指标进行具体的数据分析。

4.2.1 基于主成分分析方法的主要步骤

Step1: 标准化数据

主成分计算是从协方差矩阵出发的，它的结果会受变量单位的影响。不同的变量往往有不同的单位，对同一变量单位的改变会产生不同的主成分，主成分倾向于多归纳方差大的变量的信息，对于方差小的变量就可能体现得不够，也存在“大数吃小数”的问题。因此，为了使主成分分析能够均等地对待每一个原始变量，消除由于单位的不同可能带来的影响，我们常常将各原始变量作标准化处理。用 matlab 软件的 zscore 函数即可得到一个矩阵的标准化矩阵。（具体程序见附录 1.2）

Step2: 计算标准化理化指标相关矩阵

考虑到本题数据的复杂性，人工进行相关矩阵显然不合理，我们借助 matlab 软件 corrcoef 函数求解标准化矩阵的相关矩阵。（具体程序见附录 1.2）

处理后的相关矩阵部分数据如表 5 所示：

表 5 酿酒葡萄理化指标相关系数表

	氨基酸总量	蛋白质	L	出汁率	果皮质量
氨基酸总量	1.0000	0.0235	L	0.0075	-0.3151
蛋白质	0.0235	1.0000	L	0.4018	-0.0991
M	M	M	L	M	M
出汁率	0.0075	0.4018	L	1.0000	-0.0185
果皮质量	-0.3151	-0.0991	L	-0.0185	1.0000

Step3: 相关矩阵的特征向量和特征值统计

数学上我们可以证明，每个因子关于原来所有因子的线性函数系数的组合就是相关矩阵的特征向量矩阵，而综合得分中每个因子的权重就是与该因子系数相对应的特征值。这里我们需要借助 matlab 软件的 eig 函数来求解相关矩阵的特征值和特征向量。（具体程序见附录 1.2）

处理后的相关矩阵的特征向量和特征值及其贡献率统计的部分数据如表 6、表 7 所示：

表 6 酿酒葡萄理化指标特征向量矩阵

	因子 1	因子 2	因子 3	因子 4	L	因子 26	因子 27
--	------	------	------	------	---	-------	-------

氨基酸总量	-0.1381	-0.2637	-0.0304	0.2811	L	-0.0652	0.0109
蛋白质	-0.2489	0.2305	-0.0014	0.1634	L	-0.1991	-0.1856
M	M	M	M	M	L	M	M
出汁率	-0.1973	0.0636	0.2439	0.0612	L	0.1577	0.0779
果皮质量	0.1172	0.0727	0.3939	-0.1267	L	0.0536	0.0343

表 7 酿酒葡萄理化指标特征值和累计率

因子	特征值	百分率	累计贡献率
1	6.6114	47.26%	47.26%
2	4.6437	23.31%	70.57%
3	2.9020	9.10%	79.67%
4	2.8345	8.69%	88.36%
5	1.9676	4.19%	92.55%
M	M	M	M
26	0	0	100%
27	0.0006	0%	100%

Step4: 计算各品种葡萄在主成分下的综合得分

从表 7 可以看出，前 4 个因子的累计贡献率已经达到 88.36%，基本信息已经包含在前 4 个因子中，符合统计学的标准。所以，我们把他们作为主成分来分析是完全可行的。

所以在我们的基于主成分分析的评价体系下，由累计贡献率得到贡献率，即作为因子的综合评分的权重，不同品种葡萄的总评价得分的表达式即为：

$$W = 0.4726F_1 + 0.2331F_2 + 0.0910F_3 + 0.0869F_4 \quad (9)$$

部分葡萄的得分和排名如下表所示：（完整的数据见附录 2.7）

表 8 不同品种酿酒葡萄品质预测评价

	因子 1	因子 2	因子 3	因子 4	总评分	排名
红 1	-4.3926	-0.6892	-0.0514	-3.2468		2
红 2	-4.4591	0.5430	0.1695	1.0701	-1.7916	4

红 3	-4.1881	-3.6548	0.4231	3.0487		1
红 4	2.4579	-0.3661	-0.8512	-0.2518	0.9544	23
N	N	N	N	N	N	N
红 26	2.3909	3.6094	-0.2997	0.3308	2.1176	26
红 27	2.0190	0.2322	-0.6908	-0.7819	0.8662	22

4.3 葡萄酒质量得分

附件 1 已经给出评酒员的具体打分情况,但是百分制打分各单项指标的分数分配不一定合理。也就是说各单项指标的权重分配不一定合理。因此,首先我们以 2 组可信度较高的评分数据,对各指标的权重进行重新分配。

4.3.1 基于信息熵对权重的重新分配

(1) 检测权重的合理性

在问题 1 中通过数据可信度的评价,我们已经得到第二组的数据更可信。在此,我们可以以 2 组的可信数据,对已知权重的合理性进行检验,若权重不合理,将重新确定权重。这里为了避免客观给定权重,我们可以根据基于信息熵的确定权重的方法重新计算信息熵并比较。

(2) 基于信息熵的确定权重方法分析

信息熵法是偏于客观的确定权重的方法,它借用信息论中熵的概念。适用于多属性决策和评价。本问题中各属性是葡萄酒的 10 项单指标(外观澄清度、气味浓度等),本问题的决策方案即是对 27 种红葡萄酒和 27 种白葡萄酒进行分级,也就是说对各属性确定权重,然后计算每种葡萄酒的总得分,最后进行排序分类。

(3) 用信息熵确定各属性权重的具体步骤:

Step1: 以 2 组评酒员对红葡萄酒各项指标的评分的平均值为信息构造决策矩阵 X , 决策变量 $X_1 \cdots X_{27}$ 为 27 种红葡萄酒, 决策的属性 $\mu_1 \cdots \mu_{10}$ 。则决策矩阵 X 为 27 行 10 列矩阵如下:

$$X = \begin{matrix} & \mu_1 & \mu_2 & \cdots & \mu_{10} \\ \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_{27} \end{matrix} & \begin{bmatrix} 3.1 & 7.6 & \cdots & 8.4 \\ 3.1 & 7 & \cdots & 9.1 \\ \vdots & \vdots & \cdots & \vdots \\ 3.7 & 6.2 & \cdots & 8.8 \end{bmatrix} \end{matrix}$$

Step2: 上述 10 个指标属性都是效应型指标, 利用公式 $r_{ij} = \frac{x_{ij}}{\max_i x_{ij}}$ 对决策矩阵

进行规范化处理, 其中 $\max_i x_{ij}$ 分别为 10 个属性得分的最高值, (如) 得到规范化决策矩阵 R 。

$$R = \begin{matrix} & \mu_1 & \mu_2 & \cdots & \mu_{10} \\ \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_{27} \end{matrix} & \begin{bmatrix} 0.62 & 0.76 & \cdots & 0.38 \\ 0.62 & 0.7 & \cdots & 0.41 \\ \vdots & \vdots & \cdots & \vdots \\ 0.74 & 0.62 & \cdots & 0.4 \end{bmatrix} \end{matrix}$$

Step3: 再由 $\tilde{x}_{ij} = \frac{x_{ij}}{\sum_{j=1}^{27} x_{ij}}$ 对规范化矩阵进行归一化处理, 得到归一化决策矩阵为 (具体数据见附录 7):

$$\square R = \begin{matrix} & \mu_1 & \mu_2 & \cdots & \mu_{10} \\ \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_{27} \end{matrix} & \begin{bmatrix} 0.033 & 0.045 & \cdots & 0.036 \\ 0.033 & 0.041 & \cdots & 0.038 \\ \vdots & \vdots & \cdots & \vdots \\ 0.039 & 0.036 & \cdots & 0.037 \end{bmatrix} \end{matrix}$$

Step4: 通过公式 $E_j = -\frac{1}{\ln n} \sum_{i=1}^n \tilde{x}_{ij} \ln \tilde{x}_{ij}$, ($n=27$) 计算 10 个属性的信息熵分别为:

E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}
0.9975	0.9946	0.9974	0.9991	0.9934	0.9947	1.0004	0.9998	0.9925	0.9998

Step5: 通过公式 $\omega_j = \frac{1-E_j}{\sum_{k=1}^{10} (1-E_k)}$ 计算我们确定的各单项的新的权重为:

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
0.0145	0.2050	0.0750	0.0619	0.2519	0.0181	0.0395	0.0085	0.3176	0.0079

4.3.2 葡萄酒质量综合得分

根据以上信息熵重新确定的各个评价指标的权重分配, 得到每种葡萄酒指标的权重向量: $\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10})$
 $= (0.0145, 0.2050, 0.0750, 0.0619, 0.2519, 0.0181, 0.0395, 0.0085, 0.3176, 0.0079)$

再根据权重和评酒员的评分就可以计算出每种葡萄酒质量的总得分为：

$$G = \omega \cdot (y^n)^T = \omega_1 y_1^n + \omega_2 y_2^n + \omega_3 y_3^n + \omega_4 y_4^n + \omega_5 y_5^n + \omega_6 y_6^n + \omega_7 y_7^n + \omega_8 y_8^n + \omega_9 y_9^n + \omega_{10} y_{10}^n$$

使用 matlab 软件进行计算（具体程序见附录 1.3）得到每种红葡萄酒质量得分和排名如下表所示：

表 9 红葡萄酒得分及排名表

品种	红 1	红 2	红 3	红 4	红 5	红 6	红 7	红 8	红 9
得分	9.664	10.89	10.83	10.32	10.44	9.441	9.291	9.460	11.39
排名	19	4	5	12	10	21	24	20	1
品种	红 10	红 11	红 12	红 13	红 14	红 15	红 16	红 17	红 18
得分	9.943	8.461	9.836	6.589	10.60	9.404	10.21	4.904	9.321
排名	16	25	17	26	7	22	15	27	23
品种	红 19	红 20	红 21	红 22	红 23	红 24	红 25	红 26	红 27
得分	10.64	10.95	10.50	10.39	11.13	10.31	9.763	10.29	10.47
排名	6	3	8	11	2	13	18	14	9

4.4 基于模糊数学对酿酒葡萄等级的划分

通过以上模型我们计算得到了酿酒葡萄理化指标的综合得分和葡萄酒质量的综合得分。若把两个综合得分处理成一个综合得分，需要用层次分析法等确定两者的权重。但层次分析过于主观，而且在本问题中，酿酒葡萄的理化指标和葡萄酒的质量对葡萄等级的影响是比较模糊和复杂的。

因此我们对得分进行排序，利用模糊数学知识进行葡萄等级的划分。

表 10 得分排名模糊划分标准

葡萄模糊等级标准	葡萄理化指标排名	葡萄酒质量排名
1 级(最高等级)	1~9	1~9
2 级	1~9	10~18
3 级	10~18	10~18
4 级	10~18	19~27
5 级（最低等级）	19~27	19~27

4.5 酿酒葡萄的等级评价结果

根据 4.4 提出的酿酒葡萄等级指标的划分, 结合 4.2, 4.3 得到的葡萄理化指标排名和葡萄酒质量排名, 得到酿酒葡萄的等级划分如表 11 所示:

表 11 红葡萄等级划分表

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

其中白葡萄的等级划分方法与红葡萄酒的划分方法相同。根据以上模型对白葡萄酒的等级划分结果如表 12 所示。(程序见附录 1.3)

表 12 白葡萄等级划分表

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

4.6 酿酒葡萄等级划分标准的评价

本问中为了最终得到酿酒葡萄的等级划分标准, 分别从酿酒葡萄的理化指标和与酿酒葡萄对应的葡萄酒的质量出发。首先, 我们基于主成分分析法逐步得到了酿酒葡萄的理化指标的综合得分, 并对其进行排名, 应用主成分分析法既避免了大量数据处理的复杂同时也尽可能地获得了最大的信息量。接着, 考虑到原有的葡萄酒评分标准不一定能够完全反应各项指标在葡萄酒质量中所起的重要性, 我们又利用熵值法重新确定了各项指标的权重系数, 得到了各品种葡萄酒在新的权重下的得分, 并得到排名。最后, 综合两个排名, 我们提出了基于模糊数学对酿酒葡萄等级的划分, 这种划分方法充分尊重了两组数据, 但是当两组数据对结果的影响因素相差很大是, 评价结果将产生较大的误差。

5 问题三模型的建立与求解

5.1 问题三的分析

题目要求我们分析酿酒葡萄与葡萄酒理化指标之间的联系。初步分析得到两者之间的联系应该体现在酿酒葡萄的理化指标和葡萄酒理化指标之间的联系。由于我们在问题二的模型中已经对酿酒葡萄进行了分级,不同等级的酿酒葡萄和葡萄酒的理化指标的联系在理论上应该是不同的。由于葡萄的理化指标数量过多,处理较复杂,我们可以用问题二的模型中提出的葡萄理化指标的主成分替代众多的葡萄理化指标。

因此,本问题就简化成葡萄的主成分与葡萄酒的理化指标的联系。基于此我们就可以对各指标进行统计分析,如相关性分析、偏相关分析,并尝试建立多元回归模型。

5.2 模型的建立

5.2.1 葡萄理化指标主成分分析

在问题 2 的主成分分析中我们已经得到红葡萄的 27 个指标可以由 4 个主因子 F_1-F_4 衡量。其中 $F_i = a_i x_i$ 的表达式中的 a_i 在主成分分析中已经给出(附录 1.2 的 MATLAB 程序的输出结果),其中 x_i 为主成分法标准化后的各葡萄理化指标的数据,编号遵循附件中一级指标的排序方式。

因子 1 和所有理化指标的关系表达式为:

$$F_1 = -0.1381x_1 - 0.2489x_2 + 0.0487x_3 + \dots - 0.1973x_{26} + 0.1172x_{27} \quad (10)$$

同理 F_2-F_4 的表达式也可以表示成一次多项式的形式。每一个因子是 27 个理化指标交互的结果,问题要求我们建立起酿酒葡萄和葡萄酒理化指标的联系,指标过多将导致联系的复杂性。所以选取贡献率最高的 4 个因子中显著性指标的交互作用代替主成分,使模型更易求解又不至于影响分析的结果。

我们得到的红葡萄的四个因子可以用理化指标线性表示为:

主要表现花色苷和总酚的因子 1:

$$F_1 = -0.3218x_4 - 0.3001x_{10} - 0.3282x_{11} - 0.2811x_{12} - 0.2741x_{13} \quad (11)$$

主要表现干物质含量和总糖的因子 2:

$$F_2 = -0.3807x_{16} - 0.3014x_{17} - 0.3821x_{18} - 0.429x_{22} \quad (12)$$

主要表现百粒重量白藜芦醇的因子 3:

$$F_3 = -0.2484x_7 + 0.2478x_{12} + 0.2593x_{13} - 0.3512x_{14} - 0.3037x_{20} + 0.3301x_{24} + 0.2439x_{26} \quad (13)$$

主要表现褐变度、蛋白质和多酚氧化酶活力 E 的因子 4:

$$F_4 = 0.2811x_1 - 0.382x_2 - 0.3625x_8 - 0.4145x_9 \quad (14)$$

同理白葡萄四个主因子与理化指标的关系可表示为:

$$\begin{aligned}
 F_1' &= -0.3284x_{16} - 0.3101x_{17} - 0.3656x_{18} - 0.3692x_{22} + 0.2691x_{23} + 0.2791x_{26} \\
 F_2' &= -0.3307x_2 - 0.2079x_4 + 0.3961x_{11} + 0.2459x_{12} + 0.393x_{13} - 0.2521x_{25} \\
 F_3' &= -0.2523x_3 + 0.2241x_4 + 0.2324x_{15} + 0.2401x_{19} - 0.4491x_{20} + 0.4303x_{21} \\
 F_4' &= -0.3705x_1 - 0.4334x_6 + 0.4036x_9 + 0.2765x_{14}
 \end{aligned} \quad (15)$$

参照文献, 我们知道芳香物质在葡萄以及葡萄酒中的决定因素都是比较明显的, 因此葡萄的第五个主要影响因素 F_5 用葡萄的芳香物质的总量来表示。

5.2.3 葡萄等级数据处理

在问题二中我们对红葡萄和白葡萄划分了等级, 根据实际情况我们可分析出葡萄的等级同样会影响葡萄和葡萄酒理化指标之间的联系。不同等级的葡萄会使联系不同。所以将葡萄的等级作为葡萄理化指标第 6 个主因子 F_6 。

葡萄等级的量化就用问题 2 中已经划分的等级数, 1 级 (优质葡萄) 就量化成数字 1, 以此类推, 量化全部品种葡萄。

因此葡萄的理化指标就可用 6 个主因子 F_1-F_6 表示。

5.2.2 偏相关分析

我们将葡萄的理化指标决定因素用 5 个主成分表示, 葡萄酒的理化指标直接用附件中的 10 个指标 (单宁、总酚、芳香物质等) 表示, 用符号分别表示成 M_1-M_{10} 。首先我们对葡萄和葡萄酒理化指标两组指标之间同时进行相关分析, 但 SPSS 软件给出的结果表示相关性很小或者没有, 然而实际上有些指标显然相关性很强, 例如葡萄酒中的总酚和葡萄中的总酚。因此数据之间存在相互干扰。所以采用偏相关分析, 得到相关系数如表 13。(弱相关未列出数据)

表 13 红葡萄酒和葡萄理化指标相关系数表

葡萄 葡萄酒	F_1	F_2	F_3	F_4	F_5	F_6
M_1 单宁	—	-0.217	-0.241	—	—	—
M_2 总酚	—	-0.261	0.246	0.256	—	—
M_3 酒总 黄酮	—	0.563	—	-0.351	-	0.316
M_4 白芦	—	—	—	—	—	—

醇						
M_5 DPPH	—	—	—	—	—	—
$L^*(D65)$ M_6	—	0.40	—	-0.228	—	0.243
$a^*(D65)$ M_7	—	—	—	-	-	0.414
$b^*(D65)$ M_8	—	0.327	-0.208	-0.231	0.216	—
花色苷 M_9	—	0.454	—	-0.292	—	—
芳香物质 M_{10}	—	-0.311	0.253	0.278	—	—

从上图中可以得到葡萄和葡萄酒理化指标相关性的关系，如：葡萄酒理化指标中的酒黄酮与葡萄的等级正相关性较强，即等级数越高（葡萄越差）酿出的葡萄酒中的酒黄酮含量就越高。葡萄酒中白芦醇和 DPPH 的含量与葡萄 6 个主成分都不相关，说明白芦醇和 DPPH 可能是由于酿造发酵决定的。

5.2.3 多元线性回归

偏相关分析已经得到了葡萄和葡萄酒理化指标的相关关系，正相关性或者负相关性较强时，对两指标可以进行线性回归。以上葡萄酒 10 个理化指标除白芦醇和 DPPH 与葡萄的 6 个主成分无相关性，其余 8 个指标都与其中若干个主成分有较强的相关性，因此对葡萄酒的 8 个理化指标可以进行多元线性回归。

例如葡萄酒中理化指标 M_3 (酒总黄酮) 与葡萄理化指标主成分 F_2 , F_4 , F_6 具有较强相关性。根据附录数据以及处理得到的主成分数据表 14。

表 14 M_3 与主成分 F_2 , F_4 , F_6 数据表

酒品	F_2	F_4	F_6	M_3
红 1	-0.6214	-1.3453	2	8.020
红 2	-0.5521	-1.4004	1	13.300
⋮	⋮	⋮	⋮	⋮

红 26	1.9884	-0.5063	4	2.154
红 27	0.6222	-0.2561	3	3.284

将数据录入 SPSS 软件中，进行多元线性回归得到酒总黄酮与 F_2 , F_4 , F_6 的关系如下：

$$M_3=9.793-0.990F_2-0.340F_4-1.555F_6 \quad (16)$$

依次将葡萄酒的各项指标以及与其相关的葡萄的理化指标主因子录入 SPSS 软件得到**红葡萄酒**的各项指标与葡萄的理化指标的主因子之间的关系式为：

①、红葡萄酒中单宁含量与葡萄各因子间的线性关系：

$$M_1=7.266-0.107F_2-0.485F_3 \quad (17)$$

②、红葡萄酒中总酚含量与葡萄各因子间的线性关系：

$$M_2=6.265-0.114F_2-0.070F_3 \quad (18)$$

③、红葡萄酒中酒总黄酮含量与葡萄各因子间的线性关系：

$$M_3=9.793-0.990F_2-0.340F_4-1.555F_6 \quad (19)$$

④、红葡萄酒中色泽度表示光泽度的指标与葡萄各因子间的线性关系：

$$M_6=48.050+2.387F_2-0.290F_4-2.212F_6 \quad (20)$$

⑤、红葡萄酒中色泽度表示红/绿色的指标与葡萄各因子间的线性关系：

$$M_7=43.579+2.158F_6 \quad (21)$$

⑥、红葡萄酒中色泽度表示黄/蓝色的指标与葡萄各因子间的线性关系：

$$M_8=21.211+0.088F_2-0.1455F_3+0.803F_4+7.196F_5 \quad (22)$$

⑦、红葡萄酒中花色苷含量与葡萄各因子间的线性关系：

$$M_9=263.899-1.648F_2-20.282F_4 \quad (23)$$

⑧、红葡萄酒中芳香物质含量与葡萄各因子间的线性关系：

$$M_{10}=0.691-0.029F_2-0.057F_4-0.019F_6 \quad (24)$$

由红葡萄的联系模型，先进行偏相关分析，再进行多元线性回归，同理得到**白葡萄酒**各理化各指标与白葡萄的 6 个主因子之间的关系式为：

$$\begin{aligned}
M'1 &= -0.328F^5 + 1.929 \\
M'2 &= 0.415F^2 + 0.039F^3 - 0.021F^4 - 0.108F^5 \\
M'3 &= 0.734 + 0.925F^2 + 0.292F^3 - 0.419F^4 + 0.289F^6 \\
M'4 &= 0.282 + 0.352F^5 \\
M'5 &= 0.055 - 0.001F^4 - 0.004F^5 \\
M'6 &= 0.689 - 0.15F^2 - 0.072F^4 - 0.059F^6 \\
M'7 &= 101.422 - 0.066F^2 - 0.044F^4 - 1.28F^6 \\
M'8 &= -0.641 - 0.052F^4 \\
M'9 &= 3.456 - 0.09F^2 - 0.17F^4
\end{aligned} \tag{25}$$

6 问题四模型的建立与求解

6.1 问题四的分析

问题要求我们分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响，同时论证能否能用葡萄和葡萄酒的理化指标来评价葡萄酒的质量。

初步分析可知酿酒葡萄质量的好坏以及葡萄酒理化指标的合理会使酿出的葡萄酒的质量较好。此处葡萄酒的质量由 2 组评酒员的评分以及我们重新确定的权重计算得出，具体影响关系我们可以首先对葡萄酒的质量和葡萄和葡萄酒的理化指标进行相关性分析，对相关性强的指标可以作为自变量与葡萄的质量进行多元线性回归。回归的结果就能定量的反映葡萄和葡萄酒的理化指标对葡萄酒质量的影响。

但回归的方程只能说明现在考虑的指标可以用该方程解释，当加入新指标该回归方程就不一定能解释了，因此我们要论证葡萄和葡萄酒的理化指标是否能唯一衡量葡萄的质量。

6.2 模型的建立与求解

红葡萄酒和红葡萄的理化指标与白葡萄酒和白葡萄的理化指标对葡萄酒质量的影响是不同的，因此需分开进行分别分析。此处以红葡萄酒和红葡萄为例分析。

在问题 3 的模型中我们通过基于主成分分析的方法将红葡萄的理化指标归纳到 6 个主因子之中，6 个主因子都作为自变量，红葡萄酒的理化指标共 10 个也全部最为自变量。因变量为葡萄的质量，即量化为问题二模型中的葡萄质量的综合得分。两者之间的影响就转化成因变量与 16 个自变量之间的关系分析。

6.2.1 数据标准化

问题 3 中 6 个主因子的数据时规范化处理后的数据, 因此彼此间不会出现大数吃小数的问题, 当变量扩展到 16 个时, 红葡萄的各理化指标 $M_1 - M_{10}$ 的数据相差比较大, 会出现大数吃小数的现象, 因此首先对 $M_1 - M_{10}$ 的数据标准化处理, 处理方法同问题二中数据标准化。

6.2.2 偏相关分析

在问题三中我们为了避免相关性分析时各数据之间产生干扰采用偏相关分析, 在问题四中我们继续采用偏相关分析方法, 用 SPSS 软件分析得到各个自变量与因变量的相关关系如下表:

表 15 各自变量与因变量之间的相关系数表

自变量	F_1	F_2	F_3	F_4	F_5	F_6	M_1	M_2
r	0	0	0	0	-0.222	-0.7811	-0.258	-0.292
自变量	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
r	0	0.017	0.305	0.691	0.555	0.433	0.731	0.442

分析上表可知因变量红葡萄的质量与红葡萄主因子 $F_1 - F_4$ 以及酒总黄酮几乎无相关, 与白芦醇弱相关, 与其他指标都是强相关。相关性强即说明线性关系比较明显。

6.2.2 多元线性回归

因变量葡萄的质量与 10 个自变量有较强的线性关系, 分别将 27 个葡萄样品数据处理成 10 自变量与 1 个因变量的关系, 可进行多元线性回归。利用 SPSS 软件回归得到回归的表达式为 (葡萄酒的质量用 y 表示):

$$y=2.412-0.494M_2-0.188M_3+8.769M_5+0.110M_6+0.064M_7+0.045M_8+0.009M_9-1.32M_{10}+0.101F_1-0.805F_2$$

由线性回归方程我们得到了红葡萄和葡萄酒的理化指标对红葡萄酒的质量的影响关系式, 分析关系式我们可以得到葡萄酒的综合质量得分 (y) 与葡萄的等级 (F_2) 是负相关的, 即红葡萄等级越高 (质量越差), 葡萄所酿出的葡萄酒的质量综合评分越低, 这是符合实际情况的; 葡萄酒的质量综合得分 (y) 中决定系数最大的是 DPPH (M_5) 以及芳香物质 (M_{10})。

多元线性回归的结果用函数能表示出来, 能反映固定指标之间的联系和影响。该问题的解决也是在其他未考虑因素不变的情况下分析的。

但回归的方程只能说明现在考虑的指标可以用该方程解释, 当加入指标该

回归方程就不能解释。因此，我们有必要论证葡萄和葡萄酒的指标是否已经能完全评价葡萄酒的质量。

6.2.3 通径分析

(1) 通径分析简介

通径分析是用来研究自变量对因变量的直接重要性和间接重要性，同时能定量给出未考虑因子的量。从而为统计决策提供可靠的依据，也可对我们的问题进行论证。

通径分析在多元回归的基础上将相关系数 r_{iy} 分解为直接通径系数（某一自变量对因变量的直接作用）和间接通径系数（该自变量通过其他自变量对因变量的间接作用）。通径分析的理论已证明，任一自变量 x_i 与因变量 y 之间的简单相关系数 (r_{iy}) = x_i 与 y 之间的直接通径系数 (P_{iy}) + 所有 x_i 与 y 的间接通径系数，任一自变量 x_i 对 y 的间接通径系数 = 相关系数 r_{ij} × 通径系数 P_{jy} 。

(2) 通径分析实现步骤

①对因变量 y 实施正态性检验：

通径分析要求要对因变量进行正态性检验。本问题中因变量综合得分 y 样本容量为 27，样本容量较大采用 **K-S 检验**，由 SPSS 软件检验后得到显著性概率 $P=0.101>0.05$ ，则认定因变量服从正态性分布。

②逐步回归分析

逐步回归分析是指从所有可供选择的自变量中逐步地选择加入或剔除某个自变量，直到建立最优的回归方程为止。SPSS 逐步回归分析的部分结果如下表：

表 16 SPSS 模型汇总表

模型	R	R 方	调整R方	标准估计的误差
1	0.848	0.719	0.543	0.927029

随着自变量被逐步引入回归方程，回归方程的相关系数 R 和决定系数 R^2 在逐渐增大，说明引入的自变量对总产量的作用在增加。最后得到决定系数 $R^2=0.719$ ，则剩余因子 $e=\sqrt{1-R^2}=0.5301$ ，该值较大，说明对因变量有影响的自变量不仅有以上逐步回归的 10 个方面，还有一些影响较大的因素没有考虑到。

因此，在此论证了葡萄和葡萄酒的理化指标是不能评价葡萄酒的。

因变量与自变量的关系，由上面 SPSS 软件逐步回归分析，得到剩余因子为 0.5301，即在确定应变变量(质量综合得分)与葡萄和葡萄酒的理化指标等自变量的函数关系时，我们只用到了 47% 的指标。只确定了葡萄酒质量信息的 47%，其余 53% 的指标可能与酿造工艺等有关。

① 途径系数计算

由途径分析中逐步回归分析步骤我们已经得到不能用葡萄和葡萄酒的理化指标来评价。

在此我们继续计算直接途径系数和间接途径系数，直接途径系数反映是多元线性的系数，也就是直接对葡萄酒质量综合得分的影响，在以上多元线性回归结果中可以直接分析影响的强弱。**间接途径系数**反映的是自变量通过影响其他自变量，再去影响因变量。因此分析间接途径系数是有意义的。

逐步回归的标准回归系数，由 SPSS 给出，见附录 2.8。

通过标准回归系数计算得到间接途径如下表

表 17 间接途径系数表

	M_1	M_2	M_5	M_6	M_7	M_8	M_9	M_{10}	F_5	F_6
M_1	-0.91 74	0.369 74	0.776 69	-1.37 06	-0.18 94	0.004 016	1.1 29	0	-0.00 0171	-0.1 318
M_2	-0.82 264	0.409	0.754 69	-1.19 97	-0.16 767	-0.00 502	0.980 73	0.000 882	-0.00 0224	-0.22 74
M_5	-0.86 723	0.378 73	0.815	-1.28 69	-0.15 09	0.018 323	0.998 45	0.001 661	-0.00 0236	-0.14 933
M_6	0.729 82	-0.28 712	-0.61 369	1.709	-0.02 6082	-0.03 012	-1.21 41	-0.00 1396	0.003	0.084 138
M_7	0.277 55	-0.11 043	-0.19 804	-0.07 1778	0.621	0.078 061	-0.53 911	0.003 630	0.000 3376	-0.15 918
M_8	-0.01 456	-0.00 818	0.059 495	-0.20 508	0.193 13	0.251	-0.44 605	0.002 851	0.000 4288	0.120 52
M_9	-0.69 615	0.271 58	0.550 94	-1.40 48	-0.22 667	-0.07 5802	1.477	-0.00 0735	-0.00 0158	-0.06 5188
M_{10}	-0.00 091	-0.02 454	-0.09 2095	0.162 36	-0.15 339	-0.04 8694	0.073 85	-0.01 47	0.0 009	0.137 2
F_5	0.097 37	-0.05 726	-0.12 062	-0.05 4688	0.131 03	0.067 268	-0.14 622	-0.00 0911	0.001 6	-0.00 8338
F_6	-0.15 834	0.122 7	0.160 56	-0.18 97	0.130 41	-0.03 9909	0.127 02	0.002 6607	0.000 0176	-0.75 8

由上表分析可知：

(1) M_1 (单宁) 相对于 M_2 (总酚) 的间接途径系数为-0.91，即 M_1 (单宁) 通过 M_2 (总酚) 的传递作用对因变量产生较强的负相关。

(2) M_1 (单宁) 相对于 M_8 (b*(D65)) 的间接途径系数为-0.01，即 M_1 (单宁) 几乎不通过 M_8 (b*(D65)) 的传递作用对因变量产生影响，同时也表明了单

宁和 $b^*(D65)$ 之间的相互作用很小。

7 模型的优缺点及改进方向

7.1 模型的优缺点

7.1.1 模型的优点:

(1) 问题一模型中配对样品的 t 检验方法利用数据配对的方法将多组数据一起进行处理, SPSS 软件操作简单。

(2) 问题一的数据可信度评价中将数据的可信度比较转化为两组评酒员评价稳定性分析, 模型得以简化。

(3) 问题二模型中用模糊数学划分等级简介合理, 用信息熵充分利用了数据信息。

7.1.2 缺点:

(1) 各模型的数据使用前基本上都需要进行标准化处理。

(2) 在处理芳香物质时, 由于芳香物质二级指标众多, 我们只对该一级指标进行了处理, 部分数据被丢失。

7.2 模型的改进方向

(1) 在建模时可以考虑将红葡萄和白葡萄用相关指标统一成变量葡萄, 避免分类处理的繁琐。

(2) 在建模时我们可以将芳香物质中的小指标先进行分析, 将模型更具体化实际化一点。

8 参考文献

- [1] 张文彤主编, SPSS 统计分析高级教程[M], 北京: 高等教育出版社, 2004。
- [2] 张波, 商豪, 应用随机过程 (第二版)[M], 北京: 中国人民大学出版社, 2009。
- [3] 马莉, MATLAB 数学实验与建模[M], 北京: 清华大学出版社, 2010。
- [4] 姜启源, 谢金星, 叶俊, 数学模型(第三版)[M], 北京: 高等教育出版社, 2003。
- [5] 李运, 李计明, 姜忠军, 统计分析在葡萄酒质量评价中的应用[J], 酿酒科技, 4:79-82, 2009。
- [6] 张丽芝, 贺兰山东麓红葡萄酒等级划分客观标准的初步研究[J], 中国食物与营养, 18(3):29-32, 2012。
- [7] 王庆华, 王庆斌, 应用数理统计方法评酒 提高汾酒质量[J], 酿酒, 37(1):47-48, 2010。
- [8] 刘保东等, 葡萄酒原汁含量的多元回归分析[J], 山东大学学报, 33(2):236-240, 1998。

附录 1

附录1.1. 偏差图绘制程序

```
(1) a=[ -0.1 -0.1 0.6 -0.5 -0.6 -0.5 -0.5 -0.4 -0.6 0.2 -0.6 -0.5  
-0.7 -0.3 -0.6 -0.2 -0.4 -0.6 -0.5 -0.6 -0.2 -0.4 0.4 -0.5 -0.7  
-0.7-0.7];  
stem(a)
```

```
(2) a=[-0.1 -0.1 -0.4 -0.5 -0.7 -0.5 -0.5 -0.4 -0.6 -0.8 -0.6 -0.5  
0.3 -1.3 -0.6 -1.2 -0.4 -0.6 -0.5 -0.6 -0.2 -0.4 -0.6 -0.5 -0.7  
-0.7 -0.7];  
stem(a)
```

附录 1.2. 红葡萄酒基于主成分分析计算总得分总程序和数据

```

clear all
clc
X=[ 2027.96      2027.957      2027.957      2027.957      2.060      18.210
1.830      7.367      9.136      6.1107      7.538      7.595      7.081      3.195
17.6780      9.318      237.668      226.5      3.56      38.66      38.66
25.918      182.93      123.6      4.51      78.4      0.110
2128.82      2128.823      2128.823      2128.823      9.930      4.750
0.770      5.150      3.557      3.1589      3.955      3.557      3.557      4.889
27.4550      11.967      229.136      228.8      3.95      44.05      44.05
25.986      81.62      98.3      3.83      77.5      0.163
8397.28      8397.284      8397.284      8397.284      8.080      2.960
1.050      4.030      2.680      2.5867      3.099      2.789      2.825      4.764
164.9927      57.527      273.758      257.6      3.91      35.99      35.99
28.997      83.13      105.4      5.60      71.8      0.170
2144.68      2144.685      2144.685      2144.685      3.770      5.230
0.550      3.183      2.988      2.2404      2.804      2.677      2.574      3.412
26.9679      10.985      237.766      203.3      3.29      28.61      28.61
23.721      137.97      174.7      3.26      53.0      0.174
1844.00      1843.996      1843.996      1843.996      9.490      3.770
1.440      4.900      3.370      3.2367      3.836      3.481      3.518      0.637
6.6502      3.602      195.460      212.9      3.64      32.00      32.00
24.084      515.46      254.2      2.99      65.6      0.270
3434.17      3434.168      3434.168      3434.168      2.830      2.210      0
1.680      1.297      0.9922      1.323      1.204      1.173      2.203      7.7272
3.701      223.817      246.1      3.29      26.43      26.43      27.376
202.24      172.0      2.64      71.9      0.193
2391.16      2391.155      2391.155      2391.155      5.820      7.740
0.540      4.700      4.327      3.1889      4.072      3.862      3.708      0.623
9.8648      4.732      303.950      211.4      3.18      25.98      25.98
26.438      63.61      168.8      4.78      71.5      0.141
1950.76      1950.760      1950.760      1950.760      5.710      13.550
2.510      7.257      7.772      5.8463      6.958      6.859      6.555      5.949
115.5546      42.686      196.990      226.5      2.92      34.99      34.99
25.620      213.09      181.1      6.41      59.6      0.260
2262.72      2262.724      2262.724      2262.724      13.230      4.120
1.100      6.150      3.790      3.6800      4.540      4.003      4.074      4.907
58.5407      22.507      194.925      203.4      3.74      34.58      34.58
23.761      186.62      138.1      5.31      78.0      0.130
1364.14      1364.139      1364.139      1364.139      2.450      2.300
0.240      1.663      1.401      1.1015      1.389      1.297      1.262      12.307
28.7475      14.106      161.421      181.2      3.65      27.16      27.16
19.676      255.44      200.8      4.59      71.7      0.200
2355.69      2355.695      2355.695      2355.695      9.290      8.610
1.900      6.600      5.703      4.7344      5.679      5.372      5.262      26.851

```

25.5751	19.229	237.891	210.2	3.53	38.24	38.24
24.527	177.83	118.8	3.41	58.4	0.102	
2556.79	2556.788	2556.788	2556.788	2556.788	6.080	5.330
1.130	4.180	3.547	2.9522	3.560	3.353	3.288
2.4802	2.155	262.155	261.1	3.43	30.58	30.58
27.614	191.95	187.7	2.40	63.3	0.243	
1416.11	1416.111	1416.111	1416.111	1416.111	4.300	0.830
1.150	2.093	1.358	1.5337	1.662	1.518	1.571
40.7586	17.731	212.237	203.4	3.86	23.75	23.75
23.353	159.97	148.0	4.67	68.1	0.160	
1237.81	1237.808	1237.808	1237.808	1237.808	5.730	4.120
1.630	3.827	3.192	2.8830	3.301	3.125	3.103
134.6375	48.018	255.335	193.9	3.39	35.90	35.90
24.060	209.11	136.3	4.60	66.2	0.255	
2177.91	2177.913	2177.913	2177.913	2177.913	6.230	3.630
2.060	3.973	3.221	3.0848	3.426	3.244	3.252
9.7179	4.393	208.933	214.9	3.19	25.09	25.09
25.012	159.31	174.5	2.90	67.7	0.213	
1553.50	1553.503	1553.503	1553.503	1553.503	9.030	7.280
2.380	6.230	5.297	4.6356	5.387	5.107	5.043
8.1900	5.930	189.275	205.6	3.30	41.76	41.76
22.346	119.17	109.3	3.79	71.8	0.135	
1713.65	1713.652	1713.652	1713.652	1713.652	5.880	5.110
0.880	3.957	3.316	2.7174	3.330	3.121	3.056
43.8121	15.860	271.504	238.2	3.43	27.51	27.51
26.276	446.64	264.1	2.80	71.5	0.330	
2398.38	2398.382	2398.382	2398.382	2398.382	3.60	5.590
0.520	3.237	3.116	2.2907	2.881	2.762	2.645
6.5161	3.192	265.773	226.6	3.27	28.21	28.21
26.338	196.01	208.4	2.60	63.1	0.160	
2463.60	2463.603	2463.603	2463.603	2463.603	5.560	4.270
0.130	3.320	2.573	2.0078	2.634	2.405	2.349
31.2649	12.478	220.333	214.9	3.57	31.54	31.54
23.441	173.09	168.8	6.32	67.4	0.162	
2273.63	2273.627	2273.627	2273.627	2273.627	3.510	0.920
0.440	1.623	0.994	1.0193	1.212	1.075	1.102
9.6262	4.091	227.338	209.1	3.81	40.48	40.48
22.933	307.14	334.3	3.15	59.5	0.232	
6346.83	6346.831	6346.831	6346.831	6346.831	15.510	2.930
2.380	6.940	4.083	4.4678	5.164	4.572	4.734
47.2196	19.934	259.110	216.9	3.56	31.99	31.99
26.948	147.66	106.1	4.74	60.4	0.108	
2566.61	2566.609	2566.609	2566.609	2566.609	6.490	7.730
0.770	4.997	4.499	3.4219	4.306	4.076	3.934

```

13.8003      7.341    226.399      234.7     3.65      39.36     39.36
25.674  106.61  115.8     3.32     57.4     0.147
2380.81      2380.811    2380.811    2380.811    4.080     5.200
0.390    3.223    2.938     2.1837    2.782     2.634     2.533     9.968
44.7476      19.083    212.564      208.8     3.39      30.23     30.23
23.383  278.75  219.1     3.84     77.5     0.233
1638.83      1638.827    1638.827    1638.827    8.360     4.600
1.700    4.887    3.729     3.4385    4.018     3.728     3.728     2.935
14.3803      7.014    244.512      203.3     3.61      27.98     27.98
25.815  517.45  237.4     2.99     76.7     0.247
1409.70      1409.703    1409.703    1409.703    2.870     2.480
0.160    1.837    1.492     1.1630    1.497     1.384     1.348     2.129
30.2112      11.229    156.038      194.6     3.38      22.81     22.81
18.515  288.69  251.3     4.10     58.5     0.220
851.17  851.169    851.169    851.169     7.150     1.400     0.820
3.123    1.781    1.9081    2.271     1.987     2.055     2.086     13.9166
6.019    197.377      195.7     3.68     42.74     42.74     19.758
793.47  245.5    3.35     68.3     0.230
1116.61      1116.612    1116.612    1116.612    6.230     1.390
1.260    2.960    1.870     2.0300    2.287     2.062     2.126     1.569
15.9809      6.559    213.216      206.9     3.37      29.67     29.67
23.329  282.09  148.7     3.51     59.5     0.200 ];

```

```
%标准化处理
```

```
[a,b]=size(X);
```

```
sigmaY=corrcoef(X);
```

```
[T,lambda]=eig(sigmaY);
```

```
disp('特征根(由小到大):');
```

```
disp(lambda);
```

```
disp('特征向量:');
```

```
disp(T);
```

```
fai=[];
```

```
%方差贡献率;累计方差贡献率
```

```
Xsum=sum(sum(lambda,2),1);
```

```
fai=lambda/sum(lambda);
```

```
disp('方差贡献率:');
```

```
disp(fai);
```

```
c=[];
```

```
c=X*T(1:27,24:27)
```

```
d=[0.0338 0.0543 0.2150 0.6383];
```

```
g=zeros(27,1);
```

```
for i=1:27
```

```
    for j=1:4
```

```
        g(i)=g(i)+c(i,j)*d(j);
```

```
    end
```



```

end
l=1:27;
l=l';
m=[g,l];
sortrows(m,1)

```

附录 1.3. 重新确立指标权重后的葡萄酒感官得分及其排序程序

```

clear all
clc
a=[3.5 6.6 4.6 6.1 12.8 4.8 6.3 6.6 17.2 9.4
3.5 7.4 4.3 6.2 12.2 4.7 5.7 6 16.6 9.2
3.1 6.6 4.4 6 12.6 4.4 5.8 6.5 16.9 9.3
3.4 7 4.5 5.6 12 4.7 6.7 6.6 17.5 9.3
3.4 7.2 5.1 7 13.6 4.6 6.6 6.6 17.8 9.6
3.4 5.6 4.6 6.3 12.8 4.5 6.2 6.3 16.6 9.2
3.3 6.6 4.6 6.3 12.4 4.4 6.2 6.1 15.7 8.6
3.4 6.6 4.4 5.6 12 4.5 5.5 5.8 15.4 9.1
3.6 7.4 5 7.1 14.2 4.7 6.7 6.1 16.6 9.4
3.3 5.8 4.9 7.2 13.4 4.7 6.7 6.5 17.8 9.5
3.3 5.8 4.1 5.7 11.8 4.2 5.6 6.2 15.4 9.3
3.2 7.2 4.4 6.3 12.5 3.7 5.3 5.8 15.4 8.6
3.5 5.8 4.2 6.2 12.4 4.4 5.6 6.4 16.6 8.8
3.6 6.6 4.5 6.5 12.4 4.6 6.2 6.4 16.9 9.4
3.3 6.8 4.9 6.9 13 4.7 6.1 6.5 16.9 9.3
3 5 3.6 4.9 10.6 4.2 5.8 6.1 15.4 8.7
3.5 6.8 4.6 6.4 13.2 4.8 6.6 6.6 18.1 9.7
3.6 6.6 4.3 6 12.2 4.7 6.5 6.1 17.5 9.2
3.3 6.4 4.8 6.1 12.6 4.4 6.3 6.5 16.9 9.1
3.5 7 4.4 6.1 13 4.4 6.4 6.4 16.3 9.1
3.4 7.2 4.4 6.4 13 4.7 6.8 6.3 17.5 9.5
3.5 7.4 5.1 7.1 13.4 4.6 5.7 6.5 16.9 9.2
3.5 7.4 4.6 6.4 12.2 4.5 6.6 6.3 16.3 9.6
3.3 7.2 4 6.1 11.8 4.5 6.5 6.3 17.2 9.2
3.4 7.2 4.7 6.7 12.8 4.6 6.2 6.5 17.8 9.6
3.5 7.6 4.4 5.6 12 4 5.9 6.1 16 9.2
3.5 7.4 4.5 6.7 12.6 4.3 6.1 6.6 16 9.3
3.6 7.4 4.4 6.7 13 4.8 6.5 6.5 17.2 9.5];
b=[0.0145 0.2050 0.0750 0.0619 0.2519 0.0181 0.0395
0.0085 0.3176 0.0079]';
c=a*b
l=1:28;
l=l';
d=[c,l];

```

```

u=sortrows(d,1)
g=28:-1:1;
g=g';
v=[u,g]

```

附录 2

附录 2.1. 红葡萄酒配对样品数据表

样品	外观 澄清度 1	外观 澄清度 2	外观 色调 1	外观 色调 2	香气 纯正度 1	香气 纯正度 2	香气 浓度 1	香气 浓度 2	香气 质量 1	香气 质量 2
红 1	2.3	3.1	6.4	7.6	4.3	3.6	5.4	5.5	12.2	10.8
红 2	2.9	3.1	7.2	7	4.5	4.5	6.5	5.6	13	12
红 3	3.4	3.4	8.6	6.8	4.7	4.2	6.2	6.2	13.2	12.2
红 4	4	3.5	8	6.4	3.4	4.2	4.7	6.1	11.2	12.2
红 5	4.3	3.6	8.4	7.2	4.5	4.1	5.9	5.4	12.6	11.4
红 6	3.9	3.5	7	5.2	4.5	3.9	6	5	12.2	11.2
红 7	4	3.5	5.8	4	4.2	3.7	5.7	5.2	11.6	11.2
红 8	2.7	3.4	7	6.8	4.7	4	6.4	5	13.6	10.4
红 9	3.1	3.6	7.4	7.4	5.5	5	7.3	6.9	14.4	13.9
红 10	4	3.8	6.8	6.8	4.7	4.4	6.2	4.8	12.6	11.8
红 11	4	3.6	4.6	3.4	4.4	3.8	6.4	5.9	12.6	11
红 12	1.1	3.5	4	5.2	2.7	3.7	4.2	4.9	9	11.2
红 13	2.6	3.7	7.6	5.8	4.6	4.8	5.8	5.5	12.8	12
红 14	3.7	3.3	8.2	7.4	4	4.2	4.8	5.8	11.6	11.8
红 15	3.9	3.6	7.6	6.4	2.4	3.1	4	5.6	9	10.2
红 16	3.1	3.2	7.4	6.8	4.7	3.8	6	5.1	12.6	11.8
红 17	1.9	3.4	5	6.8	2.9	4.8	5.1	6.3	10	2.2
红 18	3.9	3.6	7.8	4.2	4.8	3.3	5.9	4.8	12.8	10.6
红 19	3.9	3.5	8	7	4.6	4.4	6.4	5.9	13	12.2
红 20	3.7	3.6	5.6	4.8	5.2	4.9	7.3	6.8	14	13.2
红 21	3.5	3.2	8	7	4.4	3.7	6.4	5.9	12.2	11.8
红 22	3.9	3.4	8	6	4.5	4.3	6.7	6.2	12.8	11.6
红 23	3.2	3.6	8.2	7.8	5.3	4.6	7.4	6.7	14.6	13.8
红 24	4.1	3.5	8	6.6	4.5	4.2	6.6	5.8	12.6	12
红 25	4	3.7	6.4	6.6	4.4	4.3	5.3	5.2	12	11.6
红 26	3.6	3.7	7.8	7.4	4.7	4.5	6	5.6	12.8	12
红 27	3.7	3.7	6.2	6.2	4.2	4.1	5.6	5.4	11.8	12

样品	外观 澄 清 度 1	外观 澄 清 度 2	外观 色 调 1	外观 色 调 2	香 气 纯 正 度 1	香 气 纯 正 度 2	香 气 浓 度 1	香 气 浓 度 2	香 气 质 量 1	香 气 质 量 2
红 1	2.9	3.8	3.9	5.1	5	5.5	12.4	14.2	7.6	8.2
红 2	4.2	4	6.1	5.5	6	5.8	15.7	15.1	9.1	8.8
红 3	3.3	4.2	5	6.1	5.4	6.3	13.6	5.4	7.9	9
红 4	4.7	3.8	6.6	5.3	6.4	5.9	17.2	15.1	9.2	8.8
红 5	4.2	3.9	6.5	5.3	6.5	6	16.3	15.7	9.2	8.7
红 6	4.4	4.2	6.4	6.1	6.2	6	16.6	16.9	9.2	9.3
红 7	4.2	4	6.3	5.8	6	6.1	16.9	15.7	9.2	9
红 8	4.6	3.9	6.2	5.6	5.8	5.7	15.7	16	9	8.9
红 9	4.8	4.4	7	6.4	7	5.8	18.1	15.1	10	8.9
红 10	4.3	3.9	6.2	5.7	5.9	5.9	16.6	15.1	9.1	8.8
红 11	4.1	3.9	4.9	4.8	5.6	5.6	14.2	13.9	8.3	8.6
红 12	4.1	4.1	5.4	5.5	5.7	5.9	14.8	14.5	8.9	8.8
红 13	4.4	4.2	6	5.2	6.1	5.9	16	16	9	8.8
红 14	2.9	3.8	3.9	5.1	5	5.5	12.4	14.2	7.6	8.2
红 15	4.2	4	6.1	5.5	6	5.8	15.7	15.1	9.1	8.8
红 16	3.3	4.2	5	6.1	5.4	6.3	13.6	5.4	7.9	9
红 17	4.7	3.8	6.6	5.3	6.4	5.9	17.2	15.1	9.2	8.8
红 18	4.2	3.9	6.5	5.3	6.5	6	16.3	15.7	9.2	8.7
红 19	4.4	4.2	6.4	6.1	6.2	6	16.6	16.9	9.2	9.3
红 20	4.2	4	6.3	5.8	6	6.1	16.9	15.7	9.2	9
红 21	4.6	3.9	6.2	5.6	5.8	5.7	15.7	16	9	8.9
红 22	4.8	4.4	7	6.4	7	5.8	18.1	15.1	10	8.9
红 23	4.3	3.9	6.2	5.7	5.9	5.9	16.6	15.1	9.1	8.8
红 24	4.1	3.9	4.9	4.8	5.6	5.6	14.2	13.9	8.3	8.6
红 25	4.1	4.1	5.4	5.5	5.7	5.9	14.8	14.5	8.9	8.8
红 26	4.4	4.2	6	5.2	6.1	5.9	16	16	9	8.8
红 27	2.9	3.8	3.9	5.1	5	5.5	12.4	14.2	7.6	8.2

附录 2.2. 白葡萄酒配对样品数据表

样品	外观 澄 清 度 1	外观 澄 清 度 2	外观 色 调 1	外观 色 调 2	香 气 纯 正 度 1	香 气 纯 正 度 2	香 气 浓 度 1	香 气 浓 度 2	香 气 质 量 1	香 气 质 量 2
白 1	3.8	3.5	3.4	6.6	5	4.6	7	6.1	13.4	12.8
白 2	3.3	3.5	7.6	7.4	4.8	4.3	6.5	6.2	12.6	12.2
白 3	3.7	3.1	7.5	6.6	4.2	4.4	6.4	6	12.5	12.6
白 4	4	3.4	5.4	7	4.7	4.5	6.4	5.6	12.8	12
白 5	2	3.4	5.4	7.2	4.7	5.1	6.4	7	12.8	13.6
白 6	3.2	3.4	6.4	5.6	3.8	4.6	5.5	6.3	11.4	12.8
白 7	3.6	3.3	6.2	6.6	4.9	4.6	6.5	6.3	13.2	12.4

白 8	3	3.4	6.6	6.6	4	4.4	5.8	5.6	11.8	12
白 9	4.1	3.6	8	7.4	4.3	5	16.5	7.1	12.6	14.2
白 10	2.2	3.3	5.4	5.8	5.1	4.9	7.1	7.2	13.6	13.4
白 11	3.6	3.3	6	5.8	4.5	4.1	6	5.7	12.2	11.8
白 12	2.1	3.2	5	7.2	4.2	4.4	5.2	6.3	11.6	12.5
白 13	2.2	3.5	5.2	5.8	3.7	4.2	5.4	6.2	11.2	12.4
白 14	3	3.6	5.6	6.6	4.4	4.5	6.3	6.5	12.8	12.4
白 15	3	3.3	6	6.8	4.5	4.9	6.6	6.9	12.6	13
白 16	2.5	3	6.8	5	4.5	3.6	6.2	4.9	12.2	10.6
白 17	3.9	3.5	7.6	6.8	4.9	4.6	7.1	6.4	13.6	13.2
白 18	3.9	3.6	5.8	6.6	4.7	4.3	6.1	6	12.4	12.2
白 19	2.9	3.3	6.4	6.4	4.1	4.8	6.3	6.1	12.2	12.6
白 20	3.8	3.5	3.2	7	4.7	4.4	6.6	6.1	13.2	13
白 21	2.6	3.4	5.6	7.2	5	4.4	7	6.4	13.6	13
白 22	3.4	3.5	7.6	7.4	4.9	5.1	6.7	7.1	13.6	13.4
白 23	3.1	3.5	6.6	7.4	4.5	4.6	6.6	6.4	12.4	12.2
白 24	3.6	3.3	7.6	7.2	3.9	4	5.8	6.1	12.2	11.8
白 25	4.4	3.4	7.8	7.2	4.9	4.7	7.1	6.7	13.6	12.8
白 26	4.1	3.5	8.2	7.6	5.1	4.4	7.3	5.6	13.4	12
白 27	2.3	3.5	6.2	7.4	3.7	4.5	5.1	6.7	11.2	12.6
白 28	4.3	3.6	8.4	7.4	4.8	4.4	6.8	6.7	13.4	13
样品	口 感 纯 正 度 1	口 感 纯 正 度 2	口 感 浓 度 1	口 感 浓 度 2	口 感 持 久 性 1	口 感 持 久 性 2	口 感 质 量 1	口 感 质 量 2	整 体 评 价 1	整 体 评 价 2
白 1	4.7	4.8	6.7	6.3	6.5	6.6	17.8	17.2	9.7	9.4
白 2	3.8	4.7	5.8	5.7	5.9	6	14.8	16.6	9.1	9.2
白 3	4.5	4.4	6.2	5.8	13.2	6.5	17.5	16.9	9.6	9.3
白 4	3.8	4.7	5.9	6.7	5.7	6.6	15.4	17.5	8.9	9.3
白 5	3.8	4.6	5.9	6.6	5.7	6.6	15.4	17.8	8.9	9.6
白 6	3.5	4.5	5.7	6.2	5.5	6.3	14.8	16.6	8.6	9.2
白 7	4.5	4.4	6.4	6.2	6.2	6.1	16.6	15.7	9.4	8.6
白 8	3.9	4.5	5.7	5.5	6.7	5.8	15.1	15.4	8.8	9.1
白 9	3.5	4.7	5.4	6.7	5.9	6.1	13.9	16.6	8.7	9.4
白 10	4.1	4.7	6.1	6.7	6.1	6.5	15.4	17.8	9.2	9.5
白 11	4.3	4.2	5.3	5.6	5.7	6.2	16	15.4	8.7	9.3
白 12	3.3	3.7	4.8	5.3	5.6	5.8	13.3	15.4	8.2	8.6
白 13	3.9	4.4	5.6	5.6	5.4	6.4	14.8	16.6	8.5	8.8
白 14	4.1	4.6	5.8	6.2	6.1	6.4	15.1	16.9	8.8	9.4
白 15	4	4.7	5.8	6.1	5.8	6.5	15.4	16.9	8.7	9.3
白 16	4.3	4.2	6.1	5.8	6.1	6.1	16	15.4	9.3	8.7
白 17	4.2	4.8	5.9	6.6	5.9	6.6	16.3	18.1	9.4	9.7
白 18	4.2	4.7	5.7	6.5	6.6	6.1	15.4	17.5	8.9	9.2

白 19	4.2	4.4	6.1	6.3	6.3	6.5	15.1	16.9	8.6	9.1
白 20	4.5	4.4	6.1	6.4	6.1	6.4	16.6	16.3	9	9.1
白 21	4.2	4.7	6.2	6.8	6.4	6.3	16.6	17.5	9.2	9.5
白 22	3.2	4.6	4.9	5.7	5.4	6.5	13	16.9	8.3	9.2
白 23	4.5	4.5	6.6	6.6	6.4	6.3	16	16.3	9.2	9.6
白 24	3.5	4.5	6.2	6.5	5.9	6.3	15.7	17.2	8.9	9.2
白 25	3.5	4.6	5.9	6.2	6	6.5	15.1	17.8	8.8	9.6
白 26	4.4	4	6.8	5.9	6.6	6.1	16	16	9.4	9.2
白 27	3.2	4.3	5.1	6.1	5.5	6.6	14.2	16	8.3	9.3

附录 2.3. K-S 正态性检验软件输出结果

		外观澄清晰度1
N		27
Normal	Mean	3.422
Parameters(a,b)	Std. Deviation	.7597
Most Extreme	Absolute	.198
Differences	Positive	.149
	Negative	-.198
Kolmogorov-Smirnov Z		1.030
Asymp. Sig. (2-tailed)		.239

a Test distribution is Normal.

b Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		外观澄清晰度2
N		27
Normal	Mean	3.493
Parameters(a,b)	Std. Deviation	.1859
Most Extreme	Absolute	.183
Differences	Positive	.097
	Negative	-.183
Kolmogorov-Smirnov Z		.949
Asymp. Sig. (2-tailed)		.329

a Test distribution is Normal.

b Calculated from data.

附录 2.4. t 检验显著性概率 p 统计表结果

2. 接收两组评酒员对白葡萄酒酒评价无显著性的概率：

指标	外观澄清度	外观色调	香气纯正度	香气浓度	香气质量	口感纯正度	口感浓度	口感持久性	口感质量	整体评价
P	0.299	0.089	0.937	0.238	0.714	0	0.005	0.863	0	0.001

2. 接收两组评酒员对红葡萄酒酒评价无显著性的概率：

指标	外观澄清度	外观色调	香气纯正度	香气浓度	香气质量	口感纯正度	口感浓度	口感持久性	口感质量	整体评价
P	0.614	0.002	0.151	0.100	0.010	0.437	0.158	0.251	0.055	0.674

附录2.5. 葡萄酒总体配对样品数据表

酒样号	第一组红酒	第一组白酒	第二组红酒	第二组白酒
1	62.7	82	68.1	73.9
2	80.3	74.2	74	77.1
3	80.4	85.3	74.6	76.5
4	68.6	79.4	71.2	76.9
5	73.3	71	72.1	81.5
6	72.2	68.4	66.3	77
7	71.5	77.5	65.3	74.2
8	72.3	71.4	66	72.3
9	81.5	72.9	78.2	80.4
10	74.2	74.3	68.8	79.8
11	70.1	72.3	61.6	71.4
12	53.9	63.3	68.3	72.4
13	74.6	65.9	68.8	80.3
14	73	72	72.6	75.5
15	58.7	72.4	65.7	77.9
16	74.9	74	69.9	67.3
17	79.3	78.8	74.5	79.6
18	60.1	73.1	65.4	78.4
19	78.6	72.2	72.6	76.4
20	78.6	77.8	75.8	76.6
21	77.1	76.4	72.2	79.2
22	85.6	71	71.6	79.4
23	78	75.9	77.1	82.1
24	69.2	73.3	71.5	76.1

25	73.8	77.1	68.2	79.5
26	73	81.3	72	74.3
27	77.2	64.8	71.5	76.7
28		81.3		79.4

附录2.6. 归一化后的决策矩阵

酒样号	外观澄清度 ₂	外观色调 ₂	香气纯正度 ₂	香气浓度 ₂	香气质量 ₂	口感纯正度 ₂	口感浓度 ₂	口感持久性 ₂	口感质量 ₂	整体评价 ₂
1	0.03 287	0.04 455	0.032 11	0.035 92	0.035 05	0.035 48	0.038 26	0.03 819	0.03 521	0.03 550
2	0.03 287	0.04 103	0.040 14	0.036 58	0.038 95	0.038 28	0.040 27	0.03 819	0.04 298	0.03 846
3	0.03 606	0.03 986	0.037 47	0.040 50	0.039 60	0.041 08	0.040 94	0.03 883	0.04 221	0.03 762
4	0.03 712	0.03 751	0.037 47	0.039 84	0.039 60	0.036 41	0.035 57	0.03 628	0.03 910	0.03 719
5	0.03 818	0.04 220	0.036 57	0.035 27	0.037 00	0.038 28	0.037 58	0.03 883	0.04 065	0.03 762
6	0.03 712	0.03 048	0.034 79	0.032 66	0.036 35	0.036 41	0.034 90	0.03 565	0.03 677	0.03 635
7	0.03 712	0.02 345	0.033 01	0.033 96	0.036 35	0.034 55	0.036 24	0.03 628	0.03 755	0.03 550
8	0.03 606	0.03 986	0.035 68	0.032 66	0.033 76	0.034 55	0.032 89	0.03 501	0.03 599	0.03 550
9	0.03 818	0.04 338	0.044 60	0.045 07	0.045 12	0.039 22	0.040 94	0.03 819	0.04 143	0.03 973
10	0.04 030	0.03 986	0.039 25	0.031 35	0.038 30	0.035 48	0.034 23	0.03 628	0.03 677	0.03 550
11	0.03 818	0.01 993	0.033 90	0.038 54	0.035 70	0.031 75	0.030 87	0.03 437	0.03 211	0.03 423
12	0.03 712	0.03 048	0.033 01	0.032 01	0.036 35	0.037 35	0.040 27	0.03 628	0.03 988	0.03 677
13	0.03 924	0.03 400	0.042 82	0.035 92	0.038 95	0.037 35	0.033 56	0.03 501	0.01 010	0.03 635
14	0.03 499	0.04 338	0.037 47	0.037 88	0.038 30	0.035 48	0.037 58	0.03 628	0.04 065	0.03 931
15	0.03 818	0.03 751	0.027 65	0.036 58	0.033 11	0.035 48	0.034 23	0.03 501	0.03 677	0.03 466
16	0.03 393	0.03 986	0.033 90	0.033 31	0.038 30	0.037 35	0.036 91	0.03 692	0.03 910	0.03 719

17	0.03 606	0.03 986	0.042 82	0.041 15	0.007 14	0.039 22	0.040 94	0.04 010	0.01 398	0.03 804
18	0.03 818	0.02 462	0.029 44	0.031 35	0.034 40	0.035 48	0.035 57	0.03 756	0.03 910	0.03 719
19	0.03 712	0.04 103	0.039 25	0.038 54	0.039 60	0.036 41	0.035 57	0.03 819	0.04 065	0.03 677
20	0.03 818	0.02 814	0.043 71	0.044 42	0.042 84	0.039 22	0.040 94	0.03 819	0.04 376	0.03 931
21	0.03 393	0.04 103	0.033 01	0.038 54	0.038 30	0.037 35	0.038 93	0.03 883	0.04 065	0.03 804
22	0.03 606	0.03 517	0.038 36	0.040 50	0.037 65	0.036 41	0.037 58	0.03 628	0.04 143	0.03 762
23	0.03 818	0.04 572	0.041 03	0.043 76	0.044 79	0.041 08	0.042 95	0.03 692	0.03 910	0.03 762
24	0.03 712	0.03 869	0.037 47	0.037 88	0.038 95	0.036 41	0.038 26	0.03 756	0.03 910	0.03 719
25	0.03 924	0.03 869	0.038 36	0.033 96	0.037 65	0.036 41	0.032 21	0.03 565	0.03 599	0.03 635
26	0.03 924	0.04 338	0.040 14	0.036 58	0.038 95	0.038 28	0.036 91	0.03 756	0.03 755	0.03 719
27	0.03 924	0.03 634	0.036 57	0.035 27	0.038 95	0.039 22	0.034 90	0.03 756	0.04 143	0.03 719

附录 2.7 重新确定指标权重后红葡萄的得分和排名

	因子 1	因子 2	因子 3	因子 4	总评分	排名
红 1	-4.3926	-0.6892	-0.0514	-3.2468	-2.4278	2
红 2	-4.4591	0.5430	0.1695	1.0701	-1.7916	4
红 3	-4.1881	-3.6548	0.4231	3.0487	-2.6198	1
红 4	2.4579	-0.3661	-0.8512	-0.2518	0.9544	23
红 5	0.3854	1.2552	1.7878	0.4518	0.7445	18
红 6	1.5563	-2.8578	2.0830	0.2229	0.2631	14
红 7	1.7860	-2.8074	-0.5750	-1.1498	0.0026	13
红 8	-3.1443	-0.3863	-0.1513	-5.3347	-2.0963	3
红 9	-5.4585	2.9392	0.8145	2.0854	-1.6647	5
红 10	1.9618	4.3900	-0.6832	-0.0806	1.7032	24
红 11	0.9435	-0.5419	-4.8466	1.0441	-0.0004	12
红 12	2.3629	-3.8522	0.8961	0.4250	0.3749	16
红 13	-0.2636	1.7569	-1.1606	1.2339	0.2885	15

红 14	-1.7438	0.8277	-0.7050	-2.1533	-0.8873	7
红 15	1.5297	-0.6788	0.8088	-1.1762	0.4492	17
红 16	-0.4274	0.7809	-2.7070	-0.3775	-0.2594	10
红 17	1.4464	-1.0828	3.3088	0.3193	0.7703	20
红 18	2.7485	-2.5054	0.2926	-0.2661	0.7460	19
红 19	-0.5816	0.3025	0.1603	0.1525	-0.1994	11
红 20	2.4161	1.5288	0.4952	1.6466	1.7826	25
红 21	-1.6432	-1.8341	-2.4030	2.1968	-1.3159	6
红 22	0.0798	-1.7251	-1.3480	0.2967	-0.4011	9
红 23	-2.4994	1.6684	3.0801	0.5259	-0.4824	8
红 24	1.3476	0.1990	1.3121	0.3633	0.8603	21
红 25	3.3695	2.9485	0.8408	-0.5952	2.2227	27
红 26	2.3909	3.6094	-0.2997	0.3308	2.1176	26
红 27	2.0190	0.2322	-0.6908	-0.7819	0.8662	22

附录 2.8 直接通径系数表

总酚	酒总 黄酮	DPP H	L*(D65)	a*(D6 5)	b*(D6 5)	花色 苷	芳香 物质	F_5	葡萄等 级 F_6
P_{1y}	P_{2y}	P_{3y}	P_{4y}	P_{5y}	P_{6y}	P_{7y}	P_{8y}	P_{9y}	P_{10y}
-0.91 0	0.409	0.8 15	1.709	0.621	0.251	1.477	-0.1 47	0.0 16	-0.758