

题目：葡萄酒的评价

摘要

本题是利用葡萄酒品尝评分表、葡萄和葡萄酒的理化指标、葡萄和葡萄酒的芳香物质对葡萄酒进行评价。

问题一，我们以评酒员对各分类指标打分的总分表示该评酒员对这种酒的总评价。由于每个评酒员评分相互独立又有随机误差，我们用方差显著性检验的方法检验了两组评酒员对每种酒的评价是否有显著性差异，我们用 EXCEL 的统计函数进行计算，得到结论：两组评酒员对第 2、11、12、13、16、24 种红葡萄酒和第 5、27 种白葡萄酒的评价具有显著性差异。又因为两组评酒员对同一种酒评分的分差是随机的且服从正态分布，用 t-检验来检验两组评酒员总体评价是否存在显著性差异，用 EXCEL 计算得到结论：两组评酒员的评价具有显著性差异。我们认为方差较小的一组更可信，即第二组的结果更可信。然后进行了误差分析。

问题二，我们用主成分分析法对每种葡萄按照其理化指标综合打分，然后引入权值 λ 对葡萄酒分数和理化指标分数进行加权得到总分，用 MATLAB 编程计算，按总分分数段分别将红葡萄和白葡萄分成 5 类，得到葡萄的分类如下：

红葡萄的分类：10、25； 4、6、7、8、11、13、15、18、26、27； 1、5、12、14、16、17、19、20、22、24； 2、21、23； 3、9。

白葡萄的分类：16； 8、11； 1、2、4、12、13、14、17、19、22、26； 3、6、7、9、10、15、18、20、21、23、24、25、28； 5、27；

问题三，我们用两种模型求解：模型一对葡萄的每种理化指标和葡萄酒的理化指标分别作相关性检验，对每个因变量取相关性较强的 5 个自变量进行多元线性拟合，用 MATLAB 的 regress() 命令求解，得到关系式见式(4.3.2)和(4.3.3)。模型二先对理化指标进行聚类分析，分成了 5 类，以每类平均相关性指标为自变量建立多元线性回归模型，仍然用 MATLAB 求解，得到关系式见式(4.3.4)和(4.3.5)，并进行统计检验，方程回归效果显著。

问题四，我们认为影响葡萄酒质量的主要因素有五类指标：酸类、糖类、酚醇类、色素和芳香物质，其中四类是理化指标。评价葡萄酒质量主要有三个方面：外观、香气和口感。我们对这五类指标进行量化处理，探究五类指标和三个方面的关系，看芳香物质对葡萄酒质量影响的大小，我们得到结论：芳香物质对葡萄酒的影响较大，不能忽略，因此不能仅用理化指标来衡量葡萄酒的质量。并用层次分析法进行验证，得到的结果一致。

关键词：显著性检验；主成分分析；聚类分析；相关性；层次分析模型；EXCEL；MATLAB；SPSS

正文

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、 基本假设：

1. 每个评酒员对葡萄酒质量的评估结果为分类指标分数求和得到的总分；
2. 评酒员对葡萄酒的评价距标准有一定的误差，这个误差是随机的；
3. 每个评酒员都是经过认真品尝后进行评分，所有分数都是可信的，且相互独立；
4. 显著性水平 α 取 0.05；
5. 每种葡萄酒是由其对应编号的葡萄酿成的；
6. 所有种类葡萄酒都是用同种方法酿制，不会因为酿制方法不同而导致酿酒葡萄的成分发生不同的变化；
7. 影响葡萄酒质量的理化指标有糖类、酸类、酚醇类和色素类，其他的指标可以认为与葡萄酒质量无关；
8. 葡萄酒的各理化指标之间是相互独立的；
9. 酿酒葡萄和葡萄酒的理化指标为线性关系。

三、 基本符号约定：

符号	意义
n	葡萄酒种类的总数，红葡萄酒为 27，白葡萄酒为 28
A_{ij}^k	第 k 组第 j 个评酒员对第 i 种红葡萄酒的总评分， $k=1,2$ ， $j=1,2,\dots,10$ ， $i=1,2,\dots,n$
B_{ij}^k	第 k 组第 j 个评酒员对第 i 种白葡萄酒的总评分， $k=1,2$ ， $j=1,2,\dots,10$ ， $i=1,2,\dots,n$
x_{ij}	第 i 种酿酒葡萄的第 j 个理化指标值， $i=1,2,\dots,n$ ， $j=1,2,\dots,30$
η	芳香物质对葡萄酒质量的影响程度

注：具体符号说明见每小题符号说明。

四、 建立模型及求解:

(一) 问题一:

1. 问题分析:

首先我们以评酒员对各分类指标打分的总分表示该评酒员对这种酒的总评价,各个评酒员对每种酒的总评分见附录表 7-1~表 7-4。由于每个评价组都有 10 个评酒员对葡萄酒进行评价,每个评酒员的评分都有随机误差,因此这 10 个分数是正态总体中的一个样本,且每个评分都相互独立,我们想到用方差显著性分析的方法检验两组评酒员对每种酒的评价是否存在显著性差异,但这种方法只能得到两组评酒员对各种酒的评价是否存在显著性差异,我们还需要得出两组评酒员对所有酒的评价结果有无显著性差异。

我们将一组评酒员对同一种酒的评价看成一个整体,那么某一组对某一种酒的评价为该组所有评酒员对该酒总评分的平均分,我们对这些平均分进行检验,看两组评酒员的评价是否存在显著性差异。但是评酒员对不同酒的评分不服从正态分布,我们无法用已知的检验方法对这些分数检验,我们需要找到一个与各种酒的分数相关的服从正态分布的数列,对这个数列进行统计检验来判定两组评价结果是否有显著性差异。

我们将两组评酒员对每种酒的评分作差,由于每个评酒员对葡萄酒的评价距离标准都有随机误差,那么两组评酒员对同一种酒评分的分差就是随机的,那这组分差数据可以看做是选自正态总体的一个样本。由于这个正态总体标准差未知且样本容量小于 30,那么样本平均数与总体平均数的离差统计量呈 t 分布,因此我们可以用单总体 t-检验来检验正态总体的均值是否为 0,若接受原假设,则表示两组评酒员的评价无显著性差异。

若两组的评价结果不存在显著性差异,那么两组评价结果可以认为在一定显著性水平内是相同的,那么两组的评价结果一样可信;若两组的评价结果存在显著性差异,因为评酒员都是有资质的评酒员,因此他们的评价结果都是在正确标准上下波动,组内的方差越小,该组整体距正确标准就越相近,所以我们认为存在显著性差异时,组内方差小的更可信。

2. 模型建立:

我们以评酒员对各分类指标打分的总分表示该评酒员对这种酒的总评价,不同评酒员对同一种葡萄酒的评分来自同一个正态总体,即:

A_{ij}^1 是来自正态总体 $N(\mu_{1i}, \sigma^2)$ 的容量为 10 的样本, $i=1,2,\dots,27$, $j=1,2,\dots,10$;

A_{ij}^2 是来自正态总体 $N(\mu_{2i}, \sigma^2)$ 的容量为 10 的样本, $i=1,2,\dots,27$, $j=1,2,\dots,10$;

B_{ij}^1 是来自正态总体 $N(\mu_{3i}, \sigma^2)$ 的容量为 10 的样本, $i=1,2,\dots,28$, $j=1,2,\dots,10$;

B_{ij}^2 是来自正态总体 $N(\mu_{4i}, \sigma^2)$ 的容量为 10 的样本, $i=1,2,\dots,28$, $j=1,2,\dots,10$;

要验证两组评酒员对同一种酒的评价是否存在显著性差异,我们引出检验问题(4.1.1)和(4.1.2):

$$H_0: \mu_{1i} = \mu_{2i} \quad \text{vs} \quad H_1: \mu_{1i} \neq \mu_{2i} \quad (4.1.1)$$

$$H_0: \mu_{3i} = \mu_{4i} \quad \text{vs} \quad H_1: \mu_{3i} \neq \mu_{4i} \quad (4.1.2)$$

这些评分之间是有差异的，对某一种红葡萄酒的评分 A_{ij}^1 和 A_{ij}^2 来说，我们分别以组内偏差平方和 S_{ei} 和组间偏差平方和 S_{Ai} 来度量一个评价组内评分和两个评价组间评分数据间差异，分别求出 S_{ei} 和 S_{Ai} ：

$$S_{ei} = \sum_{k=1}^2 \sum_{j=1}^{10} (A_{ij}^k - \overline{A_i^k})^2, \text{ 自由度 } f_e = 18, i=1,2,\dots,27; \quad (4.1.3)$$

$$S_{Ai} = 10 \sum_{k=1}^2 (\overline{A_i^k} - \overline{A_i})^2, \text{ 自由度 } f_A = 1, i=1,2,\dots,27; \quad (4.1.4)$$

我们又引入均方和 MS 来度量评分数据的离散程度：

$$MS_{ei} = \frac{S_{ei}}{f_e}, MS_{Ai} = \frac{S_{Ai}}{f_A}, i=1,2,\dots,27 \quad (4.1.5)$$

用均方和进行比较更为合理，因为均方和排除了自由度不同所产生的干扰，故用

$$F_i = \frac{MS_{ei}}{MS_{Ai}}, i=1,2,\dots,27 \quad (4.1.6)$$

作为检验统计量，而 $F_i \sim F(f_A, f_e)$ ，因此检验(4.1.1)的拒绝域为：

$$W = \{F_i \geq F_{1-\alpha}(1,18)\}$$

计算出各个 F_i ，分别与 $F_{1-\alpha}(1,18)$ 进行比较，若 $F_i \geq F_{1-\alpha}(1,18)$ 则表示对第 i 种红酒两组的评价有显著性差异。

同样地，对某种白葡萄酒的评分分别求出 S_{ei} 和 S_{Ai} ：

$$S_{ei} = \sum_{k=1}^2 \sum_{j=1}^{10} (B_{ij}^k - \overline{B_i^k})^2, \text{ 自由度 } f_e = 18, i=1,2,\dots,28;$$

$$S_{Ai} = 10 \sum_{k=1}^2 (\overline{B_i^k} - \overline{B_i})^2, \text{ 自由度 } f_A = 1, i=1,2,\dots,28;$$

检验(4.1.2)的拒绝域也为：

$$W = \{F_i \geq F_{1-\alpha}(1,18)\}$$

计算出各个 F_i ，分别与 $F_{1-\alpha}(1,18)$ 进行比较，判断两组评酒员对同一种酒的评价是否有显著性差异。

接下来我们检验两组评酒员总体的评价是否存在显著性差异。我们将一组评酒员对同一种酒的评价看成一个整体，那么第 k 组对第 i 种葡萄酒的评价为该组所有评酒员对该酒总评分的平均分，每组总评分的计算方法为：

$$\begin{cases} \overline{A_i^k} = \frac{\sum_{j=1}^{10} A_{ij}^k}{10}, i=1,2,\dots,27 \\ \overline{B_i^k} = \frac{\sum_{j=1}^{10} B_{ij}^k}{10}, i=1,2,\dots,28 \end{cases} \quad (4.1.7)$$

利用式(4.1.7)可以求出 $\overline{A_i^1}$ 、 $\overline{A_i^2}$ 、 $\overline{B_i^1}$ 和 $\overline{B_i^2}$ 。

但是评酒员对每种酒的评分都不是随机的，所以各种酒的分数并不服从正态分布，我们无法用已知的检验方法对这些分数检验。我们需要做的是找到一个与各种酒的分数相关的服从正态分布的数列，对这个数列进行统计检验来判定评价结果是否有显著性差异。于是我们将两组评酒员对每种酒的评分作差：

$$\begin{cases} \overline{\Delta A_i} = \overline{A_i^1} - \overline{A_i^2}, i=1,2,\dots,27 \\ \overline{\Delta B_i} = \overline{B_i^1} - \overline{B_i^2}, i=1,2,\dots,28 \end{cases}$$

每个评酒员对葡萄酒的评价距离标准都有随机误差，那么两组评酒员对同一种酒评分的分差就是随机的，则 $\overline{\Delta A_i}$ 和 $\overline{\Delta B_i}$ 均来自正态总体 $N(\mu, \sigma^2)$ ，由于总体方差未知，因此可以利用 t-检验来检验这组数据总体的均值是否为 0，即两组的评价结果是否有显著性差异。对红葡萄酒的评价结果，我们引出检验问题：

$$H_0: \mu = 0 \quad \text{vs} \quad H_1: \mu \neq 0 \quad (4.1.8)$$

取检验统计量 $t_1 = \frac{\sqrt{n}(\overline{\Delta A} - \mu_0)}{s_{\overline{\Delta A}}}$ ，其中 $\overline{\Delta A}$ 表示 $\overline{\Delta A_i}$ 的平均值， $s_{\overline{\Delta A}}$ 表示 $\overline{\Delta A_i}$ 的标准差

($i=1,2,\dots,27$)， n 表示样本总数，这里 $n=27$ 。在 $\mu = \mu_0$ 时 $t \sim t(n-1)$ ，从而检验问题(4.1.8)的拒绝域为：

$$W = \{|t_1| \geq t_{1-\alpha/2}(26)\}$$

求出 t_1 ，若 $|t_1| \geq t_{1-\alpha/2}(26)$ ，则拒绝原假设，两组评价结果有显著性差异。

同理，对白葡萄酒的评价结果引出检验问题：

$$H_0: \mu = 0 \quad \text{vs} \quad H_1: \mu \neq 0 \quad (4.1.9)$$

取检验统计量 $t_2 = \frac{\sqrt{n}(\overline{\Delta B} - \mu_0)}{s_{\overline{\Delta B}}}$ ，其中 $\overline{\Delta B}$ 表示 $\overline{\Delta B_i}$ 的平均值， $s_{\overline{\Delta B}}$ 表示 $\overline{\Delta B_i}$ 的标准差

($i=1,2,\dots,28$)，这里 $n=28$ ，检验问题(4.1.9)的拒绝域为：

$$W = \{|t_2| \geq t_{1-\alpha/2}(27)\}$$

求出 t_2 ，若 $|t_2| \geq t_{1-\alpha/2}(27)$ ，则拒绝原假设。

若存在显著性差异，再对每组 10 个评酒员对同一种酒的评分计算组内方差，组内方差和小的一组为可信。

3. 模型求解：

我们利用 EXCEL 电子表格处理本题的数据。先用 SUM()函数得到每个评酒员对每种酒的总评分见附录表 7-1~表 7-4。整理数据后分别按照式(4.1.3)~(4.1.6)求出 MS_{ei} , MS_{Ai} 和 F_i 的值，查阅 F 分布 0.95 分位数表知：

$$F_{0.95}(1,18)=4.41$$

用 IF()函数判断各个 F_i 的值是否在拒绝域内，检验结果见见表 4-1-1 和表 4-1-2。表 4-1-1 和表 4-1-2 表明两组评酒员对第 2、11、12、13、16、24 种红葡萄酒和第 5、27 种白葡萄酒的评价具有显著性差异。

表 4-1-1 两组评酒员对每种红葡萄酒的评价结果的显著性

样本号	MS_e	MS_A	F 值	是否显著(Y/N)	样本号	MS_e	MS_A	F 值	是否显著(Y/N)
1	87.39	145.80	1.6684	N	15	63.46	245.00	3.8610	N
2	28.01	198.45	7.0861	Y	16	19.10	125.00	6.5445	Y
3	38.27	168.20	4.3955	N	17	48.59	115.20	2.3709	N
4	74.67	33.80	0.4527	N	18	48.74	151.25	3.1033	N
5	37.83	7.20	0.1903	N	19	51.27	180.00	3.5111	N
6	40.43	174.05	4.3052	N	20	32.56	39.20	1.2041	N
7	83.14	192.20	2.3116	N	21	75.81	120.05	1.5837	N
8	54.56	198.45	3.6372	N	22	37.44	156.80	4.1875	N
9	29.34	54.45	1.8559	N	23	28.63	361.25	12.6189	Y
10	33.29	145.80	4.3798	N	24	42.81	211.25	4.9351	Y
11	54.41	361.25	6.6399	Y	25	54.18	5.00	0.0923	N
12	52.39	1036.80	19.7905	Y	26	36.42	16.20	0.4448	N
13	30.11	168.20	5.5860	Y	27	35.14	11.25	0.3202	N
14	29.58	0.80	0.0270	N					

表 4-1-2 两组评酒员对每种白葡萄酒的评价结果的显著性

样本号	MS_e	MS_A	F 值	是否显著(Y/N)	样本号	MS_e	MS_A	F 值	是否显著(Y/N)
1	59.05	84.05	1.4234	N	15	92.82	180.00	1.9392	N
2	125.07	12.80	0.1023	N	16	130.12	224.45	1.7250	N
3	253.81	470.45	1.8536	N	17	91.32	11.25	0.1232	N
4	43.41	31.25	0.7200	N	18	93.39	64.80	0.6939	N
5	76.36	551.25	7.2190	Y	19	36.22	88.20	2.4350	N
6	92.72	252.05	2.7185	N	20	57.22	7.20	0.1258	N
7	40.67	54.45	1.3388	N	21	118.56	39.20	0.3306	N
8	107.36	4.05	0.0377	N	22	96.13	352.80	3.6699	N
9	99.52	281.25	2.8262	N	23	27.63	11.25	0.4072	N
10	141.54	151.25	1.0686	N	24	74.83	39.20	0.5238	N
11	132.47	4.05	0.0306	N	25	70.19	28.80	0.4103	N
12	127.92	414.05	3.2369	N	26	87.90	245.00	2.7873	N
13	108.77	320.00	2.9421	N	27	89.98	744.20	8.2709	Y
14	65.05	130.05	1.9992	N	28	52.92	14.45	0.2731	N

下面进行两组评酒员整体评价的显著性检验：

用 AVERAGE()函数计算出每组 10 个评酒员对红葡萄酒和白葡萄酒的评价的平均得分并分别作差，如表 4-1-3 和表 4-1-4 所示。

表 4-1-3 每组 10 个评酒员对红葡萄酒的评价的平均得分以及两组的分差

样本编号	$\overline{A_i^1}$	$\overline{A_i^2}$	ΔA_i	样本编号	$\overline{A_i^1}$	$\overline{A_i^2}$	ΔA_i
样本 1	62.7	68.1	-5.4	样本 15	58.7	65.7	-7.0
样本 2	80.3	74	6.3	样本 16	74.9	69.9	5.0
样本 3	80.4	74.6	5.8	样本 17	79.3	74.5	4.8
样本 4	68.6	71.2	-2.6	样本 18	59.9	65.4	-5.5
样本 5	73.3	72.1	1.2	样本 19	78.6	72.6	6.0
样本 6	72.2	66.3	5.9	样本 20	78.6	75.8	2.8
样本 7	71.5	65.3	6.2	样本 21	77.1	72.2	4.9
样本 8	72.3	66.0	6.3	样本 22	77.2	71.6	5.6
样本 9	81.5	78.2	3.3	样本 23	85.6	77.1	8.5
样本 10	74.2	68.8	5.4	样本 24	78.0	71.5	6.5
样本 11	70.1	61.6	8.5	样本 25	69.2	68.2	1.0
样本 12	53.9	68.3	-14.4	样本 26	73.8	72.0	1.8
样本 13	74.6	68.8	5.8	样本 27	73.0	71.5	1.5
样本 14	73.0	72.6	0.4				

表 4-1-4 每组 10 个评酒员对白葡萄酒的评价的平均得分以及两组的分差

样本编号	\overline{B}_i^1	\overline{B}_i^2	ΔB_i	样本编号	\overline{B}_i^1	\overline{B}_i^2	ΔB_i
样品 1	82	77.9	4.1	样品 15	72.4	78.4	-6.0
样品 2	74.2	75.8	-1.6	样品 16	74.0	67.3	6.7
样品 3	85.3	75.6	9.7	样品 17	78.8	80.3	-1.5
样品 4	79.4	76.9	2.5	样品 18	73.1	76.7	-3.6
样品 5	71.0	81.5	-10.5	样品 19	72.2	76.4	-4.2
样品 6	68.4	75.5	-7.1	样品 20	77.8	76.6	1.2
样品 7	77.5	74.2	3.3	样品 21	76.4	79.2	-2.8
样品 8	71.4	72.3	-0.9	样品 22	71.0	79.4	-8.4
样品 9	72.9	80.4	-7.5	样品 23	75.9	77.4	-1.5
样品 10	74.3	79.8	-5.5	样品 24	73.3	76.1	-2.8
样品 11	72.3	71.4	0.9	样品 25	77.1	79.5	-2.4
样品 12	63.3	72.4	-9.1	样品 26	81.3	74.3	7.0
样品 13	65.9	73.9	-8.0	样品 27	64.8	77.0	-12.2
样品 14	72.0	77.1	-5.1	样品 28	81.3	79.6	1.7

然后分别用 AVERAGE()和 VAR()函数分别求出 $\overline{\Delta A_i}$ 和 $\overline{\Delta B_i}$ 的平均值和标准差为：

$$\overline{\Delta A} = 2.540741, s_{\overline{\Delta A_i}} = \sqrt{\text{var}(\overline{\Delta A_i})} = 5.371883$$

$$\overline{\Delta B} = -2.27143, s_{\overline{\Delta B_i}} = \sqrt{\text{var}(\overline{\Delta B_i})} = 5.503861$$

继续求出检验统计量 t_1 和 t_2 ：

$$t_1 = \frac{\sqrt{n}(\overline{\Delta A} - \mu_0)}{s_{\overline{\Delta A_i}}} = 2.457626$$

$$t_2 = \frac{\sqrt{n}(\overline{\Delta B} - \mu_0)}{s_{\overline{\Delta B_i}}} = -2.18379$$

查阅 t 分布分位数表可知 $t_{0.975}(26) = 2.0555$ ， $t_{0.975}(27) = 2.0518$ ，在检验(4.1.8)和(4.1.9)中，

$$|t_1| \geq t_{0.975}(26)$$

$$|t_2| \geq t_{0.975}(27)$$

因此均拒绝原假设，我们得到结论：两组评酒员的对红葡萄酒、白葡萄酒的评价结果都有显著性差异。

我们又对每组 10 个人对同一种酒的评分计算组内方差，结果如表 4-1-5 所示。

表 4-1-5 两组评酒员对红葡萄酒和白葡萄酒评分的组内方差

样本号	$Var(A_i^1)$	$Var(A_i^2)$	$Var(B_i^1)$	$Var(B_i^2)$	样本号	$Var(A_i^1)$	$Var(A_i^2)$	$Var(B_i^1)$	$Var(B_i^2)$
1	92.90	81.88	92.22	25.88	15	85.57	41.34	131.60	54.04
2	39.79	16.22	201.07	49.07	16	18.10	20.10	178.00	82.23
3	45.82	30.71	365.12	142.49	17	88.01	9.17	144.18	38.46
4	108.04	41.29	44.71	42.10	18	47.21	50.27	156.54	30.23
5	62.01	13.66	126.44	26.28	19	47.38	55.16	46.40	26.04
6	59.73	21.12	162.71	22.72	20	26.04	39.07	64.40	50.04
7	103.61	62.68	39.17	42.18	21	116.10	35.51	172.71	64.40
8	44.01	65.11	183.60	31.12	22	50.62	24.27	138.67	53.60
9	32.94	25.73	92.67	106.27	23	32.49	24.77	43.66	11.60
10	30.40	36.18	212.68	70.40	24	74.89	10.72	111.12	38.54
11	70.77	38.04	177.12	87.82	25	64.62	43.73	33.88	106.50
12	79.66	25.12	115.79	140.04	26	31.29	41.56	72.90	102.90
13	44.93	15.29	170.77	46.77	27	49.78	20.50	144.40	35.56
14	36.00	23.16	114.22	15.88	28			80.46	25.38

再用 SUM() 函数求出组内方差之和:

$$\sum_{i=1}^{27} Var(A_i^1) = 1582.722, \quad \sum_{i=1}^{27} Var(A_i^2) = 912.3444$$

$$\sum_{i=1}^{28} Var(B_i^1) = 3617.3, \quad \sum_{i=1}^{28} Var(B_i^2) = 1568.544$$

很明显, 第二组的方差和比第一组的小得多, 因此第二组的结果更可信。

4. 误差分析:

本题中我们是在显著性水平为 0.05 的条件下进行检验的, 我们又分别取显著性水平为 0.02, 0.01 对检验问题(4.1.8)和(4.1.9)进行检验, 则检验问题(4.1.8)和(4.1.9)的拒绝域变为:

$$\text{显著性水平 } 0.02: W = \{|t_1| \geq t_{0.99}(26)\} \quad W = \{|t_2| \geq t_{0.99}(27)\}$$

$$\text{显著性水平 } 0.01: W = \{|t_1| \geq t_{0.995}(26)\} \quad W = \{|t_2| \geq t_{0.995}(27)\}$$

查阅 t 分布分位数表知:

$$t_{0.99}(26) = 2.4786, \quad t_{0.99}(27) = 2.4727$$

$$t_{0.995}(26) = 2.7787, \quad t_{0.995}(27) = 2.7707$$

我们发现当显著性水平为 0.02 和 0.01 时均接受原假设, 即两组评价结果无显著性差异。因此, 当显著性水平取不同的值时, 检验结果可能不同。但检验时我们不能忽略犯两种错误的概率, 当显著性水平很小时, 受伪概率就增加了, 即当显著性水平为 0.02, 0.01 时, 此时的受伪概率是很大的, 这时结果显示接受原假设不排除犯了第二类错误的可能性。

(二) 问题二:

1. 符号说明:

p : 主要成分的个数;

L_i : 第 i 种葡萄的理化指标分数;

M_i : 第 i 种葡萄的总分数

2. 问题分析:

题目要求根据葡萄的理化指标和葡萄酒的质量对酿酒葡萄进行分级, 那么我们首先要确定一个分级的标准。我们想到给每种酿酒葡萄打一个“分”, 这个“分”要包含葡萄的理化指标和对应葡萄酒的质量这两个信息, 由于酿酒葡萄的好坏直接影响酿出的酒的质量, 我们认为葡萄酒的质量所占的比重更大一些, 因此引入权值 λ 表示葡萄酒质量所占的权重, $\lambda \geq 0.5$ 。葡萄酒的质量可以用题目一中第二组评酒员的平均评分直接量化, 因此我们要做的是找出量化葡萄理化指标的标准。

我们先用**主成分分析法**对这些指标进行筛选, 选出主要理化指标, 根据主要指标的信息贡献率给出每种葡萄的综合得分。

然而用这种方法得出的综合得分不是百分制的, 我们引出**差分比例法**百分化葡萄的理化指标分数, 所谓差分比例就是某个葡萄的得分与最低分的差比上最高分与最低分的差, 即:

$$\frac{L_i - \min L}{\max L - \min L}$$

得到一个 0~1 之间的数, 如果直接用这个数乘以 100, 那么葡萄理化指标的分数最高为 100, 最低为 0, 不符合常理, 因此我们将理化指标的分数分为两个部分: 基本分 50 和用差分比例乘以 50 得到的分数。

3. 模型建立:

首先我们利用葡萄各个理化指标的含量值进行主成分分析, 引入各个指标之间的相关系数矩阵 R :

$$R = (r_{jk})_{30 \times 30}$$

$$r_{jk} = \frac{\sum_{i=1}^{30} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{30} (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^{30} (x_{ik} - \bar{x}_k)^2}} \quad (4.2.1)$$

其中 r_{jk} 表示第 j 个指标和第 k 个指标之间的相关系数, $j, k=1, 2, \dots, 30$ 。

接下来计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{30} \geq 0$ 和对应的特征向量

$\mu_1, \mu_2, \dots, \mu_{30}$, 其中:

$$u_j = (u_{1j}, u_{2j}, \dots, u_{nj})^T,$$

由这些特征向量组成 m 个新的指标向量：

$$\begin{cases} w_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \cdots + u_{n1}\tilde{x}_n \\ w_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \cdots + u_{n2}\tilde{x}_n \\ \vdots \\ w_{30} = u_{1,30}\tilde{x}_1 + u_{2,30}\tilde{x}_2 + \cdots + u_{n,30}\tilde{x}_n \end{cases} \quad (4.2.2)$$

式(4.2.2)中 w_i 表示第 i 个主要成分, $i=1,2,\dots,30$, \tilde{x}_j 表示标准化指标变量, $\tilde{x}_j = \frac{x_j - \overline{x_j}}{s_j}$,

$j=1,2,\dots,n$ 。

然后我们选择前 $p(p \leq 30)$ 个主成分, 求出特征值 λ_j 的信息贡献率:

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{30} \lambda_k};$$

以及主成分 w_1, w_2, \dots, w_p 的累积贡献率:

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^{30} \lambda_k}$$

当累计贡献率和 α_p 接近 1 时则可以认为前 p 个指标变量 w_1, w_2, \dots, w_p 是主要成分, 代替原来的 30 个理化指标。分析出了主要成分我们就可以对葡萄的进行综合评价, 得到原始理化指标分数 L_{i0} , $i=1,2,\dots,n$:

$$Z_0 = \sum_{j=1}^p b_j w_j = [L_{1,0} \quad L_{2,0} \quad \cdots \quad L_{n,0}]^T$$

我们对它进行一些处理使它成为一个满分为 100 的分数:

$$L_i = \frac{L_i - \min L}{\max L - \min L} * 50 + 50, \quad i=1,2,\dots,n \quad (4.2.3)$$

以 L_i 作为第 i 种葡萄理化指标的分数进行计算。

由第一题得到第 i 种酿酒葡萄对应葡萄酒的分数 $\overline{A_i^2}$ 和 $\overline{B_i^2}$, 由式(4.2.3)可得到第 i 种酿酒葡萄理化指标的分数, 结合权值我们得到葡萄的总分:

$$\text{红酿酒葡萄: } M_i = \lambda \overline{A_i^2} + (1-\lambda)L_i$$

$$\text{白酿酒葡萄: } M_i = \lambda \overline{B_i^2} + (1-\lambda)L_i \quad (4.2.4)$$

4. 模型求解:

我们先用 MATLAB 编程进行主成分分析, 发现 $p=14$ 时, 累计贡献度达到 0.95 以上, 相当接近 1, 我们就可以认为前 14 个理化指标为主要成分, 得到主成分分析表, 如表 4-2-1 所示:

表 4-2-1 主成分分析表

序号	特征根	贡献率	累计贡献率
1	6.97	23.23	23.23
2	4.94	16.47	39.70
3	3.73	12.45	52.15
4	2.84	9.47	61.61
5	2.00	6.67	68.28
6	1.74	5.81	74.09
7	1.42	4.73	78.81
8	1.27	4.23	83.04
9	0.96	3.20	86.25
10	0.74	2.46	88.71
11	0.69	2.30	91.01
12	0.51	1.71	92.73
13	0.49	1.64	94.37
14	0.37	1.24	95.61

然后分别以这 14 个主成分的贡献率为权重, 计算出各种葡萄的综合评价分数, 利用式(4.2.3)将之百分化, 得到理化指标的分数, 进而求出总分数。

我们取权值 $\lambda = 0.7$, 得到每种红酿酒葡萄和白酿酒葡萄的总分数如表 7-2-2 所示:

表 7-2-2 每种红酿酒葡萄的分数

葡萄编号	1	2	3	4	5	6	7
总分	82.22	82.27	77.99	78.97	74.42	69.35	74.24
葡萄编号	8	9	10	11	12	13	14
总分	72.32	69.28	73.72	76.76	70.15	68.12	72.29
葡萄编号	15	16	17	18	19	20	21
总分	69.57	67.12	71.66	66.78	70.95	73.00	66.23
葡萄编号	22	23	24	25	26	27	
总分	69.56	72.49	68.92	68.29	64.01	62.74	

根据葡萄的总分得到葡萄的分类如下:

红葡萄的分类:

≤ 65 分: 10、25;

65~70 分: 4、6、7、8、11、13、15、18、26、27;

70~75 分: 1、5、12、14、16、17、19、20、22、24;

75~80 分: 2、21、23;

≥ 80 分: 3、9。

同理可得到白葡萄的分类:

≤ 65 分: 16;
 $65\sim 70$ 分: 8、11;
 $71\sim 75$ 分: 1、2、4、12、13、14、17、19、22、26;
 $76\sim 80$ 分: 3、6、7、9、10、15、18、20、21、23、24、25、28;
 ≥ 80 分: 5、27;

(三) 问题三:

模型一:

1. 符号说明:

x_i : 酿酒葡萄的第 i 个理化指标, $i=1,2,\dots,30$;

y_j : 葡萄酒的第 j 个理化指标, $j=1,2,\dots,9$;

2. 问题分析:

葡萄酒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。酿酒葡萄在酿制过程中会发生一些化学变化,其中的某些化学成分会发生改变,也有的成分会保留在葡萄酒中,我们要研究的是各种成分的变化。我们无法知道酿制过程中的化学变化,但可以通过题目中的附件 2 可以发现,酿酒葡萄与葡萄酒的理化指标之间存在着某种关系。为了更好的找出之间的关系,我们先计算出两者之间的相关性。然后找出酿酒葡萄与葡萄酒理化指标相关性最高的前 5 个的理化指标作为自变量,对应的葡萄酒理化指标作为因变量,运用多元线性回归方法找出两者之间的联系。由于白葡萄的求解方法和红葡萄类似,故我们只重点分析红葡萄。

3. 模型建立:

首先我们先对数据进行标准化处理:

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

酿酒葡萄第 i 种理化指标与葡萄酒第 j 种理化指标之间的相关程度 r_{ij} 为:

$$r_{ij} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}}, \quad i, j=1,2,\dots,30。$$

我们引入 0-1 变量 C_{ij} , 若 r_{ij} 是第 i 个酿酒葡萄的理化指标里相关性最高的前 5 个, 则

$C_{ij}=1$, 否则 $C_{ij}=0$ 。

我们可以建立如下总体回归方程的矩阵表达式:

$$E(Y) = X\beta + \beta_0 \quad (4.3.1)$$

具体的形式为:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_9 \end{bmatrix} = \begin{bmatrix} 1 & C_{1,1}X_{11} & C_{2,1}X_{21} & \dots & C_{30,1}X_{30,1} \\ 1 & C_{1,2}X_{12} & C_{2,2}X_{22} & \dots & C_{30,2}X_{30,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & C_{1,9}X_{19} & C_{2,9}X_{29} & \dots & C_{30,9}X_{30,9} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{30} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{30} \end{bmatrix}$$

其中 u_i 表示残差, $i=1,2,\dots,30$ 。

4. 模型求解:

我们用 MATLAB 的 regress() 进行线性拟合, 得到红葡萄酒与酿酒葡萄关系式为:

$$\begin{cases} y_1 = -3.3887 + 0.1718x_4 + 0.8474x_6 + 0.0116x_9 - 0.2263x_{11} + 0.4730x_{12} \\ y_2 = 0.8141 + 0.0045x_4 + 6.6868x_{10} + 0.2622x_{11} + 0.0880x_{12} - 0.1691x_{13} \\ y_3 = 1.3337 + 0.0083x_4 + 3.3559x_{10} + 0.0982x_{11} + 0.0346x_{12} + 0.1191x_{13} \\ y_4 = -0.2534 + 0.0063x_4 - 1.7316x_{10} + 0.2499x_{11} + 0.0051x_{12} + 0.1623x_{13} \\ y_5 = 0.7380 + 0.0006x_1 - 0.0467x_{10} - 0.1773x_{11} - 0.0463x_{12} + 0.5697x_{13} \\ y_6 = -0.0248 + 0.0001x_4 + 0.0639x_{10} + 0.0110x_{11} + 0.0012x_{12} + 0.0043x_{13} \\ y_7 = 79.1052 - 0.1274x_4 - 41.1418x_{10} - 0.9105x_{11} - 0.2713x_{12} + 0.8114x_{13} \\ y_8 = 81.8974 - 1.4623x_6 - 0.0718x_{14} - 0.5865x_{21} - 2.6487x_{29} - 0.4124x_{30} \\ y_9 = 2.8713 - 0.4452x_3 + 1.0992x_5 + 0.1246x_{16} + 0.1649x_{17} - 2.0594x_{22} \end{cases} \quad (4.3.2)$$

同理我们也可以得到白葡萄酒和酿酒葡萄的理化指标之间的关系:

$$\begin{cases} y_1 = 1.0552 + 0.0002x_1 - 0.1644x_{11} + 0.1490x_{12} + 0.2407x_{13} + 0.0244x_{15} \\ y_2 = 0.1939 + 0.0002x_1 + 1.0342x_{10} - 0.0032x_{11} + 0.0645x_{12} + 0.0930x_{13} \\ y_3 = -1.2330 + 0.0052x_2 + 0.1614x_{11} + 0.0653x_{13} + 0.0519x_{15} - 0.4173x_{25} \\ y_4 = 2.3314 - 0.0013x_2 - 0.0394x_5 - 0.0216x_6 - 0.0035x_{16} - 0.0072x_{28} \\ y_5 = 0.0294 - 0.0005x_8 + 0.0465x_{10} - 0.0002x_{11} + 0.0036x_{12} + 0.0029x_{13} \\ y_6 = 102.2063 - 0.0013x_{16} + 0.0030x_{17} - 0.0005x_{18} - 0.1068x_{22} + 0.0247x_{26} \\ y_7 = -2.1467 + 0.0009x_2 + 0.0101x_{13} + 0.0003x_{23} - 0.0813x_{25} + 0.0176x_{26} \\ y_8 = -2.1467 + 0.0009x_{16} + 0.0101x_{18} + 0.0003x_{22} - 0.0813x_{23} + 0.0176x_{26} \end{cases} \quad (4.3.3)$$

5. 模型检验:

我们通过统计检验来检验模型的正确度:

①拟合优度检验: 可决系数 R^2 的取值范围是 0~1。 R^2 越接近 0, 说明模型拟合度越低; 越接近 1, 说明模型的拟合度越高。通过检验, 方程的自变量与因变量的相关程度高, 模型对样本的拟合很好。

②变量的显著性检验: F 检验的伴随概率为 0.000, 反应变量间呈高度线性, 方程回归效果显著。

③参数的置信区间估计: 在给定的 $\alpha=0.05$, 得到以上的置信区间。经检验, 随 α 减小, 得到的置信区间要小很多, 说明估计结果精度随之增加, 模型越正确。

模型二:

1. 符号说明:

x_{ji} : 第 j 种酿酒葡萄的第 i 个理化指标, $i=1,2,\dots,30$;

P : 表示酿酒葡萄的理化指标可以分类个数 $p=1,2,3,\dots$

y_j : 葡萄酒的第 j 个理化指标, $j=1,2,\dots,9$;

c_{ik} : 酿酒葡萄的第 i 种理化指标是否属于第 k 类, 若属于 $c_{ik}=1$ 否则 $c_{ik}=0$;

2. 问题分析:

通过模型一的求解, 我们发现葡萄酒的理化指标的变化不只由单一的酿酒葡萄理化指标所决定, 而是更趋向于由几个酿酒葡萄理化指标所组成的集合决定的。因此, 我们认为, 可以先对酿酒葡萄的各个指标先进行分析, 然后分类, 将分类后各个类里包含理化指标求取平均值, 作为对葡萄酒理化指标影响的因子, 也就是因变量。我们仍认为每个类与葡萄酒理化指标之间呈线性关系, 运用多元线性回归拟合两者之间的关系, 最终就可以求得酿酒葡萄与葡萄酒的理化指标之间存在的关系。由于白葡萄酒的求解方法和红葡萄酒类似, 故我们仍只分析红葡萄酒。

3. 模型建立:

首先我们先对数据进行标准化处理。

我们在对葡萄酒理化指标进行聚类分析时, 首先确定变量的相似性度量, 我们用相关系数作为它们的相似性度量

$$r_{jk} = \frac{\sum_{j=1}^n (x_{ij} - \bar{a}_i)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \sum_{j=1}^n (x_{ik} - \bar{x}_k)^2}}, \quad i, j=1,2,\dots,30$$

在完成酿酒葡萄理化指标具体分类后, 我们将根据以下公式得到均值:

$$x_k = \frac{\sum_{i=1}^{m_k} x_i c_{ik}}{m_k}$$

根据 y_j 与 x_k 之间的线性关系, 我们可以建立如下总体回归方程的矩阵表达式:

$$E(Y) = X\beta + \beta_0$$

4. 模型求解:

我们运用 MATLAB 的 score() 函数进行数据标准化, 再用 SPSS 软件进行分层聚类分析, 得出酿酒红葡萄酒的理化指标可以分类成 5 类, 同理也可得到酿酒白葡萄酒的理化指标可以分类成 5 类, 其结果见表 4-3-1。

表 4-3-1 两种酿酒葡萄的理化指标种类

分类	红酿酒葡萄的理化指标种类	白酿酒葡萄的理化指标种类
1	3	2、9、11、12、13、15
2	23、24、27、28	1、6、10、16、17、18、20、22、28、30
3	5、7、14、29、30	3、4、5、8、25
4	1、16、17、18、20、22	7、14、19、21、29
5	2、4、6、8、9、10、11、12、13、15、19、21、25、26	23、24、26、27

红白两种酿酒葡萄的树状图如图 4-3-1 所示。

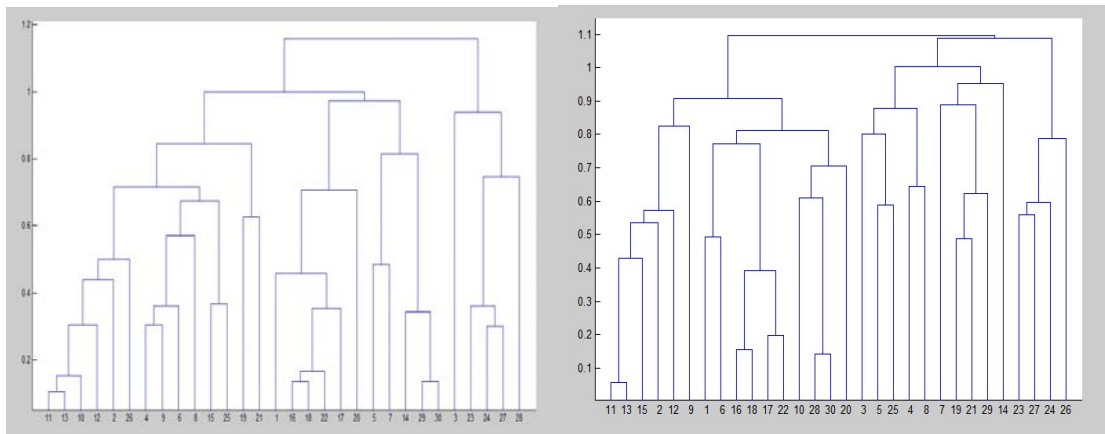


图 4-3-1 树状图

做完以上工作，我们沿用模型一中的 matlab 编程，经计算得出如下结果，即红酿酒葡萄的理化指标和红葡萄酒的理化指标的关系如下：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_9 \end{bmatrix} = \begin{bmatrix} -48.57 & 2.60 & 0.089 & -1.055 & 0.020 & 0.62 \\ -2.6643 & -0.9127 & 0.0096 & 0.1834 & 0.0065 & 0.0687 \\ -2.1597 & -0.8728 & 0.0104 & 0.2467 & 0.0042 & 0.0614 \\ -2.2648 & -1.9982 & 0.0043 & 0.3145 & 0.0031 & 0.0734 \\ -1.5895 & -1.3564 & 0.0160 & 0.1986 & 0.0050 & 0.0208 \\ -0.1895 & -0.0646 & 0.0005 & 0.0131 & 0.0002 & 0.0030 \\ 111.4715 & -0.5054 & -0.1112 & 1.7959 & -0.0323 & -0.5243 \\ 76.0748 & 2.2281 & -0.0121 & -3.8172 & -0.0034 & -0.1656 \\ 20.0605 & 0.2791 & 0.0023 & 0.0462 & 0.0102 & -0.0444 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_5 \end{bmatrix} \quad (4.3.4)$$

同理我们也可以得到白葡萄酒和酿酒葡萄的理化指标之间的关系：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_9 \end{bmatrix} = \begin{bmatrix} 0.6069 & 0.0027 & 0.0355 & -0.0420 & 0.0032 & 0.0014 \\ 0.7338 & 0.0021 & 0.0376 & -0.0700 & 0.0023 & 0.0012 \\ -1.4964 & 0.0045 & 0.1286 & -0.1389 & 0.0044 & 0.0165 \\ 0.8269 & -0.0011 & -0.0194 & 0.0107 & -0.0007 & -0.0007 \\ 0.0703 & -0.0000 & -0.0007 & -0.0029 & 0.0000 & 0.0001 \\ 102.2509 & 0.0022 & -0.0225 & -0.0168 & -0.0014 & -0.0001 \\ -0.9340 & 0.0019 & -0.0001 & -0.0088 & 0.0003 & 0.0008 \\ 2.4981 & -0.0102 & 0.0734 & 0.0488 & 0.0033 & 0.0027 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_5 \end{bmatrix} \quad (4.3.5)$$

5. 模型检验:

我们将沿用方案一的检验方法即通过统计检验来检验模型的正确度。

①拟合优度检验②变量的显著性检验③参数的置信区间估计,最终得出结论方程的自变量与因变量的相关程度高,模型对样本的拟合很好。F 检验的伴随概率为 0.000,反应变量间呈高度线性,方程回归效果显著。

(四) 问题四:

1. 符号说明:

R: 五类指标分数和三个评价方面分数的相关系数矩阵;

2. 问题分析:

本题中我们要探究酿酒葡萄、葡萄酒各种指标对葡萄酒质量的影响,从而确定理化指标对葡萄酒质量的影响大小。从第一题中给的数据中我们可以看出,葡萄酒的质量评价分成外观、香气和口感三个部分,每个部分影响程度的大小可以从每个部分的分值中看出来。理化指标对葡萄酒质量的影响用数值表示更加直观,因此我们将各种理化指标的含量化成分数,葡萄酒的质量用第一题中第二组评酒员的评分量化。

我们通过查阅各种资料^[7]得知,影响葡萄酒外观、香气、口感的主要指标有糖类、酸类、酚醇类、色素类和芳香物质的含量,其他的指标对酒的影响不大,可以认为与酒的质量无关。我们求出这五类指标对葡萄酒质量影响程度的大小,其中这五类指标里前四类都是葡萄酒的理化指标,要确定理化指标能否评价酒的质量只需确定芳香物质是否会影响酒的质量。

由于葡萄酒质量相关的五项指标中并没有全部给出,没有给出的葡萄酒的理化指标我们采用对应酿酒葡萄的指标。对这五项指标用差分比例法打分,使指标量化,然后分别对这五个指标和葡萄酒的质量分数作相关性检验,得到相关系数,对相关系数进行处理,得到相关度百分比。若芳香物质与葡萄酒质量的相关性超过我们规定的阈值,则认为芳香物质对葡萄酒质量影响较大,否则我们可以认为酿酒葡萄和葡萄酒的理化指标可以决定葡萄酒质量。

3. 模型建立:

首先将 27 种红葡萄酒的四类理化指标整合,糖类指标为总糖和还原糖的平均值,取酿酒葡萄的指标;酸类指标为酒石酸、柠檬酸和苹果酸的平均值,取酿酒葡萄的指标;酚醇类指标包括单宁、总酚、酒石黄酮和白藜芦醇,取葡萄酒指标,其中单宁为主要酚醇,含量占分值的比重较大,酚醇类指标对各种指标进行加权平均,单

宁占 50%,; 色素类指标为花色苷, 红葡萄酒的参数取葡萄酒指标, 白葡萄酒的参数取酿酒葡萄指标。芳香物质我们取其分子量总和作为其参考指标。

接下来我们对着五种指标用差分比例法进行评分, 得到红白葡萄五种指标的分数如表 4-4-1 所示。

$$R_j' = \frac{R_{ij}}{\sum_{j=1}^5 R_{ij}} \text{ 其中 } R_j' \text{ 表示 } R_j \text{ 的归一化矩阵}$$

$$\eta = \frac{15R_{15}' + 30R_{25}' + 44R_{35}'}{89} \quad (4.4.1)$$

表 4-4-1 红葡萄五种指标的分数

编号	糖类	酸类	酚醇类	色素	芳香物质	编号	糖类	酸类	酚醇类	色素	芳香物质
1	100.00	81.19	82.51	100.00	85.60	15	70.46	68.26	51.40	55.76	78.99
2	80.70	78.57	87.25	76.28	76.93	16	90.10	69.44	54.41	58.30	74.72
3	70.95	100.00	87.21	70.11	60.39	17	70.31	89.78	70.26	61.57	85.49
4	63.58	77.09	63.06	58.92	77.16	18	64.05	91.47	54.76	53.12	56.92
5	78.53	72.08	61.75	63.95	82.71	19	64.77	76.78	62.92	59.71	74.79
6	50.49	86.19	65.92	55.47	84.55	20	50.00	75.53	59.15	53.25	59.77
7	76.78	96.39	52.43	54.11	78.10	21	96.29	85.45	78.34	65.69	69.26
8	99.04	70.01	84.46	97.13	60.57	22	79.37	82.24	66.89	62.43	96.69
9	89.41	68.39	91.20	69.54	72.08	23	63.93	75.47	92.84	70.90	81.51
10	50.35	54.97	57.92	56.59	65.66	24	78.41	81.30	61.59	63.42	79.72
11	93.33	81.55	52.95	50.00	51.62	25	51.86	50.00	57.22	57.63	74.03
12	72.26	95.50	60.20	53.75	74.32	26	63.06	64.39	51.27	57.26	72.31
13	54.09	73.20	60.66	59.78	79.58	27	61.64	73.15	60.14	56.58	100.00
14	69.18	81.41	59.50	62.46	50.00						

4. 模型求解:

对于红酒: 利用MATLAB的corrcoef()函数对五类分数和三大评价方面进行计算, 求出R 矩阵。按照上两式进行模型求解, 得

$$\eta = 0.1317$$

对于白酒:

$$\eta = 0.2955$$

5. 模型检验:

由于葡萄酒的质量分成外观、香气和口感三个部分, 而影响葡萄酒外观、香气、口感的主要指标又有糖类、酸类、酚醇类、色素类和芳香物质的含量五个指标, 也就是说: 糖类、酸类、酚醇类、色素类和芳香物质的含量影响葡萄酒的外观、香气和口感, 葡萄酒的外观、香气和口感影响葡萄酒的质量, 那么对葡萄酒质量的评价就可以分为三个层次, 如图 4-4-1 所示:

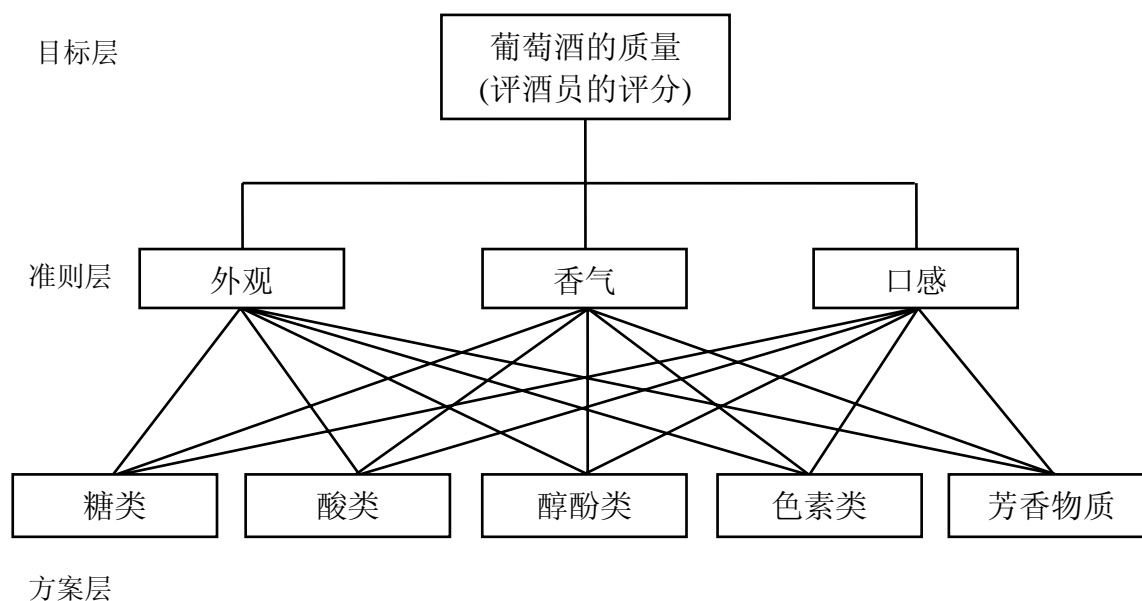


图 4-4-1 葡萄酒质量指标的层次结构

则我们可以用层次分析法分析五类指标对葡萄酒质量的影响效果，以检验我们所得到的结论是否正确。

在评酒员的打分体系中，外观，香气，口感分别占 15 分，30 分，44 分。利用 Satty 比较尺度我们为其分别赋值为 1,3,5，由此可以建立准则层成对比较阵：

$$A = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{5} \\ 3 & 1 & \frac{3}{5} \\ 5 & \frac{5}{3} & 1 \end{bmatrix}$$

通过查阅资料，我们建立方案层的判断矩阵：

$$B_1 = \begin{bmatrix} 1.00 & 0.60 & 0.43 & 0.33 & 3.00 \\ 1.67 & 1.00 & 0.71 & 0.56 & 5.00 \\ 2.33 & 1.40 & 1.00 & 0.78 & 7.00 \\ 3.00 & 1.80 & 1.29 & 1.00 & 9.00 \\ 0.33 & 0.20 & 0.14 & 0.11 & 1.00 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 1.00 & 1.67 & 0.71 & 5.00 & 0.56 \\ 0.60 & 1.00 & 0.43 & 3.00 & 0.33 \\ 1.40 & 2.33 & 1.00 & 7.00 & 0.78 \\ 0.20 & 0.33 & 0.14 & 1.00 & 0.11 \\ 1.80 & 3.00 & 1.28 & 9.00 & 1.00 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 1.00 & 1.40 & 0.78 & 7.00 & 2.33 \\ 0.71 & 1.00 & 0.56 & 5.00 & 1.67 \\ 1.29 & 1.80 & 1.00 & 9.00 & 3.00 \\ 0.14 & 0.20 & 0.11 & 1.00 & 0.33 \\ 0.43 & 0.60 & 0.33 & 3.00 & 1.00 \end{bmatrix}$$

我们再用和法求取权矩阵，得到准则层和方案层的权向量矩阵为：

$$\omega_0 = [0.1111 \quad 0.3333 \quad 0.5556]$$

$$\omega_1 = \begin{bmatrix} 0.12 & 0.20 & 0.28 \\ 0.20 & 0.12 & 0.20 \\ 0.28 & 0.28 & 0.36 \\ 0.36 & 0.04 & 0.04 \\ 0.04 & 0.36 & 0.12 \end{bmatrix}$$

然后求得层次总排序如表 4-4-3 所示：

表 4-4-2 层次总排序

准则		外观	香气	口感	总排序权值
准则层权值		0.111	0.33	0.556	
方案层单排序 权值	糖类	0.12	0.20	0.28	0.2356
	酸类	0.20	0.12	0.20	0.1733
	酚醇类	0.28	0.28	0.36	0.3244
	色素类	0.36	0.04	0.04	0.0756
	芳香物质类	0.04	0.36	0.12	0.1911

根据表 4-4-3 得到准则层矩阵一致性比例 CR_0 和方案层矩阵一致性比例 CR_1 ：

$$CR_0 = -1.1485 \times 10^{-5}$$

$$CR_1 = 10^{-5} [0.8475 \quad 0.8475 \quad 0.8475]$$

这两个数均十分接近 0，组合一致性通过，这五项指标可以作为最终决策的依据，芳香物质对葡萄酒质量的影响较大，不能仅通过理化指标来评价葡萄酒的质量。

五、 模型评价：

问题一我们分别检验了两组评酒员对同一种酒的评价的显著性和总体评价的显著性，用方差显著性分析和 t 检验对多个方面进行显著性差异分析，并在误差分析中加入对置信度的讨论。对于哪一组更可靠，我们采用方差必为主要比较元素。模型考虑较为严谨，但由于未获得相应显著性水平的标准，不免些许偏差。

问题二，我们对葡萄的理化指标进行量化，并对葡萄酒质量和葡萄理化指标采用加权分析，其中对葡萄理化指标采用主成分分析，对总分评级，较为全面。

问题三，我们对葡萄理化指标和葡萄酒理化指标进行相关性分析，取相关性较大的数据做多元线性拟合；对葡萄理化指标进行聚类分析，再与葡萄酒理化指标进

行数据拟合。两种模型都在不同方向上对葡萄理化指标进行简化，使数据处理相对简单。

问题四，我们将葡萄酒和葡萄的理化指标分为四大类，与芳香物质作为五大指标与相关分数进行相关分析处理，从而确定芳香物质对葡萄酒质量的作用。在模型检验中，我们采用层次分析法，得到的结果与模型结果相吻合。

六、 参考文献：

- [1] 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。
- [2] 茆诗松、程依明、濮晓龙，概率论与数理统计[M]，北京：高等教育出版社，2004。
- [3] 杨凌，关于葡萄品质的评价指标，中外葡萄与葡萄酒[J]，712100：P54~57，1991。
- [4] 宋廷山、尉雪波、吴风庆，应用统计学[M]，四川：西南财经大学出版社，2006。
- [5] 王金甲、尹涛、李静等，基于物理化学性质的葡萄酒质量的可视化评价研究，燕山大学报[J]，第 34 卷第 2 期：P133~137，2010。
- [6] 骆方、刘红云、黄崑，SPSS 数据统计与分析[M]，北京：清华大学出版社，2011。
- [7] 匿名，<http://wenku.baidu.com/view/f9ed71ed6294dd88d0d26b40.html>，葡萄酒的品尝，2012.9.8。

一、 附录：

附录一：一些表格

表 7-1 第一组评酒员对红葡萄酒的总评分

编号	评酒 员 1	评酒 员 2	评酒 员 3	评酒 员 4	评酒 员 5	评酒 员 6	评酒 员 7	评酒 员 8	评酒 员 9	评酒 员 10
样品 1	51	66	49	54	77	61	72	61	74	62
样品 2	71	81	86	74	91	80	83	79	85	73
样品 3	80	85	89	76	69	89	73	83	84	76
样品 4	52	64	65	66	58	82	76	63	83	77
样品 5	74	74	72	62	84	63	68	84	81	71
样品 6	72	69	71	61	82	69	69	64	81	84
样品 7	63	70	76	64	59	84	72	59	84	84
样品 8	64	76	65	65	76	72	69	85	75	76
样品 9	77	78	76	82	85	90	76	92	80	79
样品 10	67	82	83	68	75	73	75	68	76	75
样品 11	73	60	72	63	63	71	70	66	90	73
样品 12	54	42	40	55	53	60	47	61	58	69
样品 13	69	84	79	59	73	77	77	76	75	77
样品 14	70	77	70	70	80	59	76	76	76	76
样品 15	69	50	50	58	51	50	56	60	67	76
样品 16	72	80	80	71	69	71	80	74	78	74
样品 17	70	79	91	68	97	82	69	80	81	76
样品 18	63	65	49	55	52	57	62	58	70	68
样品 19	76	84	84	66	68	87	80	78	82	81
样品 20	78	84	76	68	82	79	76	76	86	81
样品 21	73	90	96	71	69	60	79	73	86	74
样品 22	73	83	72	68	93	72	75	77	79	80
样品 23	83	85	86	80	95	93	81	91	84	78
样品 24	70	85	90	68	90	84	70	75	78	70
样品 25	60	78	81	62	70	67	64	62	81	67
样品 26	73	80	71	61	78	71	72	76	79	77
样品 27	70	77	63	64	80	76	73	67	85	75

表 7-2 第二组评酒员对红葡萄酒的总评分

编号	评酒 员 1	评酒 员 2	评酒 员 3	评酒 员 4	评酒 员 5	评酒 员 6	评酒 员 7	评酒 员 8	评酒 员 9	评酒 员 10
样品 1	68	71	80	52	53	76	71	73	70	67
样品 2	75	76	76	71	68	74	83	73	73	71
样品 3	82	69	80	78	63	75	72	77	74	76
样品 4	75	79	73	72	60	77	73	73	60	70
样品 5	66	68	77	75	76	73	72	72	74	68
样品 6	65	67	75	61	58	66	70	67	67	67
样品 7	68	65	68	65	47	70	57	74	72	67
样品 8	71	70	78	51	62	69	73	59	68	59

样品 9	81	83	85	76	69	80	83	77	75	73
样品 10	67	73	82	62	63	66	66	72	65	72
样品 11	64	61	67	62	50	66	64	51	67	64
样品 12	67	68	75	58	63	73	67	72	69	71
样品 13	74	64	68	65	70	67	70	76	69	65
样品 14	71	71	78	64	67	76	74	80	73	72
样品 15	62	60	73	54	59	71	71	70	68	69
样品 16	71	65	78	70	64	73	66	75	68	69
样品 17	72	73	75	74	75	77	79	76	76	68
样品 18	67	65	80	55	62	64	62	74	60	65
样品 19	72	65	82	61	64	81	76	80	74	71
样品 20	80	75	80	66	70	84	79	83	71	70
样品 21	80	72	75	72	62	77	63	70	73	78
样品 22	77	79	75	62	68	69	73	71	69	73
样品 23	79	77	80	83	67	79	80	71	81	74
样品 24	66	69	72	73	73	68	72	76	76	70
样品 25	68	68	84	62	60	66	69	73	66	66
样品 26	68	67	83	64	73	74	77	78	63	73
样品 27	71	64	72	71	69	71	82	73	73	69

表 7-3 第一组评酒员对白葡萄酒的总评分

	评酒 员 1	评酒 员 2	评酒 员 3	评酒 员 4	评酒 员 5	评酒 员 6	评酒 员 7	评酒 员 8	评酒 员 9	评酒 员 10
样品 1	85	80	88	61	76	93	83	80	95	79
样品 2	78	47	86	54	79	91	85	68	73	81
样品 3	85	67	89	75	78	75	136	79	90	79
样品 4	75	77	80	65	77	83	88	78	85	86
样品 5	84	47	77	60	79	62	74	74	79	74
样品 6	61	45	83	65	78	56	80	67	65	84
样品 7	84	81	83	66	74	80	80	68	77	82
样品 8	75	46	81	54	81	59	73	77	85	83
样品 9	79	69	81	60	70	55	73	81	76	85
样品 10	75	42	86	60	87	75	83	73	91	71
样品 11	79	46	85	60	74	71	86	62	88	72
样品 12	64	42	75	52	67	62	77	56	68	70
样品 13	82	42	83	49	66	65	76	62	65	69
样品 14	78	48	84	67	79	64	78	68	81	73
样品 15	74	48	87	71	81	61	79	67	74	82
样品 16	69	49	86	65	70	91	87	62	84	77
样品 17	81	54	90	70	78	71	87	74	92	91
样品 18	86	44	83	71	72	71	85	64	74	81
样品 19	75	66	83	68	73	64	80	63	73	77
样品 20	80	68	82	71	83	81	84	62	87	80

样品 21	84	49	85	59	76	86	83	70	88	84
样品 22	65	48	90	58	72	77	76	70	80	74
样品 23	71	66	80	69	80	82	78	71	87	75
样品 24	82	56	79	73	67	59	68	78	86	85
样品 25	86	80	82	69	74	67	77	78	77	81
样品 26	75	66	82	75	93	91	81	76	90	84
样品 27	58	40	79	67	59	55	66	74	73	77
样品 28	66	75	89	69	88	87	85	76	88	90

表 7-4 第二组评酒员对白葡萄酒的总评分

编号	评酒 员 1	评酒 员 2	评酒 员 3	评酒 员 4	评酒 员 5	评酒 员 6	评酒 员 7	评酒 员 8	评酒 员 9	评酒 员 10
样品 1	84	78	82	75	79	84	81	69	75	72
样品 2	79	76	77	85	77	79	80	59	76	70
样品 3	85	74	71	87	79	79	80	45	83	73
样品 4	84	78	74	83	69	82	84	66	77	72
样品 5	83	79	79	80	77	87	82	73	84	91
样品 6	83	75	74	69	75	77	80	67	77	78
样品 7	78	79	74	69	69	82	80	61	72	78
样品 8	74	78	74	67	73	77	79	66	73	62
样品 9	77	78	89	88	84	89	85	54	79	81
样品 10	86	77	77	82	81	87	84	61	73	90
样品 11	79	83	78	63	60	73	81	61	60	76
样品 12	73	81	73	79	67	79	80	44	64	84
样品 13	68	78	79	81	78	72	75	62	65	81
样品 14	75	77	76	76	78	82	79	68	78	82
样品 15	83	77	88	80	84	83	80	63	76	70
样品 16	68	63	75	60	67	86	67	71	52	64
样品 17	77	69	79	83	79	87	88	75	78	88
样品 18	75	83	82	79	74	84	78	71	74	67
样品 19	76	75	78	70	81	80	83	66	78	77
样品 20	86	74	75	78	85	81	78	61	73	75
样品 21	81	80	79	85	83	76	80	58	85	85
样品 22	80	76	82	88	75	89	80	66	72	86
样品 23	74	80	80	80	74	79	75	73	83	76
样品 24	67	80	77	77	79	78	83	65	72	83
样品 25	79	76	79	86	83	88	83	52	85	84
样品 26	80	72	75	83	71	83	83	53	62	81
样品 27	72	79	84	79	76	83	77	63	79	78
样品 28	75	82	81	81	78	84	79	71	76	89

表 7-5

品种 编号	花色苷 (mg/L)	单宁 (mmol/ L)	总酚 (mmol/ L)	酒总黄 酮 (mmol/ L)	白藜芦 醇 (mg/L)	DPPH 半抑 制体 积 (IV5 0) 1/IV5 0(uL)	色泽		
红葡萄 酒							L*(D6 5)	a*(D6 5)	b*(D6 5)
1	973.8783	11.0295	9.9826	8.0199	2.4382	0.3585	2.48	16.10	3.88
2	517.5813	11.0782	9.5598	13.3001	3.6484	0.4603	14.26	45.77	24.06
3	398.7700	13.2593	8.5494	7.3675	5.2456	0.3960	16.39	48.04	27.56
4	183.5195	6.4774	5.9820	4.3063	2.9337	0.1769	42.30	59.53	26.75
5	280.1905	5.8493	6.0336	3.6437	4.9969	0.2068	34.46	60.16	24.05
6	117.0255	7.3537	5.8583	4.4449	4.4311	0.2113	56.95	54.43	23.57
7	90.8248	4.0139	3.8581	2.7653	1.8205	0.1120	59.00	48.82	32.07
8	918.6879	12.0276	10.1372	7.7476	1.0158	0.3464	8.60	38.86	14.68
9	387.7649	12.9331	11.3126	9.9049	3.8599	0.3857	14.17	46.09	24.19
10	138.7138	5.5670	4.3427	3.1454	3.2459	0.1362	57.09	58.06	8.
11	11.8378	4.5884	4.0230	2.1027	0.3816	0.1052	88.79	12.14	19.54
12	84.0786	6.4579	4.8170	2.9862	2.1628	0.1410	53.68	50.45	30.59
13	200.0803	6.3849	4.9304	3.9570	1.3388	0.1665	41.59	58.73	19.60
14	251.5701	6.0733	5.0129	3.0684	2.1659	0.1629	24.22	56.17	35.30
15	122.5916	3.9847	4.0643	1.8357	0.8886	0.0682	52.95	57.87	19.09
16	171.5019	4.8318	4.0437	2.6678	1.1620	0.1171	50.47	59.45	18.20
17	234.4200	9.1697	6.1676	4.9124	1.6504	0.3102	41.21	56.03	25.12
18	71.9017	4.4472	4.3530	3.5307	1.7396	0.1383	58.18	54.72	22.55
19	198.6145	5.9808	5.1572	3.8748	9.0269	0.1667	47.70	64.93	20.67
20	74.3767	5.8640	4.8582	4.0443	0.9641	0.1576	78.48	26.39	15.87
21	313.7843	10.0899	8.9412	4.4398	8.7937	0.3577	21.50	52.80	35.21
22	251.0172	7.1054	6.1986	5.8266	4.4666	0.2311	40.55	54.05	26.20
23	413.9404	10.8883	12.5293	12.1444	12.6821	0.5665	14.60	46.86	25.07
24	270.1084	5.7471	5.3943	3.7310	6.8689	0.1650	42.84	59.06	17.68
25	158.5686	5.4063	4.4251	3.0222	2.5789	0.1651	50.24	63.78	11.53
26	151.4805	3.6147	3.8890	2.1541	2.7369	0.0760	33.50	62.05	29.18
27	138.4546	5.9613	4.7345	3.2841	4.7758	0.1513	63.14	48.73	15.98
白葡萄 酒									
1		1.6199	1.2635	0.1047	0.3090	0.0348	102.11	-0.51	2.11
2		1.2334	1.1037	0.5105	0.2154	0.0331	101.85	-0.59	3.16
3		2.0094	1.8203	3.6693	0.3484	0.0474	101.79	-0.48	2.94
4		2.0167	1.4852	1.1320	0.1119	0.0526	101.70	-0.87	4.05

5		1.5947	1.5368	1.4145	0.3127	0.0406	101.82	-1.15	4.37
6		1.2888	1.1759	0.0790	0.1757	0.0420	102.07	-0.58	2.64
7		1.3740	1.2017	3.9313	0.3711	0.0522	101.86	-0.26	2.26
8		1.5128	0.4722	0.5772	0.5844	0.0392	102.10	-0.68	2.61
9		1.8438	1.2867	0.0996	0.1993	0.0400	101.73	-0.79	3.88
10		2.0581	1.3254	1.5634	0.0324	0.0640	102.05	-0.49	2.27
11		1.4154	1.2764	2.2568	0.1074	0.0243	101.93	-0.51	2.61
12		2.3074	1.9982	1.4915	0.4335	0.0817	101.92	-0.60	3.04
13		1.5152	1.3563	2.0360	0.5871	0.0470	102.21	-0.55	2.11
14		1.3205	1.3202	2.5445	1.2058	0.0491	102.05	-0.63	2.68
15		2.5303	1.8074	0.9419	0.3542	0.0738	101.99	-0.45	2.71
16		1.2791	1.3073	1.9230	0.5635	0.0315	101.81	-0.75	3.79
17		1.5493	1.2687	0.5002	0.1350	0.1321	101.89	-0.76	3.25
18		1.3302	1.3434	2.8783	0.4211	0.0385	102.12	-0.42	1.84
19		1.9631	1.3434	0.4077	0.0825	0.0373	101.69	-0.56	3.59
20		2.6764	1.3151	0.9008	0.4259	0.0544	102.01	-0.65	2.77
21		1.2036	1.0290	0.5413	0.3599	0.0464	101.97	-0.55	3.62
22		1.8974	1.3795	0.0893	1.2596	0.0498	101.76	-0.94	4.19
23		1.3302	1.1140	0.0996	0.1524	0.0382	101.39	-0.91	4.98
24		4.4729	3.4339	3.3047	0.2662	0.1434	101.66	-0.45	3.78
25		1.5055	1.4594	2.3339	0.2594	0.0306	101.30	-0.42	4.32
26		1.5688	1.2584	0.8649	0.7478	0.0441	101.01	-1.21	7.08
27		3.3750	2.5395	7.6552	0.1539	0.1031	101.62	-0.59	4.42
28		2.0289	1.5445	0.4231	0.0838	0.0541	100.89	-0.61	5.71

表 7-6 葡萄酒质量的三个决定因素的分数（满分 100）

编号	红葡萄酒			白葡萄酒		
	外观	香气	口感	外观	香气	口感
1	71.33	66.33	66.14	67.33	78.33	79.32
2	67.33	73.67	74.32	72.67	75.67	75.00
3	68.00	75.33	74.77	64.67	76.67	76.36
4	66.00	75.00	68.18	69.33	73.67	79.77
5	72.00	69.67	71.59	70.67	85.67	80.91
6	58.00	67.00	65.68	60.00	79.00	76.36
7	50.00	67.00	66.59	66.00	77.67	73.64
8	68.00	64.67	63.64	66.67	73.33	70.91
9	73.33	85.00	73.41	73.33	87.67	76.59
10	70.67	70.00	65.45	60.67	85.00	81.14
11	46.67	69.00	58.64	60.67	72.00	71.36
12	58.00	66.00	70.68	69.33	77.33	68.64
13	63.33	74.33	64.55	62.00	76.00	75.00
14	71.33	72.67	70.00	68.00	78.00	77.50
15	66.67	63.00	65.00	67.33	82.67	77.73

16	66.67	69.00	69.09	53.33	63.67	71.59
17	68.00	77.67	72.73	68.67	80.67	82.05
18	52.00	62.33	68.41	68.00	75.00	79.09
19	70.00	75.00	70.23	64.67	78.33	77.50
20	56.00	83.00	75.45	70.00	78.33	76.14
21	68.00	71.33	71.82	70.67	79.33	80.23
22	62.67	73.67	70.91	72.67	85.33	76.59
23	76.00	83.67	72.05	72.67	77.33	76.59
24	67.33	73.33	69.55	70.00	73.00	78.41
25	68.67	70.33	64.09	70.67	80.67	79.77
26	74.00	73.67	68.18	74.00	73.33	72.73
27	66.00	71.67	71.14	72.67	79.33	75.00
28				73.33	80.33	79.55

附录二：第一题程序1

```
clear
clc
%红
a1=[51 66 49 54 77 61 72 61 74 62
71 81 86 74 91 80 83 79 85 73
80 85 89 76 69 89 73 83 84 76
52 64 65 66 58 82 76 63 83 77
74 74 72 62 84 63 68 84 81 71
72 69 71 61 82 69 69 64 81 84
63 70 76 64 59 84 72 59 84 84
64 76 65 65 76 72 69 85 75 76
77 78 76 82 85 90 76 92 80 79
67 82 83 68 75 73 75 68 76 75
73 60 72 63 63 71 70 66 90 73
54 42 40 55 53 60 47 61 58 69
69 84 79 59 73 77 77 76 75 77
70 77 70 70 80 59 76 76 76 76
69 50 50 58 51 50 56 60 67 76
72 80 80 71 69 71 80 74 78 74
70 79 91 68 97 82 69 80 81 76
63 65 49 55 52 57 62 58 70 68
76 84 84 66 68 87 80 78 82 81
78 84 76 68 82 79 76 76 86 81
73 90 96 71 69 60 79 73 86 74
73 83 72 68 93 72 75 77 79 80
83 85 86 80 95 93 81 91 84 78
70 85 90 68 90 84 70 75 78 70
60 78 81 62 70 67 64 62 81 67
```

```

73 80 71 61 78 71 72 76 79 77
70 77 63 64 80 76 73 67 85 75
];
a2=[68 71 80 52 53 76 71 73 70 67
75 76 76 71 68 74 83 73 73 71
82 69 80 78 63 75 72 77 74 76
75 79 73 72 60 77 73 73 60 70
66 68 77 75 76 73 72 72 74 68
65 67 75 61 58 66 70 67 67 67
68 65 68 65 47 70 57 74 72 67
71 70 78 51 62 69 73 59 68 59
81 83 85 76 69 80 83 77 75 73
67 73 82 62 63 66 66 72 65 72
64 61 67 62 50 66 64 51 67 64
67 68 75 58 63 73 67 72 69 71
74 64 68 65 70 67 70 76 69 65
71 71 78 64 67 76 74 80 73 72
62 60 73 54 59 71 71 70 68 69
71 65 78 70 64 73 66 75 68 69
72 73 75 74 75 77 79 76 76 68
67 65 80 55 62 64 62 74 60 65
72 65 82 61 64 81 76 80 74 71
80 75 80 66 70 84 79 83 71 70
80 72 75 72 62 77 63 70 73 78
77 79 75 62 68 69 73 71 69 73
79 77 80 83 67 79 80 71 81 74
66 69 72 73 73 68 72 76 76 70
68 68 84 62 60 66 69 73 66 66
68 67 83 64 73 74 77 78 63 73
71 64 72 71 69 71 82 73 73 69
];
h=[];
for i=1:27
[h(i),p,ci] = ttest2(a1(i,:),a2(i,:),0.01)

end

%白
b2=[84 78 82 75 79 84 81 69 75 72
79 76 77 85 77 79 80 59 76 70
85 74 71 87 79 79 80 45 83 73
84 78 74 83 69 82 84 66 77 72

```

83	79	79	80	77	87	82	73	84	91
83	75	74	69	75	77	80	67	77	78
78	79	74	69	69	82	80	61	72	78
74	78	74	67	73	77	79	66	73	62
77	78	89	88	84	89	85	54	79	81
86	77	77	82	81	87	84	61	73	90
79	83	78	63	60	73	81	61	60	76
73	81	73	79	67	79	80	44	64	84
68	78	79	81	78	72	75	62	65	81
75	77	76	76	78	82	79	68	78	82
83	77	88	80	84	83	80	63	76	70
68	63	75	60	67	86	67	71	52	64
77	69	79	83	79	87	88	75	78	88
75	83	82	79	74	84	78	71	74	67
76	75	78	70	81	80	83	66	78	77
86	74	75	78	85	81	78	61	73	75
81	80	79	85	83	76	80	58	85	85
80	76	82	88	75	89	80	66	72	86
74	80	80	80	74	79	75	73	83	76
67	80	77	77	79	78	83	65	72	83
79	76	79	86	83	88	83	52	85	84
80	72	75	83	71	83	83	53	62	81
72	79	84	79	76	83	77	63	79	78
75	82	81	81	78	84	79	71	76	89

];

b1=[85	80	88	61	76	93	83	80	95	79
78	47	86	54	79	91	85	68	73	81
85	67	89	75	78	75	136	79	90	79
75	77	80	65	77	83	88	78	85	86
84	47	77	60	79	62	74	74	79	74
61	45	83	65	78	56	80	67	65	84
84	81	83	66	74	80	80	68	77	82
75	46	81	54	81	59	73	77	85	83
79	69	81	60	70	55	73	81	76	85
75	42	86	60	87	75	83	73	91	71
79	46	85	60	74	71	86	62	88	72
64	42	75	52	67	62	77	56	68	70
82	42	83	49	66	65	76	62	65	69
78	48	84	67	79	64	78	68	81	73
74	48	87	71	81	61	79	67	74	82
69	49	86	65	70	91	87	62	84	77
81	54	90	70	78	71	87	74	92	91
86	44	83	71	72	71	85	64	74	81

```

75 66 83 68 73 64 80 63 73 77
80 68 82 71 83 81 84 62 87 80
84 49 85 59 76 86 83 70 88 84
65 48 90 58 72 77 76 70 80 74
71 66 80 69 80 82 78 71 87 75
82 56 79 73 67 59 68 78 86 85
86 80 82 69 74 67 77 78 77 81
75 66 82 75 93 91 81 76 90 84
58 40 79 67 59 55 66 74 73 77
66 75 89 69 88 87 85 76 88 90
];
h1=[];
for i=1:28
[h1(i),p,ci] = ttest2(b1(i,:),b2(i,:),0.01)

end

```

附录二：第一题程序 2

```

clear
clc
% a1=[51      66  49  54  77  61  72  61  74  62
% 71      81  86  74  91  80  83  79  85  73
% 80      85  89  76  69  89  73  83  84  76
% 52      64  65  66  58  82  76  63  83  77
% 74      74  72  62  84  63  68  84  81  71
% 72      69  71  61  82  69  69  64  81  84
% 63      70  76  64  59  84  72  59  84  84
% 64      76  65  65  76  72  69  85  75  76
% 77      78  76  82  85  90  76  92  80  79
% 67      82  83  68  75  73  75  68  76  75
% 73      60  72  63  63  71  70  66  90  73
% 54      42  40  55  53  60  47  61  58  69
% 69      84  79  59  73  77  77  76  75  77
% 70      77  70  70  80  59  76  76  76  76
% 69      50  50  58  51  50  56  60  67  76
% 72      80  80  71  69  71  80  74  78  74
% 70      79  91  68  97  82  69  80  81  76
% 63      65  49  55  52  57  62  58  70  68
% 76      84  84  66  68  87  80  78  82  81
% 78      84  76  68  82  79  76  76  86  81
% 73      90  96  71  69  60  79  73  86  74
% 73      83  72  68  93  72  75  77  79  80
% 83      85  86  80  95  93  81  91  84  78
% 70      85  90  68  90  84  70  75  78  70

```

```

% 60      78  81  62  70  67  64  62  81  67
% 73      80  71  61  78  71  72  76  79  77
% 70      77  63  64  80  76  73  67  85  75];
% a2=[68      71  80  52  53  76  71  73  70  67
% 75      76  76  71  68  74  83  73  73  71
% 82      69  80  78  63  75  72  77  74  76
% 75      79  73  72  60  77  73  73  60  70
% 66      68  77  75  76  73  72  72  74  68
% 65      67  75  61  58  66  70  67  67  67
% 68      65  68  65  47  70  57  74  72  67
% 71      70  78  51  62  69  73  59  68  59
% 81      83  85  76  69  80  83  77  75  73
% 67      73  82  62  63  66  66  72  65  72
% 64      61  67  62  50  66  64  51  67  64
% 67      68  75  58  63  73  67  72  69  71
% 74      64  68  65  70  67  70  76  69  65
% 71      71  78  64  67  76  74  80  73  72
% 62      60  73  54  59  71  71  70  68  69
% 71      65  78  70  64  73  66  75  68  69
% 72      73  75  74  75  77  79  76  76  68
% 67      65  80  55  62  64  62  74  60  65
% 72      65  82  61  64  81  76  80  74  71
% 80      75  80  66  70  84  79  83  71  70
% 80      72  75  72  62  77  63  70  73  78
% 77      79  75  62  68  69  73  71  69  73
% 79      77  80  83  67  79  80  71  81  74
% 66      69  72  73  73  68  72  76  76  70
% 68      68  84  62  60  66  69  73  66  66
% 68      67  83  64  73  74  77  78  63  73
% 71      64  72  71  69  71  82  73  73  69
% ];
% a_m=a1-a2;
% [h,p,ci]=ttest(a_m')
%白
a1=[62.7
      80.3
      80.4
      68.6
      73.3
      72.2
      71.5
      72.3
      81.5
      74.2

```

70.1
53.9
74.6
73
58.7
74.9
79.3
59.9
78.6
78.6
77.1
77.2
85.6
78
69.2
73.8
73

];
a2=[68.10
74.00
74.60
71.20
72.10
66.30
65.30
66.00
78.20
68.80
61.60
68.30
68.80
72.60
65.70
69.90
74.50
65.40
72.60
75.80
72.20
71.60
77.10
71.50


```

68.20
72.00
71.50
];
a_m=a1-a2;
[h,p,ci]=ttest(a_m,0,0.01)
附录三：第二题程序 1
a1=zscore(a1); %数据标准化
r=corrcoef(a1); %计算相关系数矩阵
[x,y,z]=pcacov(r);
f= repmat(sign(sum(x)),size(x,1),1);
x=x.*f;
df=a1*x(:,1:14)%保留有效主成分
tf=df*z(1:14)/100;
[stf,ind]=sort(tf,'descend')

```

```

% a2=zscore(a2); %数据标准化
% r=corrcoef(a2); %计算相关系数矩阵
% [x,y,z]=pcacov(r);
% f= repmat(sign(sum(x)),size(x,1),1);
% x=x.*f;
% df=a2*x(:,1:15)%保留有效主成分
% tf=df*z(1:15)/100;
% [stf,ind]=sort(tf,'descend')

```

附录四：第二题程序 2

```

a1=zscore(a1);
y=pdist(a1);
z=linkage(y,'average')
dendrogram(z,30)
hold on

```

附录五：第二题程序 3

```

j=[68.10
74.00
74.60
71.20
72.10
66.30
65.30
66.00
78.20
68.80
61.60

```

```
68.30
68.80
72.60
65.70
69.90
74.50
65.40
72.60
75.80
72.20
71.60
77.10
71.50
68.20
72.00
71.50
];
c=[ 3
    9
   11
    1
    2
   21
   23
   22
    8
   16
   12
   14
    6
   17
    5
   24
   19
   18
    7
   13
   15
    4
   20
   27
   26
   10
   25
```

```

];
z=[ 100.0000
    92.1548
    91.5646
    90.5484
    90.3726
    89.7398
    80.0644
    79.7510
    78.5290
    76.9001
    76.0300
    75.2252
    74.7762
    72.5157
    72.5044
    71.8535
    70.8747
    70.2627
    70.2074
    70.1563
    68.8653
    64.2187
    64.2029
    62.5150
    59.7817
    56.7221
    50.0000
];
d=[];
for i=1:27
    d(i)=0.3*z(i)+0.7*j(c(i));
end
key=d'
l=[]

for i=1:27
    if key(i)<65
        l(c(i))=1;
    else
        if key(i)<70&&key(i)>=65
            l(c(i))=2
        else

```

```

        if key(i)<75&&key(i)>=70
            l(c(i))=3
        else
            if key(i)<80&&key(i)>=75
                l(c(i))=4
            else
                if key(i)<85&&key(i)>=80
                    l(c(i))=5
                else
                    l(c(i))=6;
                end
            end
        end
    end
end
end
end
end

```

附录六：第二题程序 4

```
j=[77.90
```

```
75.80
```

```
75.60
```

```
76.90
```

```
81.50
```

```
75.50
```

```
74.20
```

```
72.30
```

```
80.40
```

```
79.80
```

```
71.40
```

```
72.40
```

```
73.90
```

```
77.10
```

```
78.40
```

```
67.30
```

```
80.30
```

```
76.70
```

```
76.40
```

```
76.60
```

```
79.20
```

```
79.40
```

```
77.40
```

```
76.10
```

```
79.50
```

```
74.30
```

```
77.00
```

```

79.60
];
c= [ 27
    24
    28
     5
     3
     7
    15
     6
    20
    10
    18
    25
    23
    26
     9
    12
    14
     2
     4
    21
    11
    22
    13
    19
     1
     8
    17
    16];
z=[ 100.0000
    84.4601
    79.1374
    77.0216
    76.5584
    75.4642
    73.2349
    72.5076
    71.6516
    70.3277
    69.9057
    69.5298
    68.9856
    68.8233

```


附录七：第四题程序

%hong

```
pt_cl=[100.0000    81.1893    82.5127   100.0000    85.5962
      80.7023    78.5723    87.2481    76.2848    76.9257
      70.9518   100.0000    87.2127    70.1099    60.3872
      63.5810    77.0850    63.0599    58.9228    77.1614
      78.5258    72.0830    61.7533    63.9470    82.7115
      50.4933    86.1898    65.9200    55.4671    84.5523
      76.7847    96.3908    52.4295    54.1052    78.0985
      99.0424    70.0076    84.4628    97.1316    60.5743
      89.4080    68.3918    91.2013    69.5377    72.0799
      50.3482    54.9745    57.9248    56.5939    65.6586
      93.3256    81.5501    52.9452    50.0000    51.6194
      72.2577    95.4974    60.2023    53.7545    74.3234
      54.0917    73.1961    60.6554    59.7835    79.5776
      69.1817    81.4145    59.5032    62.4595    50.0000
      70.4585    68.2639    51.4027    55.7561    78.9917
      90.1045    69.4362    54.4108    58.2981    74.7214
      70.3134    89.7786    70.2605    61.5682    85.4913
      64.0453    91.4691    54.7580    53.1217    56.9242
      64.7707    76.7802    62.9236    59.7071    74.7872
      50.0000    75.5289    59.1455    53.2503    59.7670
      96.2855    85.4524    78.3362    65.6928    69.2648
      79.3674    82.2429    66.8878    62.4310    96.6887
      63.9292    75.4654    92.8434    70.8984    81.5129
      78.4098    81.3021    61.5876    63.4231    79.7189
      51.8572    50.0000    57.2239    57.6261    74.0295
      63.0586    64.3912    51.2705    57.2576    72.3070
      61.6367    73.1489    60.1404    56.5804   100.0000];
pf=[71.33333333 66.33333333 66.13636364
    67.33333333 73.66666667 74.31818182
    68 75.33333333 74.77272727
    66 75 68.18181818
    72 69.66666667 71.59090909
    58 67 65.68181818
    50 67 66.59090909
    68 64.66666667 63.63636364
    73.33333333 85 73.40909091
    70.66666667 70 65.45454545
    46.66666667 69 58.63636364
    58 66 70.68181818
    63.33333333 74.33333333 64.54545455
    71.33333333 72.66666667 70
```

```

66.66666667 63 65
66.66666667 69 69.09090909
68 77.66666667 72.72727273
52 62.33333333 68.40909091
70 75 70.22727273
56 83 75.45454545
68 71.33333333 71.81818182
62.66666667 73.66666667 70.90909091
76 83.66666667 72.04545455
67.33333333 73.33333333 69.54545455
68.66666667 70.33333333 64.09090909
74 73.66666667 68.18181818
66 71.66666667 71.13636364
];
%bai
% pt_cl =[ 60.3354 59.8738 57.0165 93.2990 73.8123
% 60.1985 82.1437 53.6713 51.5464 62.1437
% 100.0000 76.5805 66.7334 58.0183 100.0000
% 67.1458 78.6354 61.3513 81.2142 69.1481
% 99.0418 91.5790 59.4836 54.5246 55.3437
% 75.7358 74.4717 53.6107 62.8293 70.1973
% 70.1916 73.6145 60.2291 60.4811 50.5780
% 69.1650 57.2756 55.2695 50.0000 69.2789
% 88.3984 82.0839 58.3537 51.0882 87.7362
% 64.0315 92.2784 61.3744 53.2646 66.8965
% 64.9213 53.1248 57.5085 50.0573 53.4419
% 50.0000 91.8861 67.0865 55.3837 50.2086
% 59.6167 50.0000 60.2061 79.7251 55.5109
% 52.4983 82.3272 61.7922 57.6747 79.7733
% 61.4648 66.6698 67.0980 50.6300 57.8473
% 59.4114 68.2088 57.9905 63.4021 52.1484
% 67.9329 64.1416 56.3105 100.0000 81.4706
% 67.9671 66.6522 59.1824 58.8202 57.8401
% 79.8768 75.3476 59.3869 50.1718 64.5688
% 98.4257 90.2517 66.7992 53.0928 83.5089
% 65.6742 84.6186 53.8246 87.3998 70.4858
% 67.0773 69.9468 63.3751 79.8396 75.8966
% 63.6550 89.5941 53.6477 52.0046 79.0904
% 61.1567 90.8015 90.2094 56.9301 84.5081
% 71.2183 97.8873 59.5432 54.5246 77.7720
% 50.3080 100.0000 59.4006 59.1638 50.0000
% 93.1211 77.4087 84.0764 60.8247 82.7714
% 85.1129 91.7309 60.5953 54.9255 63.0300];
% pf=[67.33 78.33 79.32

```



```

% 72.67      75.67      75.00
% 64.67      76.67      76.36
% 69.33      73.67      79.77
% 70.67      85.67      80.91
% 60.00      79.00      76.36
% 66.00      77.67      73.64
% 66.67      73.33      70.91
% 73.33      87.67      76.59
% 60.67      85.00      81.14
% 60.67      72.00      71.36
% 69.33      77.33      68.64
% 62.00      76.00      75.00
% 68.00      78.00      77.50
% 67.33      82.67      77.73
% 53.33      63.67      71.59
% 68.67      80.67      82.05
% 68.00      75.00      79.09
% 64.67      78.33      77.50
% 70.00      78.33      76.14
% 70.67      79.33      80.23
% 72.67      85.33      76.59
% 72.67      77.33      76.59
% 70.00      73.00      78.41
% 70.67      80.67      79.77
% 74.00      73.33      72.73
% 72.67      79.33      75.00
% 73.33      80.33      79.55
% ];
xgd=[];
for i=1:3
    for j=1:5
        t=corrcoef(pf(:,i),pt_cl(:,j));
        xgd(i,j)=t(1,2);
        xgd(i,j)=abs(xgd(i,j));
    end
    xgd(i,:)=xgd(i,:)/sum(xgd(i,:))
end
es=(xgd(1,5)*15+xgd(2,5)*30+xgd(3,5)*44)/89

```