

# 葡萄酒的评价

## 摘要

葡萄酒质量由评酒员打分确定，且与酿酒葡萄有直接联系，酿酒葡萄和葡萄酒的各理化指标可在一定程度上反映葡萄及葡萄酒的质量。

针对问题一，首先我们对两组评酒员的评价进行**显著性检验**，求得葡萄酒的十个评判标准的自由度及检验统计量，并将其与检验统计量的临界值进行比较；随后我们利用**变异系数**评判两组结果的可信度；最后我们得出结论：红葡萄酒外观分析中的色调及香气分析中的质量的评价结果有显著性差异；对于前者，第一组评酒员的结果更可信，其余均是第二组评酒员的可信度更高。

针对问题二，我们以红葡萄为例，在信息保留率为 90%的前提下对它的一级指标进行**主成分分析**，随后在主成分中筛选出影响系数较大的主要指标；接着我们依据这些指标，利用SPSS Statistics 22软件对红葡萄样本进行**聚类分析**，根据第二组评酒员对红葡萄酒的评价，将已分好类的红葡萄样本分为五级；最后我们得出结论：红葡萄样本 9 的评级最高，样本 12 的评级最低；白葡萄样本 1~11、13、15~18、20~26、28 的评级最高，样本 13、21 的评级最低，具体结果见表 5。

针对问题三，我们利用上题中的主成分分析法，选取两种葡萄酒的主要指标；随后我们建立葡萄酒及酿酒葡萄的主要指标矩阵，并对它们进行**相关性分析**；紧接着我们对分析结果进行显著性检验；最后得出结论：在红葡萄酒与红葡萄的主要指标中，有 114 对指标高度相关，31 对指标非常相关；在白葡萄酒与白葡萄的主要指标中，有 8 对指标高度相关，1 对指标非常相关，具体结论见表 7。

针对问题四，我们以红葡萄为例，利用**多元线性逐步回归模型**，通过SPSS Statistics 22软件，以评分结果中的十项标准为因变量，葡萄的指标为自变量，求解出两者间的多元线性回归方程，随后对其进行残差分析，保证结果的可行性；然后利用**DEA 模型**，通过 MaxDEA 软件，求解得葡萄酒、葡萄的指标与葡萄酒质量的综合效率均趋近于“1”，最后我们得出结论：可用葡萄酒及酿酒葡萄的理化指标对葡萄酒的质量进行评价，两者的理化指标对葡萄酒质量产生的具体影响结果见 5.4.3。

关键词:变异系数 主成分分析 聚类分析 相关性分析 多元线性逐步回归 DEA

## 一、问题重述

葡萄酒质量的品评由一组有资质的评酒员打分决定，评酒员们对不同葡萄酒的各类指标打分并求和得该葡萄酒的总分，从而得到葡萄酒的质量。葡萄酒质量与酿酒葡萄的好坏有关，酿酒葡萄和葡萄酒的理化指标在一定程度上可以反映葡萄和葡萄酒的质量。根据附件 1、2、3，建立数学模型解决下列问题。

- 1.分析附件 1，两组评酒员给出的评价结果是否有显著差异，哪组结果更可靠？
- 2.根据附件中酿酒葡萄的各理化指标及葡萄酒的质量，对酿酒葡萄进行分级。
- 3.分析葡萄酒和酿酒葡萄各理化指标之间的关系。
- 4.分析葡萄酒和酿酒葡萄的各理化指标对葡萄酒质量产生的影响，并论证我们是否可以用葡萄酒和酿酒葡萄的各理化指标对葡萄酒质量进行评价。

## 二、模型假设

1. 假设附件二、三中的二级指标对分析过程没有影响

## 三、符号说明

$H_0$	两组评酒员的评价结果无显著性差异
$H_1$	两组评酒员的评价结果有显著性差异
$\mu_1, \mu_2$	第一、二组评酒员评价结果的总体均值
$n_1, n_2$	红、白葡萄酒样本数量
$\sigma_1^2$	第一组评酒员评价结果的总体方差
$\sigma_2^2$	第二组评酒员评价结果的总体方差
$f$	自由度
$s_{1i}^2, s_{2i}^2$	第一、二组评酒员对第 $i$ 个指标评价结果的方差
$\bar{x}_{1i}, \bar{x}_{2i}$	第一、二组评酒员对第 $i$ 个指标评价结果的均值
$t$	检验统计量
$a_i$	酿酒葡萄的第 $i$ 个指标
$b_j$	葡萄酒的第 $j$ 个指标
$y_i$	评酒员对葡萄酒的评分指标
$x_j$	葡萄酒与葡萄的理化指标

## 四、问题分析

### 4.1 问题一的分析

对于问题一，需对附件一中的数据进行预处理，将处理后的数据进行显著性分析，建立起原始假设及备择假设，通过比较葡萄酒各指标的检验统计量及统计量的临界值，确定各指标的评价结果是否存在显著性差异；然后求解葡萄酒各个指标的变异系数，从而确定评价结果的可信度

### 4.2 问题二的分析

对于问题二，由于酿酒葡萄的理化指标数量较多，故先采用主成分分析法，以红葡萄为例，在信息保留率为 90%的前提下对红葡萄的理化指标进行处理，随后利用聚类分析将附件二中的红葡萄样本进行分类，最后可利用附件一中红葡萄酒的得分对相应的酿酒葡萄进行评级；以同样的方法对白葡萄进行评级。

### 4.3 问题三的分析

对于问题三，首先利用主成分分析法处理葡萄酒的理化指标，然后对酿酒葡萄的主要指标及葡萄酒的理化指标进行相关性分析，从而得到酿酒葡萄和葡萄酒的理化指标的联系并对结果进行显著性检验。

### 4.4 问题四的分析

对于问题四，应求解出葡萄酒质量与葡萄酒及酿酒葡萄的理化指标之间的关系，得到结果之后对其进行检验，使之具有说服力。以综合效率为衡量因素，当其越趋近于“1”时，越能证明可用葡萄酒及酿酒葡萄的理化指标来评价葡萄酒的质量。

## 五、模型的建立和求解

### 5.1 显著性检验、变异系数分析

#### 5.1.1 数据预处理

##### ①求解平均值及标准差

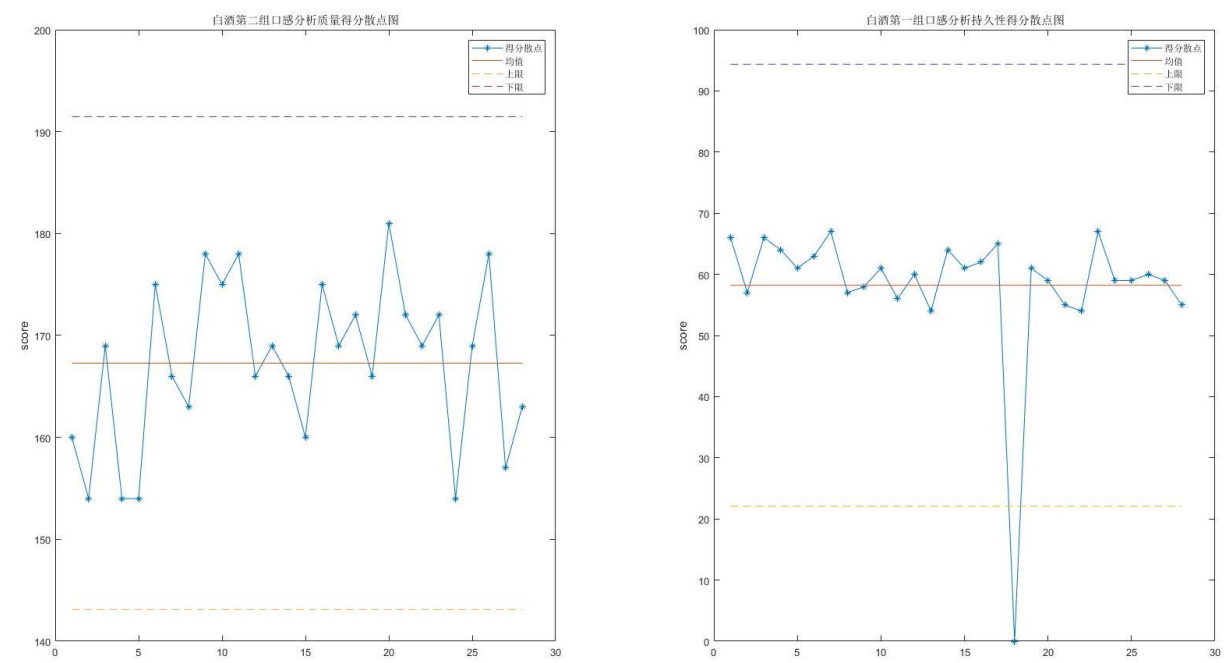
将附件 1 中十个评酒员对葡萄酒的九个指标及最后的整体评价求和，并以此求解得到不同葡萄酒这十个指标的平均值和标准差，结果如下表所示：

表 1：不同品种葡萄酒的各个指标平均值及标准差

指标 \ 种类		红葡萄酒一组		红葡萄酒二组		白葡萄酒一组		白葡萄酒二组	
		均值	标准差	均值	标准差	均值	标准差	均值	标准差
外观分析	澄清度	34.1	7.8	34.9	1.9	32.7	7.1	34.1	1.5
	色调	70.7	12.1	63.2	11.5	67.0	10.0	67.7	6.6
香气分析	纯正度	43.4	7.3	41.5	4.7	45.6	4.3	45.1	3.3
	浓度	59.3	8.6	56.7	5.9	63.6	5.9	62.9	5.3
	质量	123.3	13.5	117.7	8.5	126.5	7.6	125.9	6.9
口感分析	纯正度	40.5	5.8	39.7	2.2	40.2	4.4	45.0	2.5
	浓度	57.4	7.6	55.2	4.6	59.1	5.0	61.5	4.0
	持久性	59.3	5.0	58.2	2.3	62.9	14.1	63.3	2.4
	质量	154.4	16.5	150.4	10.4	155.6	11.5	167.3	8.1
平衡/整体评价		88.0	6.0	87.6	3.2	89.6	4.1	92.5	2.9

②特殊点处理

基于附件一，将 10 个评酒员对酒样本的各个指标的评分加和，以酒样本为横轴，该样本的指标总分为纵轴，绘制出各个指标的酒样本总分散点图（由于数量较多，下图仅以第二组白葡萄酒的第九个指标及第一组白葡萄酒的第八个指标为例）：



(a)

(b)

图 1：样本得分散点图

上图中，实线表示该组某指标的平均值，上、下虚线为平均值加、减三倍标准差所得的值。由图 1 (a) 可知，所有散点均在范围内，可认为这些数值合理有效，故保留。在图 1 (b) 中，第 17 个酒样本的数值显然不在有效范围之内，将该特殊点剔除。其余指标的散点图也做相同处理。

### 5.1.2 以 $\alpha = 0.05$ 的显著性水平进行检验

#### ①求解过程

由于检验两组评酒员对不同酒样的评价结果是否有显著性差异，并没有涉及到方向，所以本题为双侧检验。建立假设并对其进行检验：

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

其中， $H_0$  表示无显著性差异， $H_1$  表示有显著差异。组一和组二的总体均值分别用  $\mu_1$  和  $\mu_2$  表示。

根据附件 1 可知，本题中的两种葡萄酒样本数量  $n_1$  和  $n_2$  分别等于 27、28，为小样本数值；组一和组二中的总体方差  $\sigma_1^2$  和  $\sigma_2^2$  未知，也无法判断  $\sigma_1^2 = \sigma_2^2$  是否成立。故属于  $n$  较小， $\sigma_1^2$  和  $\sigma_2^2$  未知，且  $\sigma_1^2 \neq \sigma_2^2$  的情况。因此，我们采用  $t$  分布，令其自由度为  $f$ ：

$$f = \frac{\left(\frac{s_{1i}^2}{n_1} + \frac{s_{2i}^2}{n_2}\right)^2}{\frac{\left(\frac{s_{1i}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_{2i}^2}{n_2}\right)^2}{n_2 - 1}}$$

其中， $s_{1i}^2$  和  $s_{2i}^2$  分别表示第一、二组中第  $i$  个指标的方差。

此时，检验统计量  $t$ ：

$$t = \frac{(\bar{x}_{1i} - \bar{x}_{2i}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_{1i}^2}{n_1} + \frac{s_{2i}^2}{n_2}}}$$

其中， $\bar{x}_{1i}$ 、 $\bar{x}_{2i}$  分别为第一、二组中第  $i$  个指标的样本均值。

利用 *MATLAB* 软件进行求解，得到红、白葡萄酒的各个指标的自由度  $f$  及

检验统计量  $t$  的值，查询  $t$  临界分布值表可得  $t_{\frac{\alpha}{2}}(f)$ ，如下表所示：

表 2：葡萄酒各指标的自由度、检验统计量及其临界分布值

种类 指标		红葡萄酒			白葡萄酒		
		$f$	$t_{\frac{\alpha}{2}}(f)$	$t$	$f$	$t_{\frac{\alpha}{2}}(f)$	$t$
外观 分析	澄清度	28	1.7011	-0.5065	29	1.6991	-0.9842
	色调	51	2.0117	2.3496	46	1.6787	-0.3001
香气 分析	纯正度	44	1.6802	1.1545	50	2.0086	0.0699
	浓度	46	1.6787	1.3038	53	2.0057	0.4766
	质量	43	1.6811	1.8109	53	2.0057	0.3313
口感 分析	纯正度	33	1.6924	0.6866	31	1.6955	-2.7212
	浓度	42	1.6820	1.2983	51	2.0117	-2.0082
	持久性	36	1.6883	1.0864	42	1.6820	-3.5249
	质量	43	1.6811	1.0479	48	2.0106	-4.3924
平衡/整体评价		39	1.6849	0.3125	48	2.0106	-2.9939

## ②结果分析

当  $t \geq t_{\frac{\alpha}{2}}(f)$  时，拒绝原假设  $H_0$ ，接受备择假设  $H_1$ ，即两组评酒员的评价结果有显著性差异。反之，两组评酒员的评价结果没有显著性差异。

由表 2 可知，红葡萄酒的第二个指标—外观分析中的色调，以及第五个指标—香气分析中的质量的检验统计量  $t$  均大于  $t$  的临界分布值，两组结果有显著性差异，其余指标的评价结果均无显著性差异。

### 5.1.3 可信度分析

某一指标的变异系数可以反映该指标的可信度，变异系数越大，表示某组对该指标的评价结果可信度越低；反之，则代表评价结果的可信度较高。

$$\text{某指标变异系数} = \frac{\text{该指标标准差}}{\text{该指标均值}}$$

我们对两个小组测定的红、白葡萄酒的每个指标进行变异系数求解，结果如下表所示：

表 3：葡萄酒各指标的变异系数值

种类		第一组红葡萄酒	第二组红葡萄酒	第一组白葡萄酒	第二组白葡萄酒
指标					
外观分析	澄清度	22.72%	5.32%	5.32%	4.37%
	色调	17.13%	18.21%	18.21%	9.74%
香气分析	纯正度	16.71%	11.40%	11.40%	7.30%
	浓度	14.55%	10.48%	10.48%	8.45%
	质量	10.92%	7.25%	7.25%	5.50%
口感分析	纯正度	14.21%	5.59%	5.59%	5.65%
	浓度	13.27%	8.32%	8.32%	6.51%
	持久性	8.44%	3.87%	3.87%	3.77%
	质量	10.68%	6.89%	6.89%	4.82%
平衡/整体评价		6.78%	3.65%	3.65%	3.14%

#### 5.1.4 结果分析

两组评酒员对红葡萄酒的第二个指标—外观分析中的色调，以及第五个指标—香气分析中的质量的评价结果有显著性差异；其余指标的评价结果无显著性差异；对于红葡萄酒的第二个指标而言，第一组评酒员的结果可信度更高，其余指标的评价结果均是第二组评酒员的可信度更高。

### 5.2 主成分分析与聚类分析模型

#### 5.2.1 求解过程

##### ①利用主成分分析法处理酿酒葡萄的理化指标

主成分分析法可以把多个指标进行重组，形成一组全新且互相无关的若干个综合指标，在满足一定信息保留率的前提下，选取得到尽可能少量的综合指标，从而达到最大化反映原先指标信息的方法。

假设附件二、三中的二级指标对分析过程没有影响。以红葡萄为例分析，对附件二中的一级指标及附件三中的指标（共计 85 个）进行主成分分析，得到了十三类主成分，其累计贡献率达到 90%，具有统计学意义。在这十三类主成分的各指标中，以 0.6 为界，挑选出影响系数大于 0.6 的共计 45 个指标作为这些主成分的主要影响指标。分析结果如下表所示：

表 4：红葡萄指标主成分分析结果

主成分	所包含的主要指标	累计贡献率
第一主成分	乙醛、乙酸甲酯、乙酸乙酯、3-甲基丁醛、乙醇、丙酸乙酯、2-戊酮、乙酸-2-甲基丙基酯、甲苯、乙酸丁酯、2-甲基-1-丙醇、3-甲基-1-丁醇-乙酸酯、乙酸戊酯、3-甲基-1-丁醇、1-戊醇、乙酸己酯、2-辛酮、4-己烯-1-醇-乙酸盐、1-己醇、乙酸庚酯、(Z)-3-己烯-1-醇、辛酸乙酯、1-辛烯-3-醇、6-甲基-5-庚烯-2-醇、乙酸辛酯、5-甲基糠醛、3,7-二甲基-2,6-辛二烯酸甲酯、萘、甲氧基苯基丙酮肟、2-本乙基乙酸酯、苯乙醇	29.663%
第二主成分	蛋白质、花色苷、DPPH 自由基、总酚、单宁、葡萄总黄酮、(Z)-3,7-二甲基-2,6-辛二烯醛、(R)-3,7-二甲基-6-辛烯醇、(E)-3,7-二甲基-2,6-辛二烯-1-醇	40.423%
第三主成分	苹果酸、可溶性固形物、干物质含量	48.836%
第四主成分	百粒质量、柠檬烯	56.53%
第五主成分	-	62.599%
...	-	...
第十三主成分	-	90.457%

上表表明，这十三类主成分的累计贡献率已大于 90%，具有代表性，且从第五类主成分开始，不含有主要指标。

用相同的方法对白葡萄的理化指标进行处理，在满足信息保留率达到 90% 的前提下，共有十六类主成分，对其主要理化指标进行筛选，最后得到 35 个主要理化指标。

## ②利用聚类分析处理葡萄样本

聚类分析可将性质相近的成分归为同一类，从而使得同类的成分具有较为高度的同质性，且不同类的成分间具有较为高度的异质性。

利用附件二中 27 个红葡萄样本，根据①中筛选得到的 45 个主要指标，通过 SPSS Statistics 22 软件对其进行聚类分析，分析结果如下图所示：



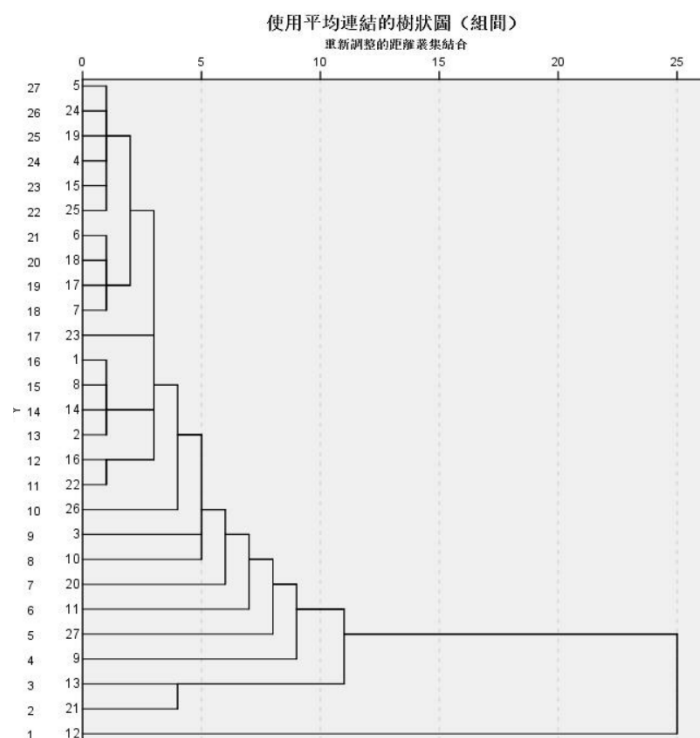


图 2：27 个红葡萄样本的聚类分析结果

依据上图结果，我们自行将 27 个红葡萄样本分为五类，其中，样本 12 为一类，样本 13、21 为一类，样本 9 为一类，样本 27 为一类，剩余样本为一类。

利用同样的方法，将 28 个白葡萄样本进行聚类分析，结果如下图所示：

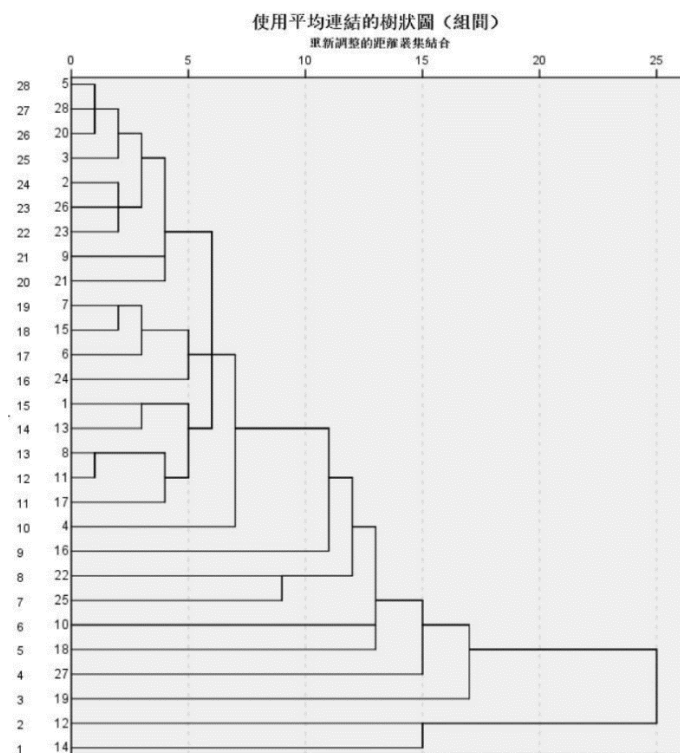


图 3：28 个白葡萄酒样本的聚类分析结果

其中，样本 14 为一类，样本 12 为一类，样本 19 为一类，样本 27 为一类，剩余样本为一类。

### ③对酿酒葡萄进行分级

由于问题一中已求得第二组评酒员的评价结果更有可信度，因此我们根据第二组评酒员对葡萄酒样品的评价结果，以该结果得分作为葡萄酒的质量评判标准，对相应的酿酒葡萄进行分级。

表 5：酿酒葡萄的分级表

类别	27 个红葡萄样本			28 个白葡萄样本		
	样本	得分	评级	样本	得分	评级
第一类	12	68.3	V	14	77.1	II
第二类	13、21	70.5	IV	12	72.4	V
第三类	9	78.2	I	19	76.4	IV
第四类	27	71.5	III	27	77	III
第五类	样本 1~8、 10~11、14~20、 22~26	73.35	II	样本 1~11、 13、15~18、 20~26、28	83.37	I

## 5.2.2 结果分析

### ①红葡萄样本

对于所有的酿酒红葡萄样本，具体分级情况为：I 级：样本 9；II 级：样本 1~8、10~11、14~20、22~26；III 级：样本 27；IV 级：样本 13、21；V 级：样本 12。其中样本 9 的评级最高，样本 12 的评级最低。

### ②白葡萄样本

对于所有的酿酒白葡萄样本，具体分级情况为：I 级：样本 1~11、13、15~18、20~26、28；II 级：样本 14；III 级：样本 27；IV 级：样本 19；V 级：样本 12。其中样本 1~11、13、15~18、20~26、28 的评级最高，样本 13、21 的评级最低。

## 5.3 基于葡萄酒与酿酒葡萄主要理化指标的相关性分析模型

### 5.3.1 求解过程

#### ①对葡萄酒的理化指标进行主成分分析

利用问题二中的主成分分析法对附件二、三中红葡萄酒的一级理化指标进行

处理，最后得到 33 个红葡萄酒的主要指标：花色苷、总酚、酒总黄酮、DPPH 半抑制体积（IV50）、丙酸乙酯、2,3-乙酰基丙酮、正十一烷、庚酸乙酯、乙酸、2-壬醇、壬酸乙酯、5-甲基糠醛、癸酸甲酯、3,7-二甲基-1,5,7-辛三烯-3-醇、二甘醇单乙醚、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、(R)-3,7-二甲基-6-辛烯醇、2-苯乙基乙酸酯、十二酸乙酯、2-吡咯烷酮、2-苯氧基-1-丙醇、4-甲基-1,1'-联苯、4-乙烯基-2-甲氧基-苯酚、十六烷酸乙酯、2-癸酸、甘油、乙基氢酸、2,3-二氢苯并呋喃、苯甲酸、正十二烷酸、异山梨糖醇。

采用相同的方法对白葡萄酒的一级理化指标进行处理，最后可得 32 个白葡萄酒的主要指标：L\*(D65)、乙酸乙酯、乙醇、丙酸乙酯、3-甲基-1-丁醇-乙酸酯、己酸乙酯、苯乙烯、乙酸己酯、2-辛酮、(Z)-3-己烯醇乙酸酯、2-己烯酸乙酯、乳酸乙酯、1-己醇、乙酸庚酯、辛酸甲酯、辛酸乙酯、乙酸辛酯、2-乙基-1-己醇、辛酸丙酯、2-壬醇、壬酸乙酯、5-甲基糠醛、癸酸乙酯、辛酸 3-甲基丁酯、反式-4-癸烯酸乙酯、3-甲硫基-1-丙醇、十二酸乙酯、苯乙醇、2-吡咯烷酮、4-乙烯基-2-甲氧基-苯酚、十六烷酸乙酯、2,5-二(1,1-二甲基乙基)-1,4-苯二醇。

## ②对葡萄酒和酿酒葡萄理化指标的相关性分析

将表 4 中所列的红葡萄的 45 个主要指标及②中红葡萄酒的 33 个主要指标分别按序编号，并建立矩阵对其进行相关性分析，用  $a_i$  表示酿酒葡萄的第  $i$  个指标， $i = 1, 2, \dots, 45$ ；用  $b_j$  表示葡萄酒的第  $j$  个指标， $j = 1, 2, \dots, 33$ 。（由于数量庞大，下列矩阵仅显示葡萄酒的及酿酒葡萄的各 10 个指标，完整矩阵见附录）。

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_{10}$
$a_{17}$	-0.14	0.88	0.75	0.93	0.52	0.61	0.76	0.54	0.73	0.88
$a_{18}$	0.33	-0.09	-0.08	-0.07	-0.05	-0.05	-0.02	-0.10	-0.03	-0.13
$a_{19}$	0.40	-0.07	-0.09	-0.05	-0.08	-0.09	-0.08	-0.09	-0.10	-0.01
$a_{20}$	-0.15	-0.28	-0.27	-0.23	-0.17	-0.24	-0.30	-0.10	-0.34	-0.30
$a_{21}$	0.50	-0.16	-0.15	-0.14	-0.16	-0.14	-0.15	-0.07	-0.13	-0.09
$a_{22}$	-0.46	0.43	0.36	0.49	0.37	0.22	0.37	0.22	0.35	0.48
$a_{23}$	-0.52	0.34	0.33	0.29	0.34	0.23	0.33	0.22	0.29	0.34
$a_{24}$	-0.44	0.69	0.65	0.64	0.54	0.55	0.67	0.46	0.60	0.61
$a_{25}$	0.11	0.17	0.12	0.26	0.09	0.03	0.11	0.06	0.11	0.23
$a_{26}$	0.41	-0.02	-0.07	0.08	-0.07	-0.09	-0.10	-0.07	-0.13	0.01

利用相同的方法对白葡萄酒及酿酒白葡萄的指标进行相关性分析。

### ③显著性检验—— $t$ 检验

利用 MATLAB 软件对上述数值进行  $t$  检验，所有结果均可通过  $t$  检验，结果合理，具有说服力（检验具体结果见附录）。

#### 5.3.2 结果分析

分析矩阵可知，在红葡萄酒与红葡萄的指标中，有 114 对指标相关系数介于 0.5~0.8 之间，为高度相关，有 31 对指标相关系数大于 0.8，为非常相关；有 8 对指标相关系数介于 0.5~0.8 之间，为高度相关，有 1 对指标相关系数大于 0.8，为非常相关，具体信息见下表：

表 6：红、白葡萄酒与红、白葡萄的指标相关性

相关性	指标（葡萄~葡萄酒）
高度相关	乙醛、乙酸甲酯、3-甲基丁醛~正十二烷酸、异山梨糖醇；乙酸乙酯、乙酸-2-甲基丙、3-甲基-1-丁醇、1-戊醇、乙酸己酯、4-己烯-1-醇-乙酸盐、1-己醇、单宁、干物质含量~异山梨糖醇；乙醇、丙酸乙酯、2-戊酮-正十二烷酸；甲苯~DPPH 半抑制体积（IV50）；2-辛酮~酒总黄酮、丙酸乙酯、2,3-乙酰基丙酮、正十一烷、庚酸乙酯、乙酸、5-甲基糠醛、十二酸乙酯、2-吡咯烷酮、4-乙烯基-2-甲氧基-苯酚、十六烷酸乙酯、甘油、乙基氢酸；辛酸乙酯~癸酸甲酯、二甘醇单乙醚、癸酸乙酯、2-苯乙基乙酸酯；1-辛稀-3-醇~花色苷；6-甲基-5-庚烯-2-醇~总酚、酒总黄酮、DPPH 半抑制体积（IV50）、丙酸乙酯、2,3-乙酰基丙酮、正十一烷、乙酸、2-壬醇、癸酸甲酯、二甘醇单乙醚、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、(R)-3,7-二甲基-6-辛烯醇、2-苯乙基乙酸酯、十二酸乙酯、4-乙烯基-2-甲氧基-苯酚、十六烷酸乙酯、乙基氢酸；甲氧基苯基丙酮肟~总酚、DPPH 半抑制体积（IV50）、2-壬醇、癸酸甲酯、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、(R)-3,7-二甲基-6-辛烯醇、2-苯氧基-1-丙醇、2-癸酸；苯乙醇~酒总黄酮、丙酸乙酯、2,3-乙酰基丙酮、正十一烷、庚酸乙酯、乙酸、5-甲基糠醛、二甘醇单乙醚、(R)-3,7-二甲基-6-辛烯醇、十二酸乙酯、2-吡咯烷酮、2-苯氧基-1-丙醇、4-甲基-1,1'-联苯、4-乙烯基-2-甲氧基-苯酚、十六烷酸乙酯、甘油、乙基氢酸；蛋白质~总酚、DPPH 半抑制体积（IV50）、2-壬醇、癸酸甲酯、癸酸乙酯、丁二酸二乙酯、(R)

	-3, 7-二甲基-6-辛烯醇、2-苯乙基乙酸酯、总酚—酒总黄酮、丙酸乙酯、2, 3-乙酰基丙酮、正十一烷、庚酸乙酯、乙酸、5-甲基糠醛、二甘醇单乙醚、(R) -3, 7-二甲基-6-辛烯醇、十二酸乙酯、2-吡咯烷酮、4-乙基-2-甲氧基-苯酚、十六烷酸乙酯、甘油、乙基氢酸；(E)-3, 7-二甲基-2, 6-辛二烯-1-醇~DPPH 半抑制体积 (IV50)、2-壬醇、癸酸甲酯、癸酸乙酯、丁二酸二乙酯、2-苯乙基乙酸酯。
非常相关	乙酸戊酯~异山梨糖醇；2-辛酮~总酚、DPPH 半抑制体积 (IV50)、2-壬醇、癸酸甲酯、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、(R) -3, 7-二甲基-6-辛烯醇、2-苯乙基乙酸酯、2-苯氧基-1-丙醇、2-癸酸；苯乙醇~总酚、DPPH 半抑制体积 (IV50)、2-壬醇、癸酸甲酯、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、2-苯乙基乙酸酯、2-癸酸；总酚~总酚、DPPH 半抑制体积 (IV50)、2-壬醇、癸酸甲酯、癸酸乙酯、2-甲基己酸、丁二酸二乙酯、2-苯乙基乙酸酯、2-苯氧基-1-丙醇、2-癸酸。
相关性	指标（葡萄~葡萄酒）
高度相关	甲苯~4-乙基-2-甲氧基-苯酚；乙酸辛酯~十二酸乙酯；2-乙基-1-己醇~2-壬醇、壬酸乙酯、4-乙基-2-甲氧基-苯酚；3, 7-二甲基-2, 6-辛二烯酸甲酯~3-甲硫基-1-丙醇、十二酸乙酯；(E)-2-己烯醛~十二酸乙酯。
非常相关	乙酸甲酯~2, 5-二(1, 1-二甲基乙基)-1, 4-苯二醇

## 5.4 多元线性逐步回归模型与 DEA 模型

### 5.4.1 利用多元线性逐步回归模型分析酒与葡萄的指标对酒质量的影响

#### ①多元线性回归方程求解

将附件一中评酒员对葡萄酒的十项评分标准按序排列，作为因变量  $y_i$  ,  $i = 1, 2, \dots, 10$ ；将葡萄酒与葡萄的理化指标仍按照问题三中的顺序排列，作为自变量  $x_j$  ,  $j = 1, 2, \dots, 45$ 。以酿酒红葡萄的理化指标对葡萄酒质量的影响为例，通过 *SPSS Statistics 22* 软件，对酿酒红葡萄的指标按照十项标准进行求解，得到的结果整理如下表所示：

表 7：酿酒红葡萄对葡萄酒质量的影响结果表

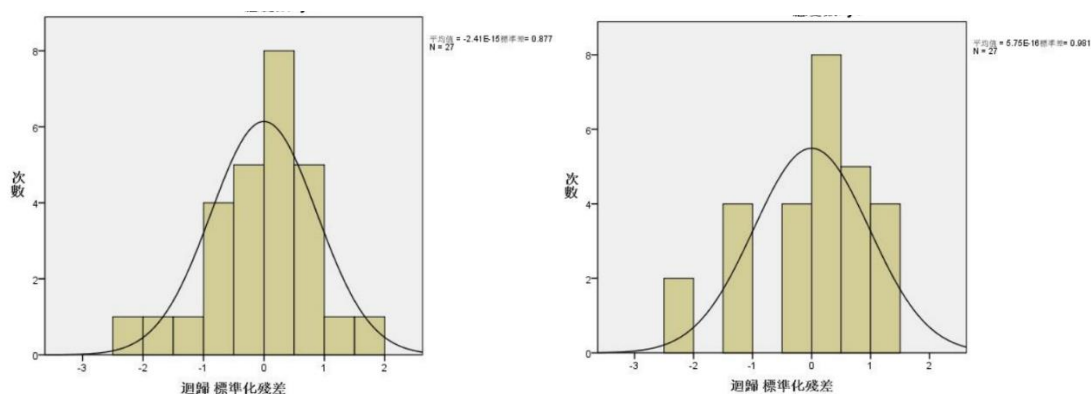
因变量	自变量	标准化系数	多元线性回归方程
-----	-----	-------	----------

$y_1$	常数、 $x_{11}$ 、 $x_{34}$ 、 $x_{38}$ 、 $x_{41}$ 、 $x_5$ 、 $x_{45}$	0.404、1.142、- 0.993、0.781、0.496、 -0.614、-0.283	$y_1=1.142x_{11}-0.993x_{34}+0.781x_{38}$ $+0.496x_{41}-0.614x_5-0.283x_{45}$ $+0.404$
$y_2$	$x_{45}$	-0.349	$y_2=-0.349x_{45}$
$y_3$	常数、 $x_4$ 、 $x_{44}$	-0.473、-0.456、-0.368	$y_3=-0.456x_4-0.368x_{44}-0.473$
$y_4$	常数、 $x_{35}$ 、 $x_{33}$ 、 $x_{44}$ 、 $x_{28}$ 、 $x_{23}$	0.396、0.958、- 0.858、-0.525、- 0.537、0.337	$y_4=0.958x_{35}-0.858x_{33}-0.525x_{44}-$ $0.537x_{28}+0.337x_{23}+0.396$
$y_5$	常数、 $x_{44}$ 、 $x_{37}$ 、 $x_{41}$ 、 $x_{45}$ 、 $x_2$ 、 $x_{22}$ 、 $x_{32}$ 、 $x_{39}$ 、 $x_{40}$	-0.515、-0.89、 0.874、-0.467、-0.3、- 0.889、0.653、 -0.649、1.419、-1.323	$y_5=-0.89x_{44}+0.874x_{37}-0.467x_{41}$ $-0.3x_{45}-0.889x_2+0.653x_{22}$ $-0.649x_{32}+1.419x_{39}-1.323x_{40}-$ $0.515$
$y_6$	常数、 $x_{35}$ 、 $x_{41}$ 、 $x_{43}$	0.492、0.527、- 0.508、0.318	$y_6=0.527x_{35}-0.508x_{41}+0.318x_{43}$ $+0.492$
$y_7$	$x_{35}$	0.521	$y_7=0.521x_{35}$
$y_8$	$x_{37}$	0.436	$y_8=0.436x_{37}$
$y_9$	常数、 $x_{37}$ 、 $x_{11}$	0.435、0.473、0.359	$y_9=0.473x_{37}+0.359x_{11}+0.435$
$y_{10}$	$x_{37}$	0.479	$y_{10}=0.479x_{37}$

酿酒白葡萄、红葡萄酒、白葡萄酒的理化指标对葡萄酒质量的影响也可由相同的方法计算得到。

## ②残差分析

利用 *SPSS Statistics 22* 软件，我们得到了酿酒红葡萄对葡萄酒质量的影响结果残差图（下图仅包含对葡萄酒质量的前四个标准的残差分析图）：



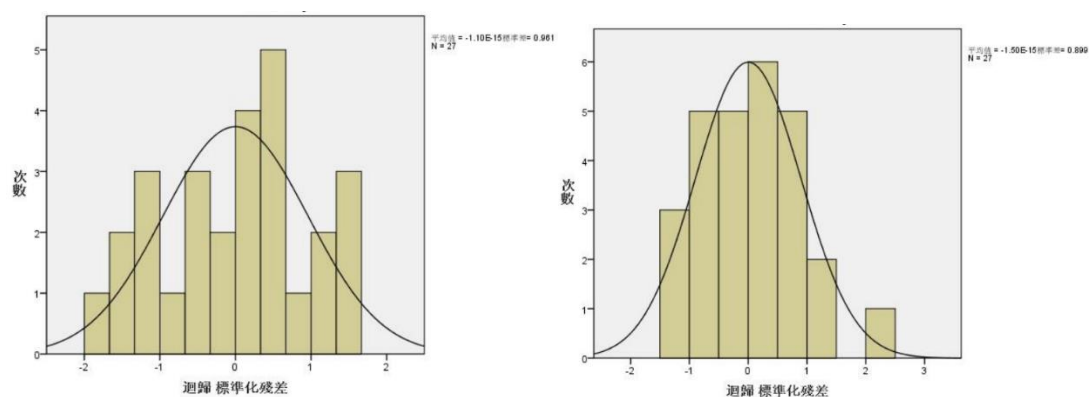


图 4：残差图

由残差图可知，以上结果具有说服力。

### 5.4.2 基于 DEA 模型论证能否用酒与葡萄的指标评价酒的质量

在该模型中，我们用综合效率作为衡量标准，设定  $m$  种类型不同的“输入量”（即葡萄酒与葡萄的理化指标），以及  $s$  种不同类型的“输出量”（即葡萄酒的质量评判标准）。因此，从实际情况考虑，若葡萄酒与葡萄的理化指标可以评价葡萄酒的质量，那么输入量对输出量的影响越大，两者的综合效率应越趋近于“1”。

通过 MaxDEA 软件，以红葡萄为例分析，将 27 个红葡萄样本的理化指标作为输入量，对应样本的红葡萄酒质量作为输出量，求解两者之间的综合效率，具体结果如下表所示：

表 8：红葡萄样本的理化指标与红葡萄酒质量的综合效率表

样本	1	...	16	17~18	19	20~21	22	23~24	25	26~27
综合效率	1	1	0.997	1	0.954	1	0.940	1	0.989	1

由上表可知，27 个红葡萄样本与红葡萄酒质量的综合效率都趋近于“1”。

### 5.4.3 结果分析

#### ①对于红葡萄酒的质量

影响红葡萄酒外观中的澄清度的 8 个主要因素依次为：葡萄中的 2-甲基-1-丙醇，葡萄酒中的酒总黄酮，葡萄中的 DPPH 自由基、(Z)-3,7-二甲基-2,6-辛二烯醛，葡萄酒中的总酚，葡萄中的乙醇、苹果酸、柠檬烯。其中第 1、4、5、7 个指标为有利因素，其余为不利因素。

影响红葡萄酒外观中的色调的 2 个主要因素依次为：葡萄酒中的庚酸乙酯，葡萄中的柠檬烯。其中前者为有利因素，后者为不利因素。

影响红葡萄酒香气中的纯正度的 5 个主要因素依次为：葡萄酒中的丙酸乙酯、2-甲基己酸，葡萄中的 3-甲基丁醛、百粒质量，葡萄酒中的乙酸。其中第 2 个指标为有利因素，其余为不利因素。

影响红葡萄酒香气中的浓度的 5 个主要因素依次为：葡萄中的总酚、花色苷、萜、百粒质量、1-辛烯-3-醇。其中第 1、5 个指标为有利因素，其余为不利因素。

影响红葡萄酒香气中的质量的 10 个主要因素依次为：葡萄的 (R)-3,7-二甲基-6-辛烯醇、(E)-3,7-二甲基-2,6-辛二烯-1-醇、百粒质量、乙酸甲酯、葡萄总黄酮、辛酸乙酯、蛋白质，葡萄酒中的庚酸乙酯，葡萄中的苹果酸、柠檬烯。其中第 1、5、6、8 个指标为有利因素，其余为不利因素。

影响红葡萄酒口感中的纯正度的 7 个主要因素依次为：葡萄酒中的酒总黄酮、十二酸乙酯，葡萄中的总酚、果酸，葡萄酒中的 4-乙烯基-2-甲氧基-苯酚，葡萄中的干物质含量，葡萄酒中的癸酸乙酯。其中第 1、3、6、7 个指标为有利因素，其余为不利因素。

影响红葡萄酒口感中的浓度的 4 个主要因素依次为：葡萄中的总酚，葡萄酒中的酒总黄酮、4-乙烯基-2-甲氧基-苯酚、2-苯乙基乙酸酯。其中第 1、2 个指标为有利指标，其余为不利指标。

影响红葡萄酒口感中的持久性的 5 个主要因素依次为：葡萄酒中的 2,3-二氢苯并呋喃，葡萄中的葡萄总黄酮，葡萄酒中的二甘醇单乙醚、总酚、3,7-二甲基-1,5,7-辛三烯-3-醇。其中第一个指标为不利因素，其余为有利因素。

影响红葡萄酒口感中的质量的 4 个主要因素依次为：葡萄酒中的 4-乙烯基-2-甲氧基-苯酚，葡萄中的葡萄总黄酮，葡萄酒中的总酚，葡萄中的 2-甲基-1-丙醇。其中第 1 个指标为不利因素，其余为有利因素。

影响红葡萄酒的平衡/整体评价的 3 个主要因素依次为：葡萄酒中的 2,3-二氢苯并呋喃，葡萄中的葡萄总黄酮，葡萄酒中的酒总黄酮。其中第 1 个指标为不利因素，其余为有利因素。

## ②对于白葡萄酒的质量

影响白葡萄酒外观中的澄清度的 2 个主要因素依次为：葡萄中的固酸比、 $b^*$  (+黄;-蓝)。这两个指标均为不利因素。

影响白葡萄酒外观中的色调的 1 个主要因素为：固酸比。该因素为不利因素。



影响白葡萄酒香气中的纯正度的 1 个主要因素为：3,7-二甲基-2,6-辛二烯酸甲酯。该因素为有利因素。

影响白葡萄酒口感中的持久性的 2 个主要因素依次为：葡萄中的(E)-2-己烯醛、邻二甲苯。两个因素都是不利因素。

影响白葡萄酒口感中的质量的 1 个主要因素为：葡萄中的 1-己醇。该因素为有利因素。

影响白葡萄酒的平衡/整体评价的 4 个主要因素依次为：葡萄中的乙醛、5-甲基糠醛、单宁、甲苯。其中第 3、4 个指标为有利因素，其余为不利因素。

对于白葡萄酒香气中的浓度、质量，口感中的纯正度、浓度这四个标准，没有主要的理化指标对它们产生影响。

通过 MaxDEA 软件的结果可知，两种葡萄酒及两种酿酒葡萄的理化指标与两种葡萄酒质量的综合效率均趋近于“1”，因此我们认为，可以用葡萄酒及酿酒葡萄的理化指标对葡萄酒质量进行评价。

## 六、模型评价

### 6.1 模型的优点：

- ①针对问题一的第二小问，我们利用变异系数分析法衡量两组评价结果的可信度，分别计算 10 个评分项目的变异系数，并将其整合，从而整体评价两组结果，与问题一的第一小问对 10 个评分项目逐个进行显著性检验有相似之处，逻辑连贯。
- ②针对问题二，我们通过主成分分析将数量繁多的理化指标减少为数量较少的主要理化指标，方便统计分析，为解决后面的问题做铺垫。
- ③针对问题四，我们运用 DEA 模型对各理化指标是否能评价葡萄酒的质量这一问题进行分析，用综合效率作为衡量指标，定量化分析评价的合理性，既直观又准确。

### 6.2 模型的缺点：

- ①忽略了附件 2 和附件 3 中二级量化指标的影响，有失准确性。
- ②问题四中的多元逐步回归的残差图呈正态分布，表明回归没有消除自相关。

## 七、参考文献

- [1] 阎克玉, 李兴波, 赵学亮, 张建荣, 阎洪洋. 河南烤烟理化指标间的相关性研究[J]. 郑州轻工业学院学报(自然科学版), 2000, 15(3): 21-23.
- [2] 阎克玉, 王建民, 屈剑波, 李兴波. 河南烤烟评吸质量与主要理化指标的相关分析[J]. 烟草科技, 2001, (10): 5-9.
- [3] 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, 2012, 18(3): 29-32.
- [4] 贺万华, 曹兴洪, 范康君, 刘京广, 周景秋, 黄胜, 张永川. 卷烟制丝和卷制过程中主要质量指标与消耗指标的关系及评价方法[J]. 中国烟草学报, 2007, 13(5): 17-22.
- [5] 薛超群, 王建伟, 奚家勤, 杨立均. 烤烟烟叶理化指标与浓香型风格程度的关系[J]. 烟草科技, 2012, (1): 52-56.
- [6] 薛超群, 尹启生, 王广山, 奚家勤, 张艳玲. 烤烟烟叶物理特性的变化及其与评吸质量的关系[J]. 烟草科技, 2008, (7): 52-55.
- [7] 李运, 李记明, 姜忠军. 统计分析在葡萄酒质量评价中的应用[J]. 酿酒科技, 2009, (4): 79-82.
- [8] 卓金武, MATLAB 在数学建模中的应用, 北京: 北京航空航天大学出版社, 2014. 9[M]
- [9] 姜启源, 数学模型, 北京: 高等教育出版社, 2011. 1[M]
- [10] 司守奎, 数学建模算法与应用, 北京: 国防工业出版社, 2011. 8[M]