葡萄酒的评价

摘　　要

现行的葡萄酒质量的评价体系是建立在人的感官感受上进行的，如何通过一些量化的理化指标来评价葡萄酒质量是一个值得研究的方向。

针对问题一，我们利用多元统计分析的相关知识，先对原始评分数据的分布进行了检验，进而通过差异性检验得出两组评论结果具有显著性差异、第二组评价结果更为可信的结论。接着对产生显著性差异的原因进行探讨，在对原始评价数据进行标准化处理后，得到两组评价结果无显著性差异的结论，确定了之前的显著性差异是来自品酒员个人因素的影响。

针对问题二，我们综合考虑了葡萄酒质量和葡萄理化指标与葡萄质量的相关性，以品酒员的感官评价为主、葡萄的理化指标为辅，采用逐步回归分析、聚类分析、判别分析的数学方法，建立了葡萄分级模型。利用此模型对酿酒红、白葡萄进行分类的结果显示，该模型与品酒员感官评价体系相似率分别为 88.9% 和 85.7% 。

针对问题三，由于葡萄的理化指标众多，我们先使用了相关系数矩阵确定了葡萄酒与葡萄理化指标中具有较大相关性的指标，从而实现了对葡萄理化指标的第一步筛选。接着利用逐步回归的方法拟合了葡萄酒理化指标与葡萄理化指标间一对多的函数关系，通过分析得到葡萄酒理化指标与葡萄理化指标之间具有较强相关性的结论。

针对问题四，首先要求分析葡萄和葡萄酒理化指标对葡萄酒质量的影响，我们通过分析葡萄酒理化指标与葡萄酒质量的函数关系，并利用第三问的结论，说明了葡萄与葡萄酒的理化指标只在一定程度上对葡萄酒质量有影响。问题第二部分要求我们论证能否用理化指标来评价葡萄酒，我们得出的结论为：不能仅仅通过葡萄与葡萄酒的理化指标对葡萄酒质量进行评价。此外，我们充分利用附件三中提供的芳香物质数据，使用多元统计分析的数学方法，通过提取对香气与口感评分相关度较大的芳香物质、建立芳香物质与葡萄酒质量的函数关系，论证了使用芳香物质作为评价葡萄酒质量的参考是可行的。

关键字：葡萄酒评价　数据处理　多元统计分析　相关系数分析

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？

2 问题分析

这是一个关于大型数据处理与分析的问题。

问题一要求我们分析两组评酒员评价结果有无显著差异。在进行差异性检验之前必须先对数据服从的分布进行检验，从而选定合适的检验方法进行检验。

问题二要求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。由题意可知除了葡萄酒的质量对葡萄的分级有比较大的影响外，酿酒葡萄的理化指标在一定程度上也会影响葡萄的质量。问题意在让我们建立一个综合葡萄酒质量与酿酒葡萄理化指标综合影响葡萄分级的模型。

问题三要求分析酿酒葡萄与葡萄酒理化指标之间的联系。由于酿酒葡萄理化指标众多，在分析两者的联系之前需要对葡萄的理化指标进行筛选。

问题四要求分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响，并论证能否用葡萄和葡萄酒的理化指标来评价葡萄酒的质量。难点在于对附件三葡萄酒和葡萄芳香物质数据的使用。

3 假设与符号

3.1 问题假设

1).假设品酒员给出的葡萄酒评价能够准确反映葡萄酒的质量；
2).假设附件三中芳香物质数据的单位不一定相同；
3).假设现有的评价体系能够准确反映葡萄酒的质量；

3.2 符号说明

- $P_{i,j}$ .....第 j 号品酒员对第 i 号酒样的评分
- y.....葡萄酒各个理化指标(一级)
- x.....酿酒葡萄各个理化指标(一级)
- S.....品酒员对葡萄或者葡萄酒的综合评分
- d.....度量酿酒葡萄与分级标准的“距离”
- w.....葡萄或葡萄酒的芳香物质

4 问题一的解答

4.1 问题一的分析

通过对附件一的数据进行观察，可知葡萄酒样品的评价项目满分为100，葡萄酒外观、香气、口感以及平衡/整体评价在其中各占一定的比例。由本题题意和附件数据分析得知，有两组品酒员，分别对27种红葡萄酒和28种白葡萄酒进行了品尝和评分。对同样的27类红葡萄酒样品，两组品酒员的打分不同，白葡萄酒亦然。为了判别两组评酒员的评价结果有无显著性差异，首先需要对每种酒的最终得分进行数据分布检验，进而选择合理的检验方式对差异进行检验。为判别两组评酒员的评价结果可信程度，选用离差平方和

4.2 数据预处理

在对附件一数据进行观察分析时，发现了某些错误的数据。例如：

(a)第一组白葡萄酒7号品酒员对3号酒的持久性评价数据为“77”，为超出其上限；

(b)第一组白葡萄酒9号品酒员对1号酒的持久度评价数据为“16”，超出其上限；

(c)第一组红葡萄酒4号品酒员对20号酒的色调评价分数为空缺。

为了减少异常数据对分析结果的影响，对异常数据做如下处理：取这一样酒评价结果中其他评酒员针对该酒该项目评分的均值，替代异常数据。

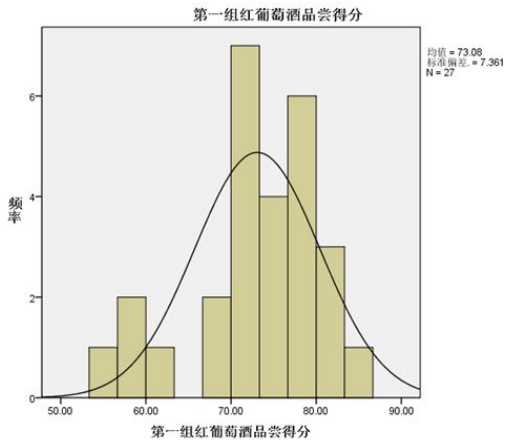
4.3 模型的建立与求解

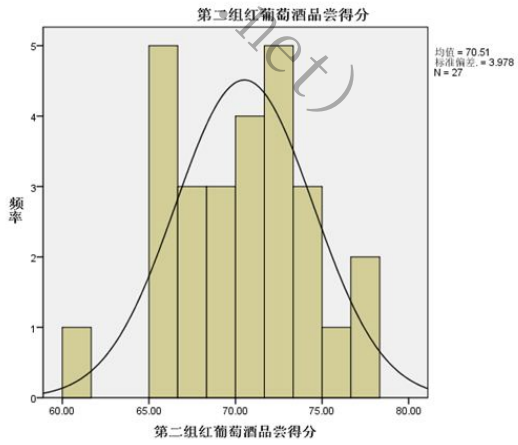
4.3.1 问题一的第一部分：显著性差异的检验

绘制两组频数分布图以进行分布初步分析

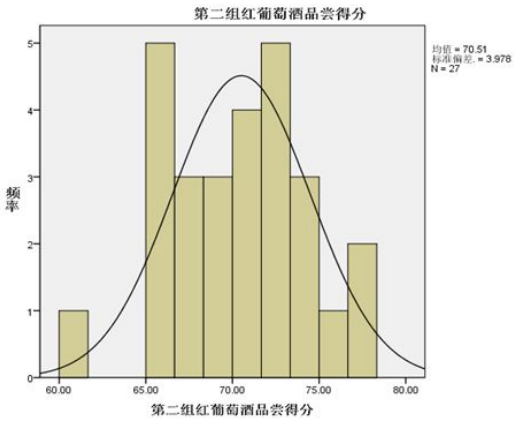
由以上数据初步得到葡萄酒质量评价数值后，对葡萄酒质量及其对应的酒样品数目分布进行分析。通过软件绘制出红、白葡萄酒在两组评酒员评价结果下的评分分布直方图：

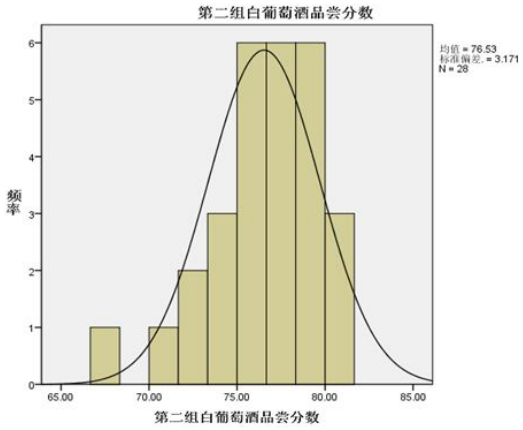
得到第一组红酒与白酒的评数分布图





得到第二组红酒与白酒的评数分布图



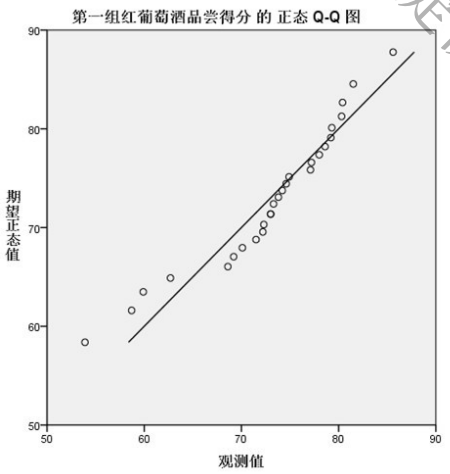


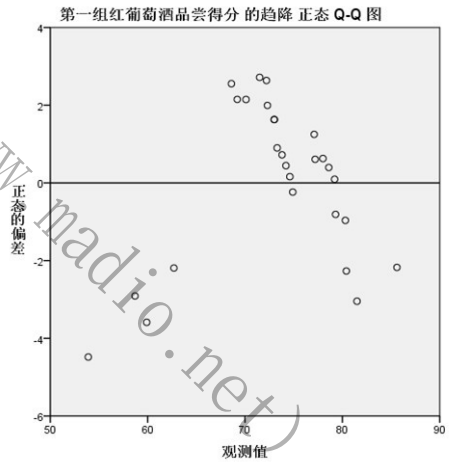
结论：通过对图像观察分析，可以大致预测葡萄酒样品质量及其对应的数量分布呈正态分布。接下来进行数据分布的正态性检验。

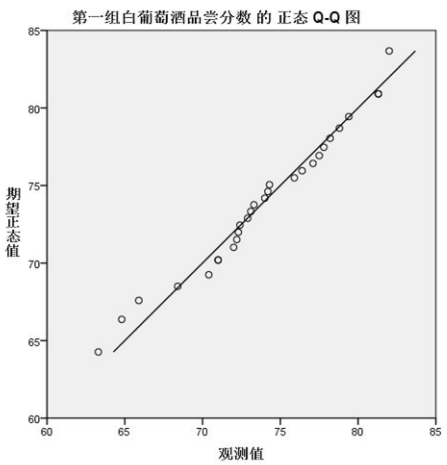
正态分布检验

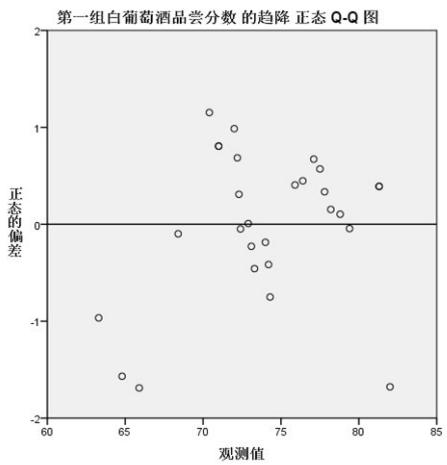
为了对这一预测进行更精确的验证，下面分别使用统计软件spss中提供的Q-Q 图检测对数据进行正态性分布检验。

QQ图检验：









由图所示，QQ图中，a图中离散情况较为一般，并不严格离散在分布线上；b 图中大量数据集中在0.5-1.0之间，虽无任何数据超过1.0，但集中在0.5之内的数据同样不足。于是得出结论，葡萄酒样感官分析评价不符合正态分布，所以接下来使用非参数检验来确定其是否有显著性差异。

非参数检验：

参数检验是在总体分布形式已知的情况下，对总体分布的参数如均值、方差等进行推断的方法。但是，在数据分析过程中，无法对总体分布形态作简单假定，此时参数检验的方法就不再适用了。非参数检验正是一类基于这种考虑，利用样本数据对总体分布形态等进行推断的方法。

秩和检验：对配对比较的资料应采用符合秩和检验，其基本思想是：若检验假设成立，则差值的总体分布应是对称的，故正负秩和相差不应悬殊。

检验的基本步骤为：

(1)建立假设；

H_0：差值的总体中位数为0；

H_1：差值的总体中位数不为0；检验水准为0.05。

(2)算出各对值的代数差；

(3)根据差值的绝对值大小编秩；

(4)将秩次冠以正负号，计算正、负秩和；

(5)用不为“0”的对子数n及T（任取T+或T-）查检验界值表得到P值作出判断。

白葡萄酒秩和检验结果：

假设检验汇总

	原假设	测试	Sig.	决策者
1	第一组白葡萄酒品尝分数 与 第二组白葡萄酒品尝分数 之间差异的中位数等于 0。	相关样本 Wilcoxon 符号秩检验	.016	拒绝原假设。

显示渐进显著性。显著性水平是 .05。

红葡萄酒秩和检验结果：

假设检验汇总

	原假设	测试	Sig.	决策者
1	第一组红葡萄酒品尝得分 与 第二组红葡萄酒品尝得分 之间差异的中位数等于 0。	相关样本 Wilcoxon 符号秩检验	.011	拒绝原假设。

显示渐进显著性。显著性水平是 .05。

由上秩和检验结果可知，红葡萄酒、白葡萄酒的两组评价结果具有显著性差异。

模型改进:使用标准化数据进行分析

数据预处理：

对附件一中数据再分析，很明显看到对葡萄酒的质量评定为品酒员的感官评定，这种评价为典型的人为评定，所以将会不可避免的产生误差。

因此我们需将数据进行标准化处理.

设P_{ij}为第i号酒样被第j品酒员评价的分数,P'_{ij}为标准化后的评价分数，$\overline{P_j}$为第j号品酒员的平均打分，σ_j为第j 号品酒员打分的方差。

标准化公式：

$$P'_{ij} = \frac{P_{ij} - \overline{P_j}}{\sigma_j}$$

标准化后各项打分详见附件。

显著性检验 标准化后，评分数据满足均值为0正态分布，于是接下来采用参数检验其是否显著性差异，使用SPSS进行f检验，我们得到以下检验结果：

表 1 显著性检验

	平方和	df	均方	F	显著性
组间	0.000	1	0.000	0.000	1.000
组内	28.144	52	0.541		
总数	28.144	53			

上述结果显示，两组数据无显著性差异。

由上文分析可知，对原始评分数据进行标准化处理后，原来具有显著性差异的两组评价结果变得没有显著性差异，原因分析如下：即使是专业的评酒员在对葡萄酒质量进行评价时，难免掺杂个人的主观因素在其中。由于个人评分习惯在评分起始点、评分区间等方面会有些许差别，我们分析正是由于评酒员的这些个人因素导致了两组评价结果的显著性差异。在对数据进行过标准化处理过后，相当于消除了评酒员个人主观因素的影响。

因此处理后的两组评价结果没有显著性的差异。

4.3.2 问题一的第二部分：评价结果可靠性判断

由上文可知，当两组数据未经标准化处理的情况下，两组数据在非参数检验中显示出差异性显著。在此基础上判断两组数据的可信性。

对两组数据的可靠性进行分析，在葡萄酒的感官评价中，品酒员的评价尺度等方面的差异，导致不同品酒员对同一酒样的评价差异很大，从而不能真实的反应酒样的差异。

于是，我们假设P_{ij}为第j号评酒员对第i号酒样的评价分数，每一酒样的评价误差程度为：

$$\varepsilon_{1i} = \frac{1}{10}\sum_{j=1}^{10}(P_{ij} - \overline{P_i})^2 \quad i = 1, 2...27$$

此处的ε_{1i}为10名品酒员对第i号酒样品评价离散程度的衡量指标。

$$\varepsilon_{2j}=\frac{1}{27}\sum_{i=1}^{27}(P_{ij}-\overline{P_j})^2 \quad j=1,2...10$$

此处的ε_{2j}为第j名品酒员对第$1-10$号酒样品评价离散程度的衡量指标。

我们有理由认为，当不同专家对同一样品酒的评分差距小时，说明该评分是能被大多数人所接受的。当一名专家对不同品质的样本酒给出的评价差异较明显时，我们更能接受该专家的评分。

综上所述，我们建立我们的可靠性评价指标：

1.同一样品酒的得分离散度ε_1应越小越好.

2.同一品酒员对不同品质的酒的评分离散度ε_2应越大越好。

可以得到：

$$\varepsilon_1=\frac{1}{27}\sum_{i=1}^{27}\varepsilon_{1i}$$

表 2

ε_1	红葡萄酒	白葡萄酒
第一组	52.4063	106.1114
第二组	30.41148	50.4175

结论：第二组的样品酒的离散程度更小，第二组对样品酒一致性更高。

$$\varepsilon_2=\frac{1}{10}\sum_{j=1}^{10}\varepsilon_{2j}$$

表 3

ε_2	第一组	第二组
红葡萄酒	52.17654	15.2383
白葡萄酒	22.4449	9.695753

结论：第一组的品酒员分数离散度更大，第一组品酒员更具备区分好对各酒样品的能力。

综上所述，两组在可靠性评价指标上各具有一定优势。

一方面，由于世界上红酒品种共计上百种，而附件中酒样本品种仅有28组，所以存在酒样本过少导致的品酒员对不同酒样品的区分明显。另一方面，正如前面已知，若将附件一中数据标准化，两组数据没有显著差别。由此可见，对每种酒的质量而言，专家的打分离散程度更重要，更具有参考意义。最终，由于第二组的样品酒的离散程度更小，第二组对样品酒一致性更高，第二组分数更加可信。

模型改进：可靠性评价指标

通过建立新的可靠性评价指标，品酒员评分离差指标。

$$\sigma(P)=\frac{1}{27}\sum_{i=1}^{27}(P_{ij}-\overline{P_i})^2$$

其现实意义为，每位品酒员将其评价分数减去同组成员对该品种酒的均值，得到其每次打分的偏度，最后将其偏离程度统一想加后求均值，得到每组品酒员的打分能力以及其对每组酒的平均分数的贡献程度，在一定程度上概括了信度评价的两重标准。

表 4 可信指标

	$\sigma(P_1)$	$\sigma(P_2)$	$\sigma(P_3)$	$\sigma(P_4)$	$\sigma(P_5)$	
第一组	43.70111	39.36037	69.96037	64.81222	83.79	
第二组	13.17501	17.51279	58.98923	50.84849	57.37945	
	$\sigma(P_6)$	$\sigma(P_7)$	$\sigma(P_8)$	$\sigma(P_9)$	$\sigma(P_10)$	均值
第一组	53.76778	23.18259	34.58259	64.23444	46.67148	52.4063
第二组	15.49975	21.07486	30.07901	17.81204	11.86597	29.42366

由新可信指标分析得出，第二组数据更可信。

5 问题二的解答

5.1 问题分析

问题二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。

由常识可以知道，葡萄酒的质量很大程度上取决于酿酒葡萄的质量，优质的葡萄酒对应优质的酿酒葡萄，劣质的葡萄酒对应的酿酒葡萄质量也相应较差，因此我们考虑利用附件一中所给的葡萄酒质量评分作为参考标准对葡萄进行分级。

同时，葡萄酒的质量也并不是完全取决于酿酒葡萄。根据题意我们知道，葡萄酒和酿酒葡萄的理化指标会在一定程度上反映葡萄和葡萄酒的质量，因此我们对酿酒葡萄进行分级的同时加入酿酒葡萄的理化指标作为参考。

5.2 数据处理

附件二中提供的酿酒葡萄的理化指标一共有30种(一级指标)，经过初步观察我们得知，固酸比指标值=可溶性固形物指标/可滴定酸指标，可以认为固酸比指标能够反映后两者的综合影响，因此我们选择剔除可溶性固形物与可滴定酸指标。

特别地，我们发现附件二中存在某些异常的数据值，例如酿酒葡萄理化指标中白葡萄百粒质量的第三次检测值为2226.1g，而前两次检测值分别为225.8g 和224.6g，为了避免异常数据对数据分析产生的影响，我们利用前两次的检测值均值代替异常数据。

由于不同的理化指标数值大小不同、数据波动范围不同，为了消除这些因素的影响，

我们先对理化指标数据进行标准化处理。公式如下：

$$x_i' = \frac{x_i - \overline{x}}{\sigma_x}$$

x 为某种理化指标原始数据，σ_x 为该理化指标标准差。

5.3 模型的建立与求解

模型思想

利用第一问中得出的标准化处理后的葡萄酒评分，我们建立葡萄的理化指标与葡萄酒评分之间的函数关系，通过该函数关系得出某一酿酒葡萄酿出的葡萄酒的预期评分；接着利用葡萄酒得分对葡萄酒进行聚类分析，划分出所给葡萄酒样品的等级，进而通过分析酿酒葡萄所酿葡萄酒预期得分与葡萄酒样品各等级之间的“距离”来对酿酒葡萄进行分级。从而既考虑到了所酿葡萄酒评分对葡萄分级的影响，又考虑到了酿酒葡萄的理化指标对葡萄分级的影响。

模型建立与求解

由于酿酒葡萄的理化指标(一级)多达30种，全部给予考虑将使计算十分繁琐，也不能较清晰地判断出主要影响酿酒葡萄质量的理化指标，因此我们对酿酒葡萄的理化指标进行筛选。

利用多元统计分析中逐步回归的思想，把葡萄酒评分作为因变量，对应的酿酒葡萄理化指标作为自变量，将酿酒葡萄的理化指标逐个加入到函数中进行拟合，若相应的统计量是检验显著的则保留该变量，检验不显著则剔除该变量。

该方法能够筛选出对葡萄酒评分有显著影响的酿酒葡萄理化指标，同时能够有效地减少酿酒葡萄理化指标之间的多重共线性。

令 y 为标准化处理后的葡萄酒评分，$x_1, x_2, ..., x_{30}$ 分别对应酿酒葡萄的30种理化指标(一级)，建立如下形式的函数：

$$y = B_0 + B_1 x_1 + B_2 x_2 + ... + B_p x_p + \varepsilon$$

其中B_0,B_1,B_2和B_p为待估参数，ε为残差。

由x对y进行逐步回归，得到结果如下：

红葡萄酒：$R - squared = 0.96$

$$y = -0.097x_2 + 0.054x_3 + 0.153x_4 - 0.302x_6 - 0.236x_8 + 0.319x_{10} + 0.212x_{13} - 0.150x_{14} + 0.266x_{17} + 0.287x_{21}0.148x_{22} - 0.091x_{24} - 0.137x_{26} + 0.161x_{27} - 0.144x_{29}$$

白葡萄酒：$R - squared = 0.87$

$$y = 0.131x_1 - 0.296x_2 + 0.192x_7 + 0.439x_9 + 0.164x_{10} + 0.914x_{11} + 0.156x_{12} - 0.922x_{13} + 0.117x_{14} - 0.406x_{15} - 0.641x_{17} - 0.240x_{21} + 0.732x_{22} - 0.184x_{23} + 0.373x_{24} + 0.177x_{29}$$

通过R-squared可知拟合程度较好，说明所选理化指标能够解释大部分葡萄酒评分的变化。

因此将以下酿酒葡萄理化指标作为影响葡萄酒评分的主要指标：

红葡萄：

蛋白质、VC、花色苷、苹果酸、多酚氧化酶活力、DPPH 自由基、酒总黄酮、白藜芦醇、还原糖、固酸比、干物质含量、百粒质量、出汁率、果皮质量、色泽(a)；

白葡萄：

氨基酸总量、蛋白质、柠檬酸、褐变度、DPPH自由基、总酚、单宁、酒总黄酮、白藜芦醇、黄酮醇、还原糖、固酸比、干物质含量、果穗质量、百粒质量、色泽(a)。

利用上述函数关系，可以获得各个酿酒葡萄样本所酿葡萄酒的预期评分。

接下来对各葡萄酒的得分进行聚类分析。这里我们采用对样本进行聚类的Q型聚类，利用软件SPSS进行聚类，再根据各类平均得分高低进行分级，结果如下表：

表 5 红葡萄酒样本分级

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

表 6 白葡萄酒样本分级

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

考虑分级的合理性，我们将红白两种葡萄酒各分为四级(一级最优、四级最劣)，在各级别内，对各葡萄酒样品的评分求均值得到该评级葡萄酒的基准分。结果如下：

表 7 葡萄酒分类标准

评级基准分 / 葡萄酒	一级	二级	三级	四级
红葡萄酒	1.139	0.338	-0.538	-1.559
白葡萄酒	0.533	-0.08	-0.652	-1.646

通过上文建立的酿酒葡萄理化指标与葡萄酒评分的函数关系，得到各个葡萄样本所酿葡萄酒的预期评分，我们参考几何学中欧氏空间距离的概念，利用预期评分与葡萄酒

各评级基准分之间的“距离”来判断酿酒葡萄与各个等级的“远近关系”，进而实现对酿酒葡萄的分级。“距离”的计算公式如下：

$$d_{i,j} = \sqrt{[y(i) - y(j)]^2}$$

其中 $d_{i,j}$ 为第i个葡萄样本所酿葡萄酒预期评分与第j级葡萄酒基准分之间的“距离”，y_i 为第i个葡萄样本所酿葡萄酒预期评分，y_j 为第j级葡萄酒基准分。$d_{i,j}$ 越小，代表该葡萄样本越接近此评级，选择“距离”最小的评级，即为该样本酿酒葡萄所对应的评级。

经判别，各葡萄样本分级结果如下：

表 8 红葡萄样本分级

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

表 9 白葡萄样本分级

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

5.4 模型的检验

经过分析，该分级模型的误差可能来自以下几个方面：

(1).利用逐步回归建立函数关系时产生的误差：回归分析的目的是提高拟合程度，因此为了提高拟合程度有可能将过多的理化指标变量选入函数中进行拟合。虽然显示的拟合程度很高，但由于样本量个数不大，引入过多的变量将对模型效果产生影响；

(2).对葡萄酒聚类和分级时产生的误差：附件中提供的葡萄酒样品质量并没有覆盖所有可能的范围，在此基础上直接对葡萄酒进行分级难免欠缺妥当；并且红葡萄酒样本与白葡萄酒样本质量未必处于同一个档次，两种葡萄酒样本都分成四级可能会产生误差。

检验思路：评价酿酒葡萄好坏的最主要因素是所酿葡萄酒的质量好坏，因此在大样本的基础上，酿酒葡萄的分级与葡萄酒的分级不应该有显著性的差异，据此可以检验该分级模型的有效性。

可知红葡萄酒与红葡萄有 24 个样本即 88.9% 的样本评级相同，白葡萄酒与白葡萄有 24 个样本即 85.7% 的样本评级相同，可知该评价模型效果较好。

5.5 模型的评价与改进

该分级模型从葡萄酒与酿酒葡萄质量的直接关系和酿酒葡萄理化指标对酿酒葡萄质

量两方面因素进行了考虑，首先通过建立函数关系确立了酿酒葡萄理化指标对葡萄酒质量的影响，接着利用聚类分析对葡萄酒进行分级，进而通过判别分析对酿酒葡萄的等级进行划分，经检验结果较好。

该分级模型也有需要进行改进的地方。如在对葡萄酒进行聚类分析时分级标准的确定略显主观，没有考虑到红葡萄酒与白葡萄酒之间质量的差异；同时在影响葡萄酒质量的主要理化指标的筛选上也有改进的空间，可以通过各理化指标之间的相关关系与理化指标和葡萄酒质量之间的相关关系进一步筛选理化指标。

6 问题三的解答

6.1 问题分析

问题三要求我们分析酿酒葡萄与葡萄酒的理化指标之间的联系，由于葡萄酒与酿酒葡萄有多个理化指标，因此简单的两指标间相关分析不再适用。分析可知酿酒葡萄的理化指标影响了葡萄酒的理化指标，它们之间并不是互相影响而是一种因果关系，因此考虑建立模型，描述一个葡萄酒理化指标与酿酒葡萄的多个理化指标之间的联系，通过这种联系分析酿酒葡萄理化指标对葡萄酒理化指标的影响。

根据附件二可知酿酒葡萄理化指标数量较多，而样本量较小，取过多的酿酒葡萄指标进行分析难免产生较大的误差，因此必须先对酿酒葡萄的理化指标进行筛选。

6.2 模型的建立与求解

数据预处理

考虑到各个理化指标的指标值之间绝对值大小程度、离散程度等的不同，为了避免这些因素对分析的结果产生影响，先对各个理化指标数据进行标准化处理。

酿酒葡萄理化指标的筛选

由问题分析可知，由于样本量较小，将过多的酿酒葡萄理化指标纳入考虑范围可能产生较大的误差，因此必须对酿酒葡萄的理化指标进行筛选。

以红葡萄酒及红葡萄样本为例进行说明。

记红葡萄酒经过标准化处理后各个理化指标为：$y_1, y_2, y_3, ..., y_9$，红葡萄样本各个理化指标为 $x_1, x_2, x_3, ..., x_{30}$。

为了建立单个葡萄酒理化指标与多个酿酒葡萄理化指标之间的函数关系，利用这多个酿酒葡萄理化指标与某个葡萄酒理化指标之间的相关程度对酿酒葡萄理化指标进行筛选：相关程度大代表两个理化指标之间有比较密切的联系，反之则代表联系不够紧密，可以考虑将其排除在考虑范围。

利用 Matlab 软件计算每个葡萄酒的理化指标与每个酿酒葡萄理化指标之间的相关系数，生成相关矩阵：

考虑将相关系数绝对值大于均值的理化指标纳入考虑范围，进而可以得到针对葡萄酒的每一个理化指标，与其有比较大相关性的酿酒葡萄的理化指标。

筛选后结果如下，其中 X 代表两变量间相关度不大。

表 10 葡萄与葡萄酒理化指标相关矩阵

	花色苷	单宁	总酚	酒总黄酮	白藜芦醇	DPPH	L	a	b
氨基酸总量	X	0.21	0.05	X	0.05	0.11	X	X	0.07
蛋白质	0.01	0.19	0.15	0.15	X	0.10	0.20	X	X
VC	X	X	X	X	X	X	X	X	0.08
花色苷	0.64	0.43	0.49	0.42	X	0.38	0.55	0.06	X
酒石酸	X	X	X	X	X	X	X	X	0.18
苹果酸	0.41	0.01	0.07	X	X	X	0.06	0.27	0.02
柠檬酸	0.09	X	X	X	X	X	X	X	X
多酚氧化酶活力	0.19	X	X	X	X	X	0.12	X	X
褐变度	0.48	0.16	0.17	0.16	X	0.09	0.28	0.05	X
DPPH自由基	0.27	0.46	0.52	0.47	0.13	0.49	0.42	X	X
总酚	0.33	0.53	0.59	0.60	0.17	0.59	0.47	X	X
单宁	0.37	0.43	0.46	0.41	0.03	0.41	0.39	X	X
总黄酮	0.15	0.40	0.53	0.54	0.28	0.53	0.32	X	X
白藜芦醇	X	X	X	X	X	X	X	0.16	X
黄酮醇	0.12	0.29	0.12	0.01	X	0.14	0.24	X	X
总糖	X	0.03	X	X	X	X	X	X	0.09
还原糖	X	X	X	X	X	X	X	X	0.28
可溶性固形物	X	0.12	X	X	X	0.03	X	X	X
PH	X	X	X	X	X	X	X	X	X
可滴定酸	X	X	X	X	X	X	X	X	X
固酸比	0.03	X	X	0.04	X	X	X	0.15	X
干物质含量	X	0.13	0.01	X	X	0.04	X	X	0.11
果穗质量	X	X	X	X	X	X	X	X	X
百粒质量	X	0.04	X	X	X	X	0.02	X	X
果梗比	0.21	0.19	0.12	0.01	X	0.05	0.19	X	X
出汁率	0.04	0.07	0.11	0.20	X	0.14	0.15	X	X
果皮质量	X	X	X	X	X	X	X	0.05	X
L	0.08	0.18	0.16	0.06	X	0.11	0.21	X	X
a	0.08	0.01	0.00	X	X	0.01	0.31	0.26	X
b	X	X	X	X	X	X	0.06	0.34	X

针对酿酒葡萄的这些理化指标，利用逐步回归法建立其与葡萄酒各理化指标之间的函数关系。

利用软件 $Eviews$ 进行逐步回归，结果如下：

红葡萄酒：

$$
\begin{aligned}
y_1 =& 0.441x_4 - 0.085x_{26} - 0.085x_{29} + 0.209x_{12} + 0.461x_6 + 0.201x_{28} + 0.122x_{15} \\
&+ 0.057x_7 + 0.306x_{10} - 0.199x_{13} - 0.104x_{25} \\
& R - squared = 0.965 \\
y_2 =& 0.347x_9 + 0.360x_1 + 0.396x_{18} + 0.360x_{10} - 0.250x_{16} + 0.138x_{22} - 0.081x_{25} \\
&+ 0.059x_{24} \\
& R - squared = 0.932 \\
y_3 =& 0.386x_{13} + 0.215x_4 + 0.358x_1 - 0.183x_{26} + 0.789x_{10} - 0.189x_2 + 0.111x_{22} \\
&+ 0.335x_9 - 0.156x_{25} - 0.195x_{15} - 0.262x_{11} + 0.072x_{12} \\
& R - squared = 0.955 \\
y_4 =& 0.671x_{11} + 0.160x_9 - 0.155x_{25} + 0.136x_{28} + 0.194x_{13} + 0.144x_4 \\
& R - squared = 0.832 \\
y_5 =& 0.951x_{13} + 0.334x_1 - 0.479x_{11} \\
& R - squared = 0.431 \\
y_6 =& -0.303x_2 + 0.413x_1 + 0.375x_9 + 0.249x_{13} - 0.316x_{25} + 0.817x_{10} + 0.077x_{18} \\
&+ 0.086x_{28} - 0.055x_{29} \\
& R - squared = 0.927 \\
y_7 =& -0.280x_4 + 1.100x_{29} - 0.221x_{15} - 0.638x_{10} - 0.666x_{30} + 0.274x_{24} - 0.216x_{28} \\
&+ 0.206x_{13} + 0.415x_{25} + 0.102x_2 \\
& R - squared = 0.955 \\
y_8 =& -0.917x_{29} - 0.527x_4 - 0.177x_{21} + 0.226x_{14} - 0.123x_9 \\
& R - squared = 0.697 \\
y_9 =& 0.540x_{17} - 0.383x_6 + 0.324x_5 - 0.137x_3 \\
& R - squared = 0.626
\end{aligned}
$$

白葡萄酒：

$$
\begin{aligned}
y_1 =& 0.572x_{12} + 0.194x_{27} + 0.717x_{15} + 0.420x_{16} + 0.169x_1 + 0.169x_1 - 1.668x_{11} \\
&+ 1.111x_{13} - 0.415x_{22} - 0.274x_{25} + 0.320x_{26} + 0.135x_{10} + 0.162x_7 - 0.193x_8 \\
& R - squared = 0.803 \\
y_2 =& 0.540x_{13} + 0.305x_{18} + 0.201x_1 + 0.171x_{27} - 0.156x_4 - 0.119x_{26} \\
& R - squared = 0.675 \\
y_3 =& 0.249x_{11} + 0.168x_2 + 0.262x_6 + 0.332x_{15} - 0.247x_{25} \\
& R - squared = 0.742 \\
y_4 =& -0.348x_{16} - 0.254x_6 - 0.161x_5 \\
& R - squared = 0.242 \\
y_5 =& 1.291x_{13} + 0.435x_{16} + 0.507x_3 - 0.250x_8 - 0.960x_{11} + 0.309x_{15} \\
& R - squared = 0.573 \\
y_6 =& -0.690x_{22} + 0.254x_{26} - 0.345x_{14} + 0.402x_{17} + 1.128x_{20} + 1.078x_{21} - 0.520x_{18} \\
&+ 0.162x_{19} \\
& R - squared = 0.821 \\
y_7 =& 0.486x_{24} - 0.387x_{19} - 0.697x_{25} + 0.513x_{26} - 0.219x_7 - 0.219x_{30} + 0.126x_2 \\
&+ 0.529x_5 - 0.510x_{27} - 0.255x_{22} + 0.324x_{18} + 0.161x_{13} \\
& R - squared = 0.798 \\
y_8 =& -0.756x_{26} - 0.796x_{24} + 0.251x_{22} - 1.812x_{20} + 0.313x_{27} + 0.288x_{16} + 0.449x_{23} \\
&- 1.866x_{21} + 0.123x_5 - 0.075x_{19} \\
& R - squared = 0.914
\end{aligned}
$$

由以上结果可知红葡萄酒理化指标中的白藜芦醇与酿酒葡萄理化指标的相关程度较弱，仅与蛋白质、总酚和总黄酮理化指标拟合有 43.1% 的拟合度；色泽(b) 理化指标与酿酒葡萄理化指标的相关程度一般，仅与VC含量、苹果酸和柠檬酸有 62.6% 的拟合程度；白葡萄酒理化指标中的白藜芦醇与酿酒葡萄理化指标的相关程度弱，仅与溜石酸、苹果酸和总糖有些许相关关系。而葡萄酒的其他理化指标，大多数能够与酿酒葡萄的某几种理化指标之间建立起函数关系，且拟合程度不错。

6.3 模型的评价与改进

模型的评价

本问中我们首先利用葡萄酒理化指标与酿酒葡萄理化指标之间相关矩阵，筛选出对葡萄酒某一理化指标相关程度较大的酿酒葡萄理化指标，接着通过逐步回归的方法建立了葡萄酒理化指标与酿酒葡萄理化指标之间的函数关系。可以看到，理化指标之间的大部分函数关系拟合效果都不错，能够反映理化指标之间的变化关系。

模型的改进

针对不同的理化指标之间，函数关系的形式可能不同，由于理化指标众多、时间有限等因素，我们没有进一步探讨理化指标之间可能的函数关系形势。可以通过对理化指标之间两两做散点图，大致判断两者的函数关系形势，进而再代入模型中进行拟合，也许会得到更好的解释效果。

7 问题四的解答

7.1 问题四的分析

问题四首先要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，并要求我们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。

经过第三问的分析与求解，我们得出的结论是：葡萄酒理化指标与酿酒葡萄的理化指标之间具有比较高度的相关性，从各回归方程的可决系数(R-squared)可以看出。并且，分析可知葡萄酒的理化指标对葡萄酒质量的影响更为直接，而酿酒葡萄的理化指标必须通过葡萄酒的理化指标来间接影响葡萄酒的质量，因此我们考虑分析葡萄酒的理化指标对葡萄酒质量的影响，进而利用葡萄理化指标与葡萄酒理化指标之间高度的相关性来分析葡萄的理化指标对葡萄酒质量的影响。

通过查阅文献与网络资料，我们得知葡萄中的芳香物质对所酿出的葡萄酒的气味、口感等方面有比较大的影响，初步分析可以通过葡萄酒或葡萄中的芳香物质来评价葡萄酒的质量。

同时，由于附件三中的芳香物质种类众多，必须对芳香物质进行筛选。葡萄酒的香气与口感占评分体系的比重较大，且通过文献资料可知芳香物质对香气与口感确实有比较大的影响，因此考虑利用芳香物质对香气分析、口感分析评分的相关程度作为筛选的标准。

7.2 模型的建立与求解

分析葡萄酒理化指标对葡萄酒质量的影响

以红葡萄酒为例，建立葡萄酒质量与葡萄酒理化指标之间的函数关系，分析葡萄酒理化指标对葡萄酒质量的影响程度。

随机抽取若干个葡萄酒理化指标，画出其对葡萄酒质量的散点图，可以看出大致上成线性关系，因此我们采用线性函数来描述葡萄酒理化指标对葡萄酒质量的影响。

使用软件 Eviews 拟合以下方程：

$$s = a_1y_1 + a_2y_2 + a_3y_3 + ... + a_9y_9$$

结果如下：

$$\begin{aligned} s = &-0.820y_1 + 0.417y_2 - 0.269y_3 + 0.302y_4 + 0.231y_5 - 0.188y_6 - 0.767y_7 - 0.144y_8 \\ &- 0.157y_9 + 2.47E - 16 \\ &R - squared = 0.588 \end{aligned}$$

表 11 红葡萄酒回归参数

自变量	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	C
Std.Error	0.396	0.351	0.512	0.329	0.183	0.527	0.429	0206	0.169	0.105
t-Statistic	-2.073	1.188	-0.525	0.917	1.258	-0.357	-1.787	-0.697	-0.927	2.36E-15
Prob.	0.054	0.251	0.606	0.372	0.226	0.726	0.092	0.495	0.367	1

由回归结果可知，虽然回归可决系数接近了0.6，但由于大部分自变量系数在统计上不显著，可以说明，红葡萄酒理化指标在一定程度上影响了红葡萄酒的质量，但影响有限，不能仅仅通过葡萄酒的理化指标来评价葡萄酒的质量。

同理，可得到白葡萄酒理化指标与白葡萄酒质量之间函数关系为：

$$s = 0.038y_1 - 0.011y_2 - 0.142y_3 - 0.075y_4 + 0.124y_5 - 0.170y_6 - 0.076y_7 - 0.141y_8 + 7.14E - 11$$

$$R - squared = 0.149$$

表 12 白葡萄酒回归参数

自变量	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	C
Std.Error	0.257	0.282	0.151	0.127	0.175	0.464	0.280	0.588	0.112
t-Statistic	0.146	-0.037	-0.936	-0.588	0.708	-0.366	-0.271	-0.239	6.35E-10
Prob.	0.885	0.971	0.361	0.564	0.488	0.718	0.789	0.814	1

可知对白葡萄酒来说，葡萄酒理化指标对葡萄酒质量的拟合程度很低，大多数自变量系数统计不显著。

综上可知，葡萄酒理化指标对葡萄酒质量有一定的影响，但是不能仅仅依靠葡萄酒的理化指标对葡萄酒的质量作出评价。由于葡萄的理化指标与葡萄酒的理化指标之间高度的相关性，可以推断葡萄的理化指标对葡萄酒质量的影响程度与葡萄酒的理化指标基本一致。

建立葡萄酒芳香物质对葡萄酒质量影响的函数关系

由前文的分析与论证可知，单纯使用葡萄的理化指标或者葡萄酒的理化指标对葡萄酒质量进行评定是不够合理的。

通过查阅文献与网络资料，我们得知葡萄与葡萄酒中的芳香物质对葡萄酒的香气、口感等方面有一定的影响，因此考虑利用附件三中提供的葡萄酒的芳香物质数据来建立芳香物质对葡萄酒质量的函数关系。

在分析葡萄酒芳香物质对葡萄酒质量的影响时，根据附件一的资料，我们将葡萄酒评价指标细分为外观分析、香气分析、口感分析与平衡/整体评价，根据相关文献资料的研究，芳香物质主要影响葡萄酒的香气和口感，且根据附件一的评价体系，香气分析和口感分析的比重占到74%，可以比较全面地反映葡萄酒的大体质量。

下面以红葡萄酒为例进行说明评价体系的构建。

1).对附件一中的评分数据分别按照外观分析、香气分析、口感分析和平衡/整体评价做数据的标准化处理，对芳香物质同样进行标准化处理。分别计算四方面评分与葡萄酒各芳香物质数据的相关系数，得到相关系数矩阵(见附录)，矩阵中第 i 行第 j 列元素表示芳香物质 i 对 j 方面评分的相关系数。

2).相关系数矩阵数据处理：对所有相关系数求均值，每个相关系数减去均值，得到处理后的相关系数矩阵。接着对每一种芳香物质进行检验，若该种芳香物质对香气分析评分与口感分析评分的相关系数都大于 0，则将该芳香物质纳入考虑范围，剔除其他不满足条件的芳香物质。

最终筛选出来的芳香物质为：乙酸乙酯、乙酸正丙酯、丁酸乙酯、柠檬烯、正十三烷、乙酸庚酯、乙酸辛酯、2-乙基-1-己醇、辛酸丙酯、3,7-二甲基-1,6-辛二烯-3-醇、5-甲基糠醛、二甘醇单乙醚、辛酸3-甲基丁酯、2-甲基己酸、丁二酸二乙酯、反式-4-癸烯酸乙酯、3-甲硫基-1- 丙醇、辛酸、7-甲氧基-2,2,4,8-四甲基-三环[5.3.1.0(4,11)]十一碳烷、2,3-二氢苯并呋喃。

3).以经过筛选的葡萄酒芳香物质为自变量，葡萄酒总评分为因变量建立线性函数关系，得到函数：

$$s = -0.205w_{60} + 0.380w_{29} + 0.680w_5 - 0.260w_{36} + 0.532w_{41} - 0.433w_{53} - 0.935w_{43} + 0.390w_{38} - 0.678w_{34} - 0.127w_{28} + 0.557w_{45} + 0.501w_7 - 0.736w_2 - 0.290w_{42} - 0.242w_{12} + 0.112w_{31}$$

$$R - squared = 0.917$$

表 13 红葡萄酒回归参数

自变量	w_{60}	w_{29}	w_5	w_{36}	w_{41}	w_{53}	w_{43}	w_{38}
Prob	0.082	0.011	0.001	0.063	0.013	0.004	0.002	0.116
自变量	w_{34}	w_{28}	w_{45}	w_7	w_2	w_{42}	w_{12}	w_{31}
Prob	0.005	0.286	0.005	0.008	0.005	0.060	0.079	0.318

可知芳香物质对总评分的拟合程度相当高，因此葡萄酒芳香物质可以作为评价葡萄酒质量的参考因素。

同理，对白葡萄酒做同样的数据处理与分析，经过筛选后的芳香物质为：乙醇、丁酸乙酯、3- 甲基-1-丁醇-乙酸酯、乙酸庚酯、辛酸乙酯、香叶基乙醚、辛酸丙酯、3,7-二甲基-1,6-辛二烯-3-醇、二甘醇单乙醚、辛酸3-甲基丁酯、丁二酸二乙酯、2-苯乙基乙酸酯、7-甲氧基-2,2,4,8-四甲基-三环[5.3.1.0(4,11)]十一碳烷、4-乙烯基-2-甲氧基-苯酚。

得到的回归方程为：

$$s = 1.774w_{40} + 2.683w_{29} + 7.497w_{37} + 6.772w_{48} + 0.960w_{35} - 0.318w_8 + 2.309w_6 - 0.356w_{21} + 34.455w_{19} + 0.177w_3 - 22.574w_{26}$$

$$R - squared = 0.741$$

表 14 白葡萄酒回归参数

自变量	w_{40}	w_{29}	w_{37}	w_{48}	w_{35}	w_8
Prob	0.098	0.002	0.027	0.028	0.280	0.093
自变量	w_6	w_{21}	w_{19}	w_3	w_{26}	
Prob	0.059	0.051	0.007	0.142	0.257	

分析可知白葡萄酒的芳香物质对白葡萄酒的质量也具有显著性的影响，因此可以用葡萄酒的芳香物质来作为评价葡萄酒质量的参考。

7.3 模型的评价与改进

模型的评价

优点：本模型在处理芳香物质这一大数据的方法上，通过相关矩阵分析和处理，提取评价项目中对质量影响最大－－－－香气、口感评价对指标进行了筛选，既在理论上符合了芳香物质在葡萄酒中的化学作用，又能在数学上有效避免了在芳香物质与质量关系的逐步回归中由于指标过多而导致结论不准确的情况。

其次，本模型最后逐步回归的结果中，数据相关性程度高，R^2在70%以上。在采用本模型的数据处理方法前，我们尝试了聚类分析对数据进行筛选，筛选后的数据回归模型拟合程度很低，均在50%一下。可见，本模型的筛选指标方法是比较可行且有专业依据的。

缺点：在芳香物质与葡萄酒质量的逐步回归中数据的相关程度虽然较高，但其中一些芳香物质指标的显著性较低，在5%的置信度中没法通过检验（由解答中的eviews回归结果可直观看出），最后归回模型的可信度有待提高。

模型的改进

①本题在模型的建立中，由于默认了第三问的结论，认为酿酒葡萄与葡萄酒理化指标有比较强的相关性，因此在第四问中只采用数据量较少的葡萄酒理化指标与质量进行回归分析，可以再对酿酒葡萄的理化指标与质量进行回归分析，得出结论，与已有结论相互验证。由于筛选酿酒葡萄理化指标较多，若想采用逐步回归分析的方法，首先可以参照第三问中相关矩阵筛选法对理化指标进行筛选，具体过程不再赘述。以红葡萄为例得到回归结果如下：

$$s = 0.899y_7 - 0.260y_3 + 0.370y_{12} + 0.225y_{11} - 0.591y_5 + 0.210y_9 - 0.169y_1$$

$y_1 \sim y_7$分别为红葡萄理化指标中的蛋白质、苹果酸、总酚、总黄酮、PH、固酸比、果梗比。

检验统计显著数据如下:

表 15 检验

自变量	y7	y3	y12	y11	y5	y9	y1
Prob	0.0016	0.0859	0.0038	0.0856	0.0856	0.1639	0.246
R-aquard	0.70921						
Adjusted R-squared	0.6219						

可见酿酒葡萄理化指标对葡萄酒质量统计显著性也很差，进一步说明纯粹用理化指标不能评价葡萄酒质量。

②本模型在得出“葡萄酒理化指标对葡萄酒质量有一定的影响，但是不能仅仅依靠葡萄酒的理化指标对葡萄酒的质量作出评价。”的结论后，可以采用联合回归检验的方法，进一步论证理化指标与芳香物质在统一回归中对葡萄酒质量影响程度的大小。具体方法是将理化指标和芳香物质都作为自变量进行筛选，对葡萄酒质量进行逐步回归。观察回归结果中所引用的自变量分属于芳香物质和理化指标的多少来检验这两种指标的重要性。

参考文献

[1] 杨小平，统计分析方法与SPSS应用教程[M]，北京：清华大学出版社，2008年。

[2] 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，第4期：2009年。

[3] 朱建平，应用多元统计分析[M]，北京，科学出版社，2006年。

附录

1.程序1

```
%本程序用于分解附件一中每种酒的四种评价项目，并计算其该项目的得分
clear
clc
tic
Y=[];     %该矩阵中填入excel处理后的10种酒得分矩阵，红葡萄酒为（270,10），白葡萄酒为（280,10
[m,n]=size(Y);
for i=0:m/10-1
    for j=1:n
        A(i+1,j)=Y(i*10+1,j)+Y(i*10+2,j);
        B(i+1,j)=Y(i*10+3,j)+Y(i*10+4,j)+Y(i*10+5,j);
        C(i+1,j)=Y(i*10+6,j)+Y(i*10+7,j)+Y(i*10+8,j)+Y(i*10+9,j);
        D(i+1,j)=Y(i*10+10,j);
    end
end
A
B
C
D
toc
```

2.程序2

```
%本程序用于计算第三问、第四问中使用到的相关矩阵
%输出相关矩阵以及中解题方法所需要的处理后的矩阵，
%Y中有a项指标，X中有m项指标，输出一个[m,a]的矩阵
%每列对应的是X的各项指标对于Y中一项指标的相关性
clear
clc
tic
Y=[];
X=[];

[a,b]=size(Y);
[m,n]=size(X);
for i=1:a
    for j=1:m
       A= corrcoef(Y(i,:),X(j,:));
       C(i,j)=A(1,2);
```

```
    end
end
S=abs(C);
SUM=sum(sum(S));
[h,v]=size(C);
MEAN=SUM/(h*v);
for i=1:h
    for j=1:v
        B(i,j)=S(i,j)-MEAN;
    end
end
C=C'          %相关矩阵
S=S'          %相关矩阵取绝对值
B=B'          %均对值相关矩阵减去均值后

toc
```

3.程序3

```
%本程序用于矩阵 行向量 的归一化处理
clear
clc
tic
load x.txt % 导入需要处理的矩阵
[m,n]=size(x);
for i=1:m
    A(i,1)=max(x(i,:));
    A(i,2)=min(x(i,:));
end

for i=1:m
    for j=1:n
        x(i,j)=(x(i,j)-A(i,2))/(A(i,1)-A(i,2));
    end
end
toc
```

4.程序4

```
%本程序用于矩阵 列向量 的归一化处理
clear
clc
```

```
tic
load x.txt % 导入需要处理的矩阵
[m,n]=size(x);
for i=1:n
    A(1,i)=max(x(:,i));
    A(2,i)=min(x(:,i));
end

for j=1:n
    for i=1:m
        x(i,j)=(x(i,j)-A(2,j))/(A(1,j)-A(2,j));
    end
end
toc
```

5.对葡萄酒评分的标准化处理(以第一组红葡萄酒为例)

品酒员1 品酒员2 品酒员3 品酒员4 品酒员5 品酒员6 品酒员7 品酒员8 品酒员9 品酒员10 平均分μ

红酒1	-2.32	-0.82	-1.77	-1.71	0.17	-1.12	0.02	-1.22	-0.79	-2.64	-1.22
红酒2	0.24	0.53	0.91	1.11	1.26	0.61	1.46	0.65	0.89	-0.43	0.72
红酒3	1.39	0.89	1.13	1.39	-0.45	1.42	0.15	1.07	0.74	0.17	0.79
红酒4	-2.20	-1.00	-0.61	-0.02	-1.30	0.79	0.54	-1.01	0.58	0.37	-0.38
红酒5	0.62	-0.10	-0.10	-0.58	0.71	-0.93	-0.50	1.17	0.28	-0.83	-0.03
红酒6	0.37	-0.55	-0.17	-0.72	0.56	-0.39	-0.37	-0.91	0.28	1.78	-0.01
红酒7	-0.79	-0.46	0.19	-0.30	-1.22	0.97	0.02	-1.43	0.74	1.78	-0.05
红酒8	-0.66	0.08	-0.61	-0.16	0.09	-0.12	-0.37	1.27	-0.64	0.17	-0.09
红酒9	1.01	0.26	0.19	2.24	0.79	1.52	0.54	2.00	0.12	0.77	0.94
红酒10	-0.28	0.62	0.69	0.27	0.02	-0.03	0.41	-0.49	-0.49	-0.03	0.07
红酒11	0.49	-1.36	-0.10	-0.44	-0.91	-0.21	-0.24	-0.70	1.66	-0.43	-0.22
红酒12	-1.94	-2.99	-2.42	-1.57	-1.69	-1.21	-3.25	-1.22	-3.25	-1.23	-2.0
红酒13	-0.02	0.80	0.41	-1.00	-0.14	0.34	0.67	0.34	-0.64	0.37	0.11
红酒14	0.11	0.17	-0.25	0.55	0.40	-1.30	0.54	0.34	-0.49	0.17	0.03
红酒15	-0.02	-2.27	-1.70	-1.14	-1.84	-2.11	-2.07	-1.32	-1.87	0.17	-1.42
红酒16	0.37	0.44	0.48	0.69	-0.45	-0.21	1.06	0.13	-0.18	-0.23	0.21
红酒17	0.11	0.35	1.27	0.27	1.72	0.79	-0.37	0.75	0.28	0.17	0.53
红酒18	-0.79	-0.91	-1.77	-1.57	-1.77	-1.48	-1.29	-1.53	-1.41	-1.44	-1.3
红酒19	0.88	0.80	0.77	-0.02	-0.53	1.24	1.06	0.55	0.43	1.18	0.64
红酒20	1.13	0.80	0.19	1.11	0.56	0.52	0.54	0.34	1.04	1.18	0.74
红酒21	0.49	1.34	1.64	0.69	-0.45	-1.21	0.93	0.03	1.04	-0.23	0.43
红酒22	0.49	0.71	-0.10	0.27	1.41	-0.12	0.41	0.44	-0.03	0.97	0.45

红酒23 1.77 0.89 0.91 1.96 1.57 1.79 1.19 1.90 0.74 0.57 1.33
红酒24 0.11 0.89 1.20 0.27 1.18 0.97 -0.24 0.23 -0.18 -1.03 0.34
红酒25 -1.17 0.26 0.55 -0.58 -0.37 -0.57 -1.03 -1.12 0.28 -1.64 -0.54
红酒26 0.49 0.44 -0.17 -0.72 0.25 -0.21 0.02 0.34 -0.03 0.37 0.08
红酒27 0.11 0.17 -0.75 -0.30 0.40 0.25 0.15 -0.60 0.89 -0.03 0.03

6.红葡萄酒指标对葡萄指标的相关系数矩阵(其他相关系数矩阵的处理形式以此为例)

花色苷单宁总酚酒总黄酮白藜芦醇DPPH半抑制体积L a b
氨基酸总量0.11 0.50 0.34 0.20 0.33 0.40 -0.24 -0.10 0.36 1.00
蛋白质0.30 0.47 0.44 0.44 0.00 0.38 -0.48 -0.03 0.05 2.00
VC -0.09 -0.09 -0.13 -0.10 -0.03 -0.12 0.12 0.11 -0.37 3.00
花色苷0.92 0.72 0.77 0.71 0.20 0.67 -0.83 -0.35 -0.24 4.00
酒石酸0.03 0.28 0.27 0.16 0.22 0.24 -0.24 0.01 0.46 5.00
苹果酸0.69 0.30 0.35 0.27 -0.19 0.24 -0.35 -0.56 -0.31 6.00
柠檬酸0.38 0.15 0.14 -0.08 -0.20 0.02 -0.25 -0.27 -0.01 7.00
多酚氧化酶活力0.48 0.14 0.15 0.12 -0.13 0.07 -0.41 -0.01 0.10 8.00
褐变度0.77 0.45 0.46 0.44 -0.09 0.38 -0.56 -0.33 -0.24 9.00
DPPH自由基0.56 0.75 0.81 0.76 0.42 0.77 -0.71 -0.12 -0.05 10.00
总酚0.61 0.82 0.87 0.88 0.46 0.87 -0.75 -0.17 0.05 11.00
单宁0.66 0.72 0.74 0.70 0.31 0.70 -0.68 -0.09 -0.20 12.00
总黄酮0.44 0.68 0.82 0.82 0.57 0.81 -0.61 -0.07 0.05 13.00
白藜芦醇-0.03 0.05 0.08 0.05 0.01 0.07 0.16 -0.45 -0.11 14.00
黄酮醇0.41 0.58 0.41 0.30 0.07 0.42 -0.52 -0.05 0.22 15.00
总糖0.05 0.32 0.19 0.19 0.15 0.27 -0.06 -0.19 0.38 16.00
还原糖-0.07 0.09 -0.01 -0.02 0.00 0.08 0.02 -0.20 0.57 17.00
可溶性固形物0.19 0.41 0.24 0.25 0.01 0.31 -0.15 -0.18 0.25 18.00
PH -0.02 0.23 0.14 0.28 0.18 0.23 -0.13 -0.09 0.02 19.00
可滴定酸-0.22 -0.07 -0.12 -0.18 0.11 -0.06 0.20 0.26 0.07 20.00
固酸比0.32 0.24 0.24 0.32 -0.09 0.22 -0.25 -0.44 0.07 21.00
干物质含量0.23 0.42 0.30 0.25 0.08 0.33 -0.20 -0.25 0.39 22.00
果穗质量-0.10 -0.27 -0.18 -0.24 0.08 -0.20 0.02 0.22 -0.04 23.00
百粒质量-0.26 -0.33 -0.25 -0.25 -0.04 -0.23 0.31 0.15 -0.18 24.00
果梗比0.50 0.47 0.40 0.30 0.19 0.33 -0.47 -0.06 -0.09 25.00
出汁率0.33 0.36 0.40 0.48 0.26 0.42 -0.44 -0.01 -0.10 26.00
果皮质量-0.03 -0.08 -0.11 -0.11 -0.02 -0.04 -0.03 0.34 0.05 27.00
L -0.37 -0.46 -0.44 -0.34 -0.03 -0.39 0.49 0.05 -0.19 28.00
a -0.37 -0.30 -0.29 -0.28 -0.28 -0.30 0.59 -0.54 -0.06 29.00
b -0.13 -0.12 -0.08 -0.18 -0.25 -0.13 0.35 -0.63 0.03 30.00