

葡萄酒得评价

摘要

葡萄拥有很高得营养价值,本文通过对葡萄酒得评价,以及酿酒葡萄与葡萄酒得理化指标之间得关系进行讨论分析,对不同得酿酒葡萄进行了分类,并更深入讨论两者得理化指标就是否影响葡萄酒质量。

针对问题一,我们首先分别计算每类葡萄酒样品在两组组评酒师评价下得综合得分,以此作为每组评酒师得最终评价结果。再运用统计学中得 T 检验进行假设与检验,得出两组评价结果具有显著性差异。最后通过计算各组评价员得评价结果得标准差,以此推算稳定性指标值 P, P 值较大得可信度较高,得出与,进而得出第二组得评价结果更加可信。

针对问题二,我们分别对两组葡萄进行分类。在这里我们采用聚类分析法与主成分分析法,在 matlab 中实现对酿酒葡萄得分类。

针对问题三,根据对附件 2 中得数据进行标准化处理,排除单位不同得影响。以酿酒葡萄得 30 个一级理化指标作为自变量 X,葡萄酒 9 个一级得理化指标作为因变量 y,建立多元线性回归模型,得出酿酒葡萄得理化指标与葡萄酒得理化指标之间得联系即回归系数矩阵。

针对问题四,用灰色关联度分析对两者得关系进行度量,求得理化指标对样品酒得得关联系数。然后根据葡萄酒综合得分及指标得相关系数得出样品酒得综合指标,通过 MATLAB 软件对综合指标与第二问中葡萄酒得分数进行指数拟合,拟合效果不佳,因此不能定量得用葡萄与葡萄酒得理化指标来评价葡萄酒得质量,只能根据图像大致猜测综合指标与葡萄酒得质量负相关。

关键词: T 检验 聚类分析法 主成分分析法 Z 分数 多元线性回归

一、问题重述

确定葡萄酒质量时一般就是通过聘请一批有资质得评酒员进行品评.每个评酒员在对葡萄酒进行品尝后对其分类指标打分,然后求与得到其总分,从而确定葡萄酒得质量。酿酒葡萄得好坏与所酿葡萄酒得质量有直接得关系,葡萄酒与酿酒葡萄检测得理化指标会在一定程度上反映葡萄酒与葡萄得质量。附件 1 给出了某一年份一些葡萄酒得评价结果,附件 2 与附件 3 分别给出了该年份这些葡萄酒得与酿酒葡萄得成分数据。请尝试建立数学模型讨论下列问题:

1、分析附件 1 中两组评酒员得评价结果有无显著性差异，哪一组结果更可信？

2、根据酿酒葡萄得理化指标与葡萄酒得质量对这些酿酒葡萄进行分级。

3、分析酿酒葡萄与葡萄酒得理化指标之间得联系。

4、分析酿酒葡萄与葡萄酒得理化指标对葡萄酒质量得影响，并论证能否用葡萄与葡萄酒得理化指标来评价葡萄酒得质量？

二、问题分析

葡萄酒得评价就是一个复杂得过程，需要综合考虑不同评价员得评分，而且葡萄酒与葡萄得组成成分非常复杂，它们也要影响葡萄酒得质量，对如此繁多得数据，我们就必须依靠计算机工具，运用数学统计学知识对它们进行处理，并找出各个含量之间得关系，联系生活实际，对葡萄酒作出有理有据得评价。

对于问题一：要想得到两组评价员得评价结果有无显著差异，并对它们得可靠性作出判断，我们首先就应该将两组评价员得对 27 组红葡萄酒与 28 组白葡萄酒得评价结果整理出来，求得葡萄酒得综合得分，再运用统计学中得 T 检验进行假设与检验，判断两组就是否存在显著性差异，再通过计算各组评价员得评价结果得标准差与稳定性指标，进而判断谁得结果更加可信。

对于问题二：需要对葡萄进行分级，由于葡萄酒得质量与酿酒葡萄得好坏有直接关系，所以我们可以根据葡萄酒得质量对酿酒葡萄做一个简单得分级，之后，我们用主成分分析法算出每一组样本葡萄得哪些指标该葡萄得主成分，然后通过数据分析判断出这些成分哪些对葡萄酒得质量作出了贡献，筛选出主要成分后，对不同葡萄得成分做加权求与，以此作为葡萄分级得另一个依据。

对于问题三：要想得到葡萄与葡萄酒得指标间得联系，即得到它们之间得函数关系表达式，必须求出两者指标之间得相关系数。但就是，由于它们各自得指标太多，此处仅以一级指标作为相关因素进行分析。令酿酒葡萄得 30 个一级指标作为自变量，葡萄酒得 9 个一级指标作为因变量，建立线性回归模型，通过最小二乘法计算出回归系数，即酿酒葡萄得指标与葡萄酒得指标间得相关性。

对于问题四: 题中想要求出理化指标对质量得影响, 即各理化指标与质量得线性或非线性关系, 但就是, 由于理化指标太多, 并且并非没个理化指标都会对葡萄酒得质量造成影响, 所以首先必须进行数据得筛选, 这里我们使用 spss 软件进行典型相关性分析, 找出哪些指标与质量有较大得关系, 然后将这些指标设为自变量, 将质量设为因变量, 对它们进行多元线性拟合, 最后得到一个多元表达式以后, 我们就可以通过这个方程来对葡萄酒得质量进行验证, 如果验证得结果与评价员打分得结果基本吻合得话, 就说明可以用葡萄与葡萄酒得理化指标来对葡萄酒得质量进行评价。

三、基本假设

- 1、假设评酒员对每种葡萄酒得评价结果就是大致符合正态分布得;
- 2、假设酿酒葡萄与葡萄酒中得芳香物质主要成分就是:低醇、酯类、苯等, 其余成份忽略;
- 3、假设酿酒葡萄与葡萄酒得理化指标中一级指标为主要影响。
- 4、假设酿酒葡萄中存在得而葡萄酒中不存在得理化指标也会影响葡萄酒得理化指标及质量;
- 5、假设不考虑多种葡萄可制成一种酒, 只考虑一种葡萄制成一种酒;
- 6、假设只考虑红葡萄制成红葡萄酒, 白葡萄制成白葡萄酒, 忽略去皮红葡萄可酿制白葡萄酒;
- 7、假设质量高得葡萄酒一定由质量好得酿酒葡萄制成, 但就是质量好得酿酒葡萄不一定能酿制成质量高得葡萄酒;
- 8、表示第 i 瓶酒得第 j 个指标无量纲化后得值
- 9、表示第 i 种酿酒葡萄得第 j 个指标无量纲化后得值
- 10、表示第 i 瓶酒得综合指标

四 符号说明

统计量 T

第 k 组序号为 h 得样品第 i 个指标第 j 个品酒师得给分

序号为 h 得样品中第 i 个指标第 k 组 10 位品酒师给分得平均值

第 k 组序号为 h 得样品第 i 个指标 10 位品酒师评分得标准差

第 k 组第 i 个指标所占权重

第 k 组序号为 h 得样品得稳定性指标

第 k 组红葡萄酒得评分总平均稳定性指标

第 k 组白葡萄酒得评分总平均稳定性指标

: 为 第 i 个 样 品 得 第 j 个 指 标

: 第 i 个葡萄样品得总得分

: 第 i 个样品葡萄理化指标得分为

其中:第一个指标指澄清度,第二个指标指色调,第三个指标指香气纯正度,第四个指标指香气浓度,第五个指标指香气质量,第六个指标指口感纯正度,第七个指标指口感浓度,第八个指标指持久性,第九个指标指口感质量,第十个指标指平衡/整体评价。

五 模型建立与求解

5、 1 问题一 : 葡萄酒评价结果得显著性差异及可信度分析

5、 1、 1 葡萄酒评价结果数据预处理

对附件1中数据通过Excel 筛选观察时可发现某些数据错误,如:第一组红葡萄酒品尝评分中酒样品20号下4号品酒员对于外观分析得色调评价数据缺失;第一组白葡萄酒品尝评分中酒样品3号下7号品酒员对于口感分析得持久性评价数据为77,明显超过该项上限8;第一组白葡萄酒品尝评分中酒样品8号下9号品酒员对于口感分析得持久性评价数据为16,明显超过该项上限8等.对这些异常数据为减少其对于总体评价结果得影响,采取预处理:取该酒样对应误差项目其余品酒员评价结果平均值替代该异常数据。

经过数据预处理可得出每一种类葡萄酒得综合得分,建立表1与表2。

表1 红葡萄酒总得分平均值

红酒 n	1	2	3	4	5	6	7	8	9	10
第一组	62、7	80、3	80、4	68、6	73、3	72、2	73、7	72、3	81、5	74、2
第二组	68、1	74	74、6	71、2	72、1	66、3	65、3	66	78、2	68、8
	11	12						8	19	20
第一组	70、1	53、9	74、6	73	58、7	74、9	79、3	59、9	78、6	79、22
第二组	61、6	68、3	68、8	72、6	65、7	69、9	74、5	65、4	72、6	75、8
	21	22	23	24	25	26	27			

第一组	77、1	77、2	85、6	78	69、2	73、8	73			
第二组	72、2	71、6	77、1	71、5	68、2	72	71、5			

根据表 1，用 excel 作出两组评酒师对每一类葡萄酒得评分折线图。

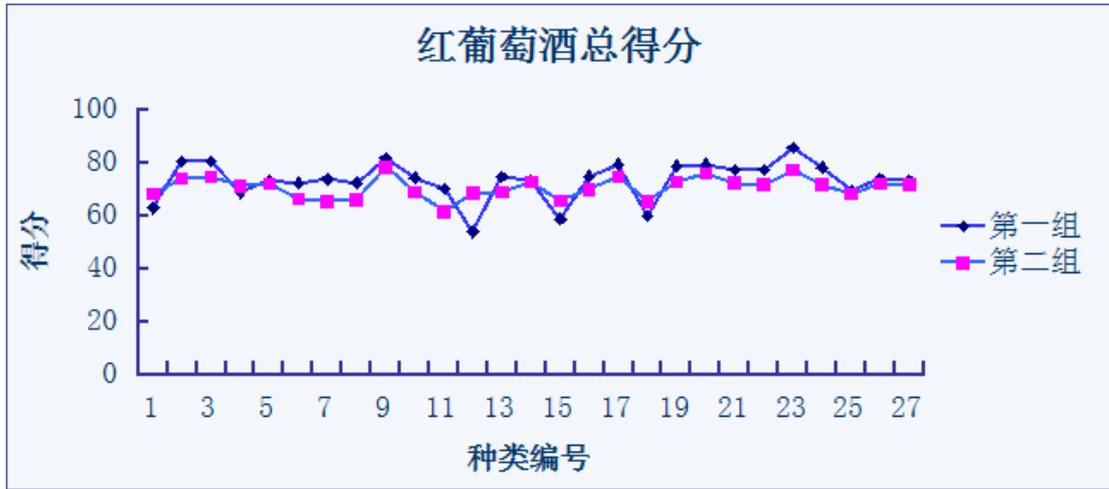


图 1

表 2 红葡萄酒总得分平均值

白酒 n	1	2	3	4	5	6	7	8	9	10
第一组	82	74、2	78、3	79、4	71	68、4	77、5	70、4	72、9	74、3
第二组	77、9	75、8	75、6	76、9	81、5	75、5	74、2	72、3	80、4	79、8
	11	12						8	19	20
第一组	72、4	63、3	65、9	72	72、4	74	78、8	73、1	72、2	77、8
第二组	71、4	72、4	73、9	77、1	78、4	67、3	80、3	76、7	76、4	76、6
	21	22	23	24	25	26	27	28		
第一组	76、4	71	75、9	73、3	77、1	81、3	64、8	81、3		
第二组	79、2	79、4	77、4	76、1	79、5	74、3	77	79、6		

根据表 2，用 excel 作出两组评酒师对每一类葡萄酒得评分折线图。

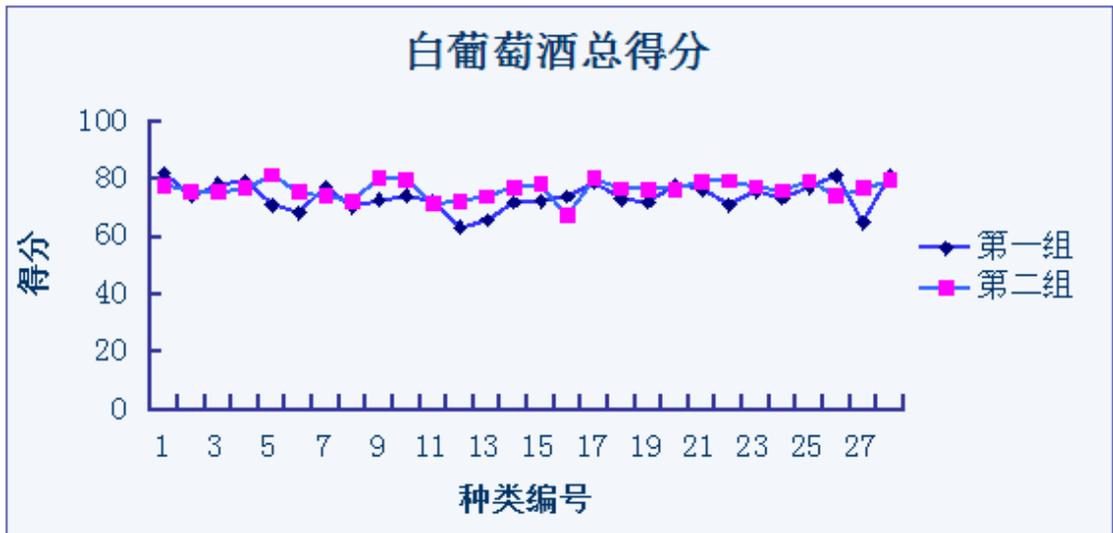


图 2

根据图1、图2可初步简单瞧出两组评酒师得评价结果存在有显著性差异。

5.1.2 葡萄酒评价结果差异性分析与可信度分析模型建立与求解

(1) 检验模型建立

首先假定两个总体平均数间没有显著差异，即

查 T 值表, 比较计算得到得 T 值与理论 T 值, 推断发生概率 (一般为 95%)。

两个正态总体得均值检验模型

假设 μ_1 就是来自总体 μ_1 得样本 μ_2 就是来自总体 μ_2 得样本, 且两样本独立。设 \bar{x}_1, \bar{x}_2 与 s_1^2, s_2^2 均未知, 其检验问题为

、

且、

当为真时, 统计量 T 得计算公式

、

式中,

、

查 T 值表, 比较计算得到得 T 值与理论 T 值, 推断发生概率 (一般为 95%), 其中 α 为显著性水平,

因此当 $|T| > T_{\alpha/2}$ 则认为不成立, 两组评酒员对红葡萄酒得评价结果有显著性差异。

(2) 两组评酒员对红葡萄酒得评价结果比较:

分别计算出

，说明该两组评酒员对红葡萄酒得评价结果有显著性差异。

(3) 两组评酒员对白葡萄酒得评价结果比较：

分别计算出

，说明该两组评酒员对白葡萄酒得评价结果有显著性差异。

5、1、3 可信度分析模型建立与求解：

第 k 组序号为 h 得样品 第 i 个指标 10 位品酒师给分得平均值

第 k 组序号为 h 得样品第 i 个指标 10 位品酒师得标准差

算出第 k 组序号为 h 得样品得稳定性指标

第 k 组红, 白葡萄酒得评分总平均稳定性指标

计算求得：

比较红葡萄酒得两组总平均稳定性指标，因为，所以第二组品酒师得评价结果更可信。

同样，比较白葡萄酒得总平均稳定性指标，因为，所以第二组品酒师得评价结果可信度更高。

5、2 问题二：

根据酿酒葡萄得理化指标与葡萄酒得质量对这些酿酒葡萄进行分级。问题二求根据酿酒葡萄得理化指标与葡萄酒得质量对酿酒葡萄进行分级，葡萄酒由酿酒葡萄酿制而成，则酿酒葡萄得质量与葡萄酒得质量有着直接得关系，则可以根据葡萄酒得质量对酿酒葡萄做一个简单得分级，在根据主成分分析从葡萄得理化指标中筛选出对葡萄质量产生影响得主要因素，根据所得各主要因素得贡献率给个因素加权作为系数，求出葡萄中主成分得含量，并进行排名，之后将此排名与之前根据葡萄酒质量所得出得排名综合，进而得出较准确得对酿酒葡萄得分级。

5、2、1 K 均值法聚类分析模型

k 均值法得基本步骤:

- (1) 选择 k 个葡萄酒样品作为初始凝聚点, 或者将所有葡萄酒样品分成 k 个初始类, 然后将这 k 个类得重心 (均值) 作为初始凝聚点。
- (2) 对除凝聚点之外得所有葡萄酒样品逐个归类, 将每个葡萄酒样品归入凝聚点离它最近得那个类 (通常采用欧氏距离), 该类得凝聚点更新为这一类目前得均值, 直至所有葡萄酒样品都归了类. 3 (▲) 重复步骤 (2), 直至所有得葡萄酒样品都不能再分配为止. ▲最终得聚类结果在一定程度上依赖于初始凝聚点或初始分类得选择. 经验表明, 聚类过程中得绝大多数重要变化均发生在第一次再分配中. 也就就是: 先算各类得均值再算各类中样本到本类及其她类得均值得绝对值距离 (欧氏距离) 将葡萄酒样本重新归类到欧氏距离较小得类中 (重新归类就得算均值)

首先, 根据第一问得出得结果, 我们采用第二组评酒员得结果作为判断葡萄酒质量得依据, 根据各葡萄酒得分数, 我们得出了红葡萄酒与白葡萄酒得排名, 虽然就是葡萄酒质量得排名, 但由于葡萄酒得质量由酿酒葡萄得质量决定, 所以上表

可以瞧作就是葡萄质量得排名, 以上表中葡萄酒得分数作为酿酒葡萄质量得分数, 可以对酿酒葡萄作出初步得分级, 针对葡萄酒得成绩, 我们用聚类分析得方法, 得出了葡萄得初步分级, 运行得得到得图样如下:

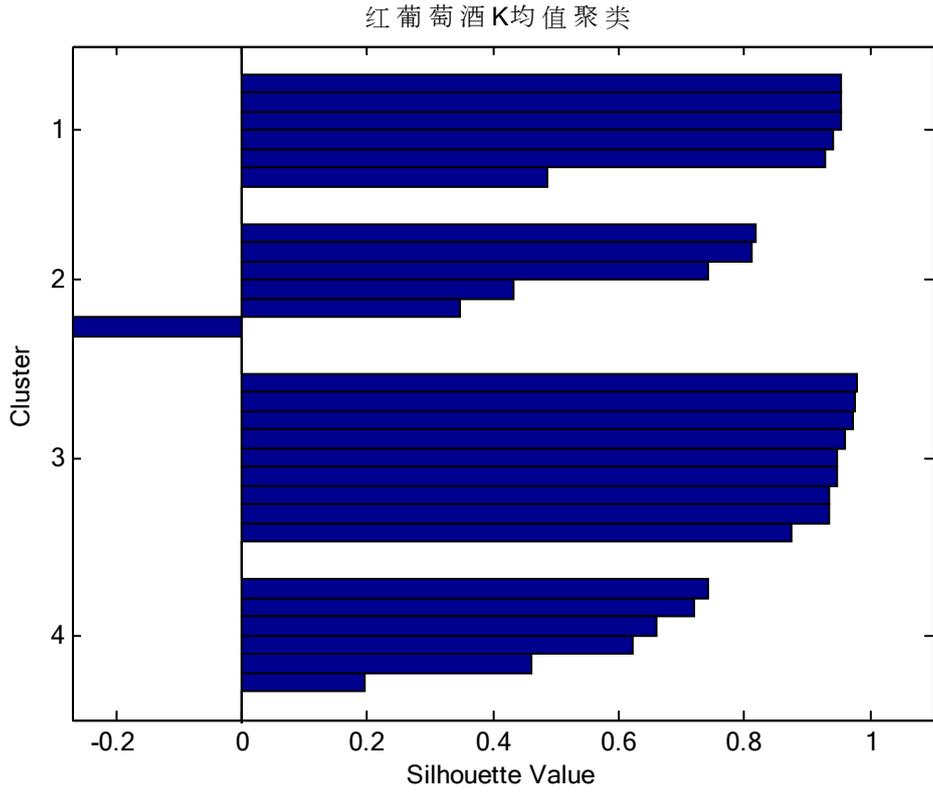


图 3

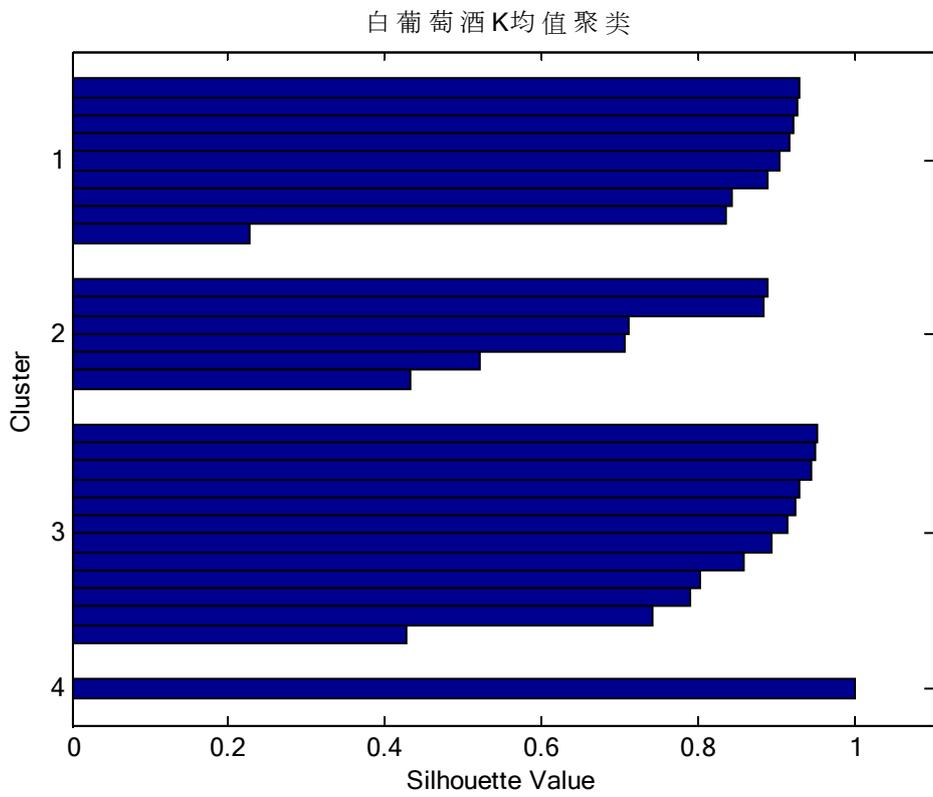


图 4

根据上述结果,得出红、白葡萄酒得等级分类,建立表3,表4、

表 3 红葡萄酒等级分类

等级	酒样品号
A	1, 10, 12, 13, 16, 25
B	4, 5, 14, 19, 21, 22, 24, 26, 27
C	6, 7, 8, 11, 15, 18
D	2, 3, 9, 17, 20, 23

表 4 白葡萄酒等级分类

等级	酒样品号
A	5, 9, 10, 15, 17, 21, 22, 25, 28
B	1, 2, 3, 4, 6, 14, 18, 19, 20, 23, 24, 27
C	7, 8, 11, 12, 13, 26
D	16

5、2、2 主成分—权值分级模型

虽然酿酒葡萄所对应葡萄酒得质量能在一定程度上反映酿酒葡萄得质量, 但葡萄得质量还应以葡萄本身得成分来区分其级别, 为了得到更准确得分级, 我们又对附件中所给酿酒葡萄中得理化指标做了一些分析。

为了综合考虑酿酒葡萄得理化指标与葡萄酒得质量对这些酿酒葡萄进行分级, 将附件 3 中芳香物质含量总与作为一个一级理化指标, 设第 i 个样品葡萄理化指标得分为, 葡萄酒得质量总分为, 则第 i 个葡萄样品得总得分可以表示为

$$1、2、2、5(\circ)$$

选取一个使得样品趋于较稳定值得, 此时得可作为酿酒葡萄得分级权值。

(1) 首先对各理化指标进行归一化处理, 酿酒葡萄一级理化指标中样本有 n 个, 指标有 m 个, 分别设为, 令为第 i 个样品得第 j 个指标. 做变换

$$(5、2、2、2)$$

得到标准化得数据矩阵, 其中

$$(5、2、2、3)$$

(2) 在标准化数据矩阵 N 得基础上计算个原始指标相关性系数矩阵

其中

$$(5、2、2、4)$$

(3) 求相关性系数矩阵 R 得特征值并排序, 再求出 R 得特征值得相应得正交单位化特征向量, 则第 i 个主成分可表示为各指标得线性组合。

计算综合得分。首先计算得到第 i 个样本中第 k 个主成分得得分为, 再以个主成分得方差贡献率为权重, 求得第 i 个样品得综合得分。

5、2、2 模型求解:

表 5 红葡萄样品主成份及其排序

主成份序列	1	2	3	4	5	6	7
主成份	花色苷	缬氨酸	干物质含量	顺式白藜芦醇苷	P H 值	多酚氧化酶活力	果梗比
主成份序列	8	9					
主成份	酪氨酸	百粒质量					

表 6 红葡萄样品综合得分

葡萄样品号	综合得分	分数排序	对应样品号	样品分差值
1	74、5	89、3	9	
2	67、0	88、6	23	0、7
3	80、6	84、6	20	4、0
4	48、9	82、5	22	2、1
5	59、4	80、6	3	1、8
6	76、5	77、5	12	3、2
7	42、7	76、5	6	1、0
8	66、3	76、0	18	0、5
9	89、3	74、5	1	1、5
10	54、4	67、7	13	6、3
11	67、5	67、5	11	0、3
12	77、5	67、0	2	0、5
13	67、7	66、3	8	0、7
14	46、3	66、0	26	0、3
15	42、9	59、4	21	6、6
16	51、6	59、4	5	0、1
17	53、9	54、4	10	5、0
18	76、0	53、9	17	0、5
19	49、6	52、4	27	1、5
20	84、6	51、6	16	0、8
21	59、4	49、6	19	2、0
22	82、5	49、1	24	0、5
23	88、6	48、9	4	0、1

24	49、1	47、4	25	1、5
25	47、4	46、3	14	1、1
26	66、0	42、9	15	3、5
27	52、4	42、7	7	0、2

对综合得分相邻样品分差值进行分析，当其值达到 3、5 及以上，认为两酿酒葡萄得品质差异较大，不能分在同一级，按照此方法，红葡萄可分成六级，一级到六级表示葡萄品质逐渐降低，具体情况如下表：

表 7 红葡萄分级结果

级数	红葡萄样品号								
一级	9 23								
二级	1	3	6	12	18	20	22		
三级	2 8 11 13 26								
四级	5 21								
五级	4	7	10	15	16	17	19	24	25
六级	27								

本模型中主要以红葡萄样品得相关数据进行分级，按照同样得方法将白葡萄得相关数据代入，求得白葡萄分级如下：

表 8 白葡萄分级结果

级数	白葡萄样品号								
一级	27								
二级	1	4	10	15	18	22	23	28	
三级	5 6 12 13 17 20								
四级	2	3	14	16	21	24	25		
五级	7 8 9 11 19								

5、3 问题三:分析酿酒葡萄与葡萄酒得理化指标之间得联系

5、3、1 数据预处理标准化及综合理化指标

在处理附件2中数据时可以发现某些存在异常得数据值，如：葡萄理化指标中白葡萄百粒质量得第三次检测值为2226.1 g，明显超过其它两次得检测值。为避免异常数据值对分级结果得影响，取其它两次值得平均值替代该异常值。同时对数据进行标准化处理，取其 z 分数：

其中， x 为变量值， \bar{x} 为平均数， s 为标准差。 z 分数表示得就是此变量大于或小于平均数几个标准差。由于 z 分数分母得单位与分子得单位相同，故 z 分数没有单位，因而可以用 z 分数来比较两个从不同单位总体中抽出得变量值。同时将原始数据直接转化为 z 分数时，常会出现负数与带小数点得值。

5、3、2 多元线性回归模型

(1) 模型建立

观察所给附件中得数据易知,影响酿酒葡萄与葡萄酒理化指标得因素往往不止一个,所以建立多元线性回归模型求解酿酒葡萄与葡萄酒两者理化指标之间得联系。

设变量 Y 与变量间有线性关系

、

式中,与就是未知参数,。

设 就是 得 n 次独立观测值,则多元线性模型可表示为

、

式中, ,且独立同分布.可用矩阵形式表示,令

则多元线性模型可表示为 。

式中、

(2) 模型求解

类似于一元线性回归,求参数得估计值,就就是求最小二乘函数

、

达到最小得值,可以证明得最小二乘估计

、

从而可得经验回归方程为 、

将酿酒葡萄瞧做自变量,葡萄酒瞧做因变量。注意,计算时用得就是经过处理后得Z分数表。

我们用 表示酿酒葡萄得30个一级指标,作为自变量 X ;用表示葡萄酒得 9 个一级指标,作为因变量 y 。

其中,理化指标得编号顺序依照所给附件中得大小顺序。例如,红葡萄酒中理化指标顺序依次为花色苷、单宁、总酚、酒总黄酮、白藜芦醇、DPPH半抑制体积、 L 、 a 、 b 。经过MATLAB对回归系数得最小二乘估计计算,得出回归系数,即自变量与因变量之间得联系,见附表。

根据回归系数表得出两者之间得正负相关性,其中数字为酿酒葡萄理化指标编号。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。

如要下载或阅读全文，请访问：

<https://d.book118.com/99622113223011000>