

重庆大学 2013 年暑期第四次模拟竞赛题

2012_A 题：基于统计分析的葡萄酒质量评价问题

姓名	学院	年级专业	学号	联系电话	相关学科成绩						
					高等数学	线性代数	概率统计	数学模型	数学实验	英语四级	英语六级
杨阳	动力工程	2011 核工程与核技术	2011****	186*****	96	94	78	100	95	480	450
刘蛰	自动化	自动化	2011****	151*****	89	83	69	100	95	/	/
熊莹	机械工程	2011机自	2011****	185*****	90	91	88	100	90	507	/

2013.7.31

基于统计分析的葡萄酒质量评价问题

摘要

简要叙述所给问题的意义和要求，并分别分析每个小问题的特点。针对这些特点对问题一用.....的方法解决；对问题二用.....的方法解决；对问题三用.....的方法解决。

对于问题一用.....数学中的.....首先建立.....模型 I。在对.....模型改进的基础上建立了.....模型 II。对模型进行了合理的理论证明和推导，所给出的理论证明结果大约为.....然后借助于.....数学算法和.....软件，对附件中所提供的数据进行了筛选，去除异常数据，对残缺数据进行适当补充，并从中随机抽取了 3 组数据（每组 8 个采样）对理论结果进行了数据模拟，结果显示，理论结果与数据模拟结果吻合。（方法、软件、结果都必须清晰描述，可以独立成段，不建议使用表格）

对问题二，利用问题一中已经得出的品酒师可靠性的结果，得到 27 组红葡萄酒和 28 组白葡萄酒的得分。将红葡萄酒的得分和红葡萄酒的各项理化指标一起构成一个 27×64 大小的数据矩阵，即样本观测矩阵。将白葡萄酒的得分和白葡萄酒的各项理化指标一起构成一个 28×64 大小的数据矩阵。通过以上步骤完成数据导入。由于各指标间的量纲不同并且数值的差异也较大，先对样本观测矩阵进行方差标准化处理，得到标准化后的样本观测矩阵，利用标准化的样本观测矩阵画出数据的箱线图，形象的展示出数据在各个指标上的分布情况。由于指标较多有 64 个，不同的指标之间可能共线性，这将对聚类结果产生不良影响，在聚类分析之前先对数据进行主成分分析到达降低变量维数的目的，在累积贡献率不小于 90% 的前提下保留前几个主成分，在主成分分析过程中发现大量的变量之间存在共线性关系。最后利用保留下来的前几个主成分进行系统聚类分析，聚类分析采用欧式距离和离差平方和法（Ward 方法），最后得到动态聚类树，并结合葡萄酒的得分和葡萄的特征性化学成分的含量，将酿酒葡萄分为 A（优质）、B（良好）、C（中等）、D（差）四个等级，客观的反映出了酿酒葡萄理化指标与葡萄酒质量之间的关系。

对问题三用.....

（注意：如果题目单问题，则至少要给出 2 种模型，分别给出模型的名称、思想、软件、结果、亮点详细说明。并且一定要在摘要对两个或两个以上模型进行比较，优势较大的放后面，这两个模型一定要有具体结果）

如果在.....条件下，模型可以进行适当修改.....（这种条件的改变可能来自你的一种猜想或建议。要注意合理性。此推广模型可以不深入研究，也可以没有具体结果。）

关键词：秩和检验法；主成分分析；系统聚类分析；数理统计；偏最小二乘回归

（注：字数：700—1000 之间；摘要中必须将具体方法、结果写出来；摘要写满几乎一页，不要超过一页）

目录

摘要	1
一. 问题的重述	3
二. 模型假设及符号说明	3
2.1 模型的假设	3
2.2 符号说明	4
三. 问题分析	5
四. 模型的建立和求解	6
4.1 问题一模型的建立及求解	错误!未定义书签。
4.2 问题二的求解	错误!未定义书签。
4.3 问题二的求解	错误!未定义书签。
五. 结果分析与检验	7
六. 模型的优缺点	29
6.1 模型的优点	29
6.2 模型的缺点	29
参考文献	29
附录	29
1.1..... (程序文件说明)	29
1.2..... (程序文件说明)	30
1.3..... (程序文件说明)	30
1.4..... (程序文件说明)	32
1.5..... (程序文件说明)	33

一. 问题的重述

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. 模型假设及符号说明

2.1 模型的假设

- (1) 假设题目所给的数据真实可靠；
- (2) 忽略酿酒过程中人为因素的影响；
- (3) 主成分的累积贡献率达到 90% 以上能够较好的解释原始变量；
- (4)
- (5)

注：假设对整篇文章具有指导性，有时决定问题的难易。一定要注意假设的某种角度上的合理性，不能乱编，完全偏离事实或与题目要求相抵触。注意罗列要工整。

2.2 符号说明

对文章中所用到的主要数学符号进行解释

尽可能借鉴参考书上通常采用的符号，不宜自己乱定义符号，对于改进的一些模型，符号可以适当自己修正（下标、上标、参数等可以变，主要符号最好与经典模型符号靠近）。

对文章自己创新的名词需要特别结实。其他符号要进行说明，注意罗列要工整。如“ $x_{ij} \sim$

第 i 种疗法的第 j 项指标值”等，注意格式统一，不要出现零乱或前后不一致现象，关键是容易看懂。

符号	说明	单位
x_i	取自总体 \mathcal{X} 的样本	
X	样本观测矩阵	
x_i^*	取自标准化总体 \mathcal{X}^* 的样本	
n	样本容量	
p	样本变量维数	
S	样本协方差据矩阵	
\hat{R}	样本相关系数矩阵	
μ	总体的期望	
Σ	总体的协方差	
t_i	矩阵 Σ 的特征向量	
λ_i	矩阵 Σ 的特征值	
G	类别	
D_{KL}	类别 G_K 与类别 G_L 之间的距离	
W	类内离差平方和	

三. 问题分析

3.1 问题一的分析

分析附件 1，对数据进行适当处理后，求出两组评酒员分别对红葡萄酒与白葡萄酒的各样品的评价总分数。将第一组评酒员对红葡萄酒的评价结果与第二组评酒员对红葡萄酒的评价结果看作两个总体分布，判断两组评酒员的评价结果有无显著性差异，就变成检验这两个总体分布是否有明显差异，根据数理统计知识，可选用秩和检验法。同样可将第一组评酒员对白葡萄酒的评价结果与第二组评酒员对白葡萄酒的评价结果看作两个总体分布，进行秩和检验，判断两组评酒员的评价结果有无显著性差异。评价结果与评价的期望越接近，即其方差越小，其结果更可信。由此可求出各总体的方差来判断哪组评酒员的评价结果更可信。

3.2 问题二的分析

问题二中要求根据葡萄酒的理化指标和品酒员通过评价给出的葡萄酒的质量对葡萄酒进行分级，酿酒葡萄的成分直接影响到葡萄酒的质量。优质的葡萄酒不但要具有纯正的口感还要具有美观的色泽和芬芳的酒香。而这些感官质量只能通过品酒师的评价得到通过查阅相关文献^{[1][2]}可以得知单宁和花色苷可以影响葡萄酒的色泽，单宁也会影响葡萄酒的口感，氨基酸含量代表了葡萄酒的营养价值，不同氨基酸之间，氨基酸与氨基酸总量存在着一定的关系。

所以，对酿酒葡萄进行分级，不仅仅只根据酿酒葡萄的理化指标进行分级还应该结合酿出的葡萄酒的质量，即品酒师给出的评价进行分级。根据葡萄的理化指标和葡萄酒质量对葡萄进行分级可以使用系统聚类法，通过系统聚类做出聚类树后再根据葡萄酒的评价指标和葡萄酒的特征性化学成分单宁做出最后的分级结果。

在聚类分析之前，变量的量纲不同数值差异较大，应该对样本进行标准化。同时由于不同的变量之间存在共线性，所以应该对数据进行主成分分析削弱数据的共线性，降低数据变量的维数。

3.3 问题三的分析

对问题三研究的意义进行分析。

问题三属于。。。。。。数学问题，对于解决此类问题一般数学方法的分析。

对附件中所给数据特点的分析。

对问题三所要求的结果进行分析。

由于以上原因，我们可以将首先建立一个。。。。。。的数学模型 I，然后将建立一个。。。。。。的模型 II，。。。。。。对结果分别进行预测，并将结果进行比较。

主要是表达对题目的理解，特别是对附件的数据进行必要分析、描述（一般都有数据附件），这里需要提到分析数据的方法、理由。如果有多个小问题，可以对每个小问题进行分别分析。

四. 模型的准备

4.1 秩和检验法

秩和检验是检验两个总体分布是否有明显差异或两个独立样本是否来自同一总体的方法。

设 $F_X(x), F_Y(y)$ 分别为连续型总体 X, Y 的分布函数, $f_X(x), f_Y(y)$ 为它们的密度函数, 这些函数都未知, X_1, X_2, \dots, X_n 和 Y_1, Y_2, \dots, Y_m 是分别来自 X 和 Y 的样本, 且相互独立, 样本值分别为 $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$, 统计假设是:

$$H_0: F_X(x) = F_Y(x), H_1: F_X(x) \neq F_Y(x)$$

(1) 秩的概念: 将样本 X_1, X_2, \dots, X_n 的样本值 x_1, x_2, \dots, x_n 按由小到大的顺序排成一排, 使得

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

如果 $x_i = x_{(k)}$, 则称 k 是 X_i 的秩 ($i=1, 2, \dots, n$)。将 X_1, X_2, \dots, X_n 与 Y_1, Y_2, \dots, Y_m 混合, 且按观测值的大小顺序排列, 可得每个变量的秩。

(2) 秩和检验法: 设 $n \leq m$, 将 X_1, X_2, \dots, X_n 在 X_1, X_2, \dots, X_n 与 Y_1, Y_2, \dots, Y_m 的混合样本中的秩相加, 记其和为 T , 则有:

$$\frac{1}{2}n(n+1) \leq T \leq \frac{1}{2}n(n+2m+1)$$

在 H_0 成立的条件下, 两个独立样本 X_1, X_2, \dots, X_n 与 Y_1, Y_2, \dots, Y_m 应来自同一总体, 这时

X_1, X_2, \dots, X_n 随机分散在 Y_1, Y_2, \dots, Y_m 中, 因而 T 不应太大, 也不应太小, 否则认为 H_0 不成立。于是选择拒绝域的形式为:

$$\{T < t_1\} \cup \{T > t_2\} \quad (t_1 < t_2)$$

且

$$P(T < t_1 \text{ 或 } T > t_2 | H_0 \text{ 成立}) \leq \alpha$$

其中 α 为给定的显著性水平, 由于

$$P(T < t_1 \text{ 或 } T > t_2 | H_0 \text{ 成立}) = P(T < t_1 | H_0 \text{ 成立}) + P(T > t_2 | H_0 \text{ 成立})$$

令

$$P(T < t_1 | H_0 \text{ 成立}) = P(T > t_2 | H_0 \text{ 成立}) = \frac{\alpha}{2}$$

由此可以查表求出临界值 $t_1(n, m), t_2(n, m)$ 。

4.2 聚类分析（进行采样）

4.3 预测的准备工作

五. 模型的建立与求解

5.1 问题一

5.1.1 红葡萄酒评价结果的差异显著性分析

分别计算出两组评酒员对红葡萄酒的 27 个酒样品的评价总分，如表 1 所示。

表 1 两组评酒员对红葡萄酒的评价结果

评酒员	酒样品 1	酒样品 2	酒样品 3	酒样品 4	酒样品 5	酒样品 6	酒样品 7
第一组	77	96	94	84	86	86	89
第二组	84	91	89	88	89	86	84
评酒员	酒样品 8	酒样品 9	酒样品 10	酒样品 11	酒样品 12	酒样品 13	酒样品 14
第一组	84	97	88	84	79	89	87
第二组	84	94	84	81	87	86	93
评酒员	酒样品 15	酒样品 16	酒样品 17	酒样品 18	酒样品 19	酒样品 20	酒样品 21
第一组	76	91	92	79	92	92	92
第二组	82	88	90	88	87	93	90
评酒员	酒样品 22	酒样品 23	酒样品 24	酒样品 25	酒样品 26	酒样品 27	
第一组	90	100	91	83	89	90	
第二组	89	89	88	86	88	88	

将第一组评酒员对红葡萄酒的评价结果与第二组评酒员对红葡萄酒的评价结果看做两个总体分布。使用 Matlab 软件编程(程序见附录 1.1)，调用 ranksum 函数对两总体分布进行秩和检验，在显著水平 $\alpha = 0.05$ 下，程序输出结果为如表 2 所示。

表 2 红葡萄酒显著性差异的秩和检验结果

样品	stats		p	h
	zval	ranksum		
红葡萄酒	0.6862	782.5	0.4926	0

分析结果知，一方面，两总体分布相似的概率 $p = 0.4926 > 0.05$ ，不趋近于 0；另一

方面， $h=0$ 也说明可以接受原假设 H_0 ，即认为两组评酒员对红葡萄酒的评价结果无显著性差异。红葡萄酒评分结果直方图如图 1 所示。由图 1 可以得到基本相似的结论。

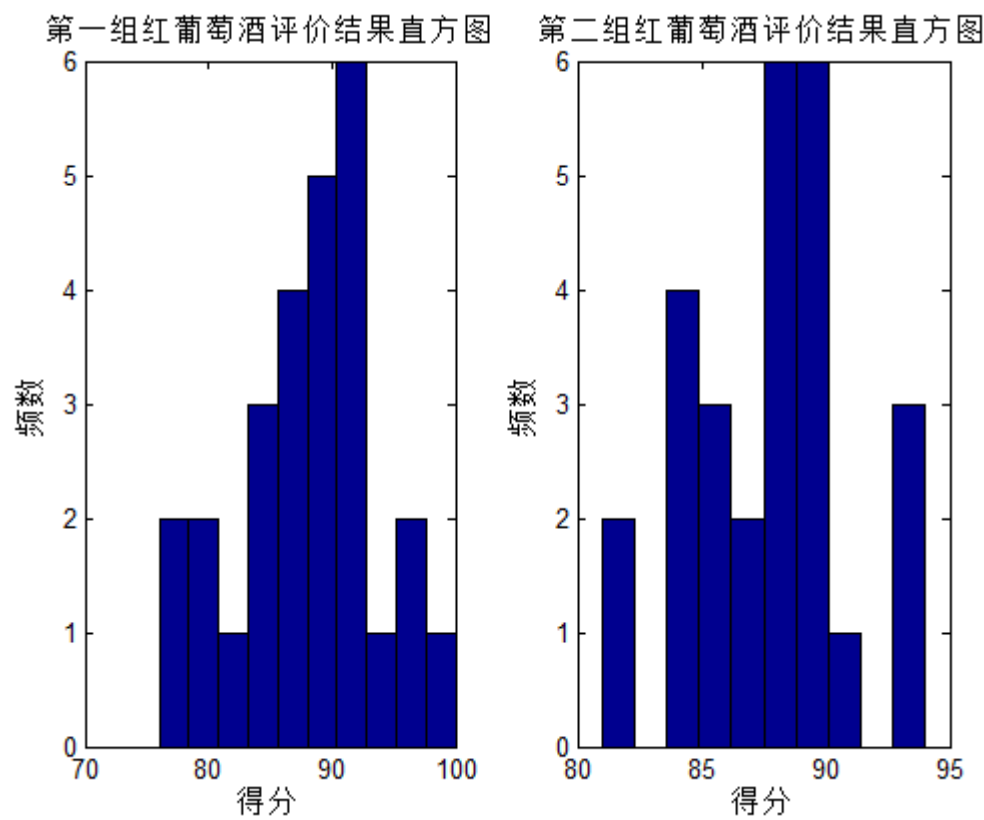


图 1 红葡萄酒评分结果直方图

计算两样本总体的方差，第一组评酒员的评价结果方差为 35.6524，第二组评酒员的评价结果方差为 10.2422。评价结果与期望越接近，其方差越小，其更可信。因此第二组评酒员的评价结果更可信。

5.1.1 白葡萄酒评价结果的差异显著性分析

分别计算出两组评酒员对白葡萄酒的 28 个酒样品的评价总分，如表 3 所示。

表 3 两组评酒员对白葡萄酒的评价结果

评酒员	酒样品 1	酒样品 2	酒样品 3	酒样品 4	酒样品 5	酒样品 6	酒样品 7
第一组	97	91	96	94	89	86	94
第二组	94	92	93	93	96	92	86
评酒员	酒样品 8	酒样品 9	酒样品 10	酒样品 11	酒样品 12	酒样品 13	酒样品 14
第一组	88	87	92	87	82	85	88
第二组	91	94	95	93	86	88	94
评酒员	酒样品 15	酒样品 16	酒样品 17	酒样品 18	酒样品 19	酒样品 20	酒样品 21
第一组	87	93	94	89	86	90	92
第二组	93	87	97	92	91	91	95
评酒员	酒样品 22	酒样品 23	酒样品 24	酒样品 25	酒样品 26	酒样品 27	酒样品 28
第一组	83	92	89	88	94	83	94

第二组	92	96	92	96	92	93	95
-----	----	----	----	----	----	----	----

将第一组评酒员对白葡萄酒的评价结果与第二组评酒员对白葡萄酒的评价结果看做两个总体分布。在显著水平 $\alpha = 0.05$ 下，秩和检验结果如表 4 所示。

表 2 白葡萄酒显著性差异的秩和检验结果

样品	stats		p	h
	zval	ranksum		
白葡萄酒	-2.5366	643.5	0.0112	1

分析结果可知，两总体分布相似的概率 $p = 0.0112 < 0.05$ ，趋近于 0，又 $h = 1$ ，所以两组评酒员对白葡萄酒的评价结果有显著性差异。白葡萄酒评分结果直方图如图 2 所示。由图 2 可以得到基本相似的结论。

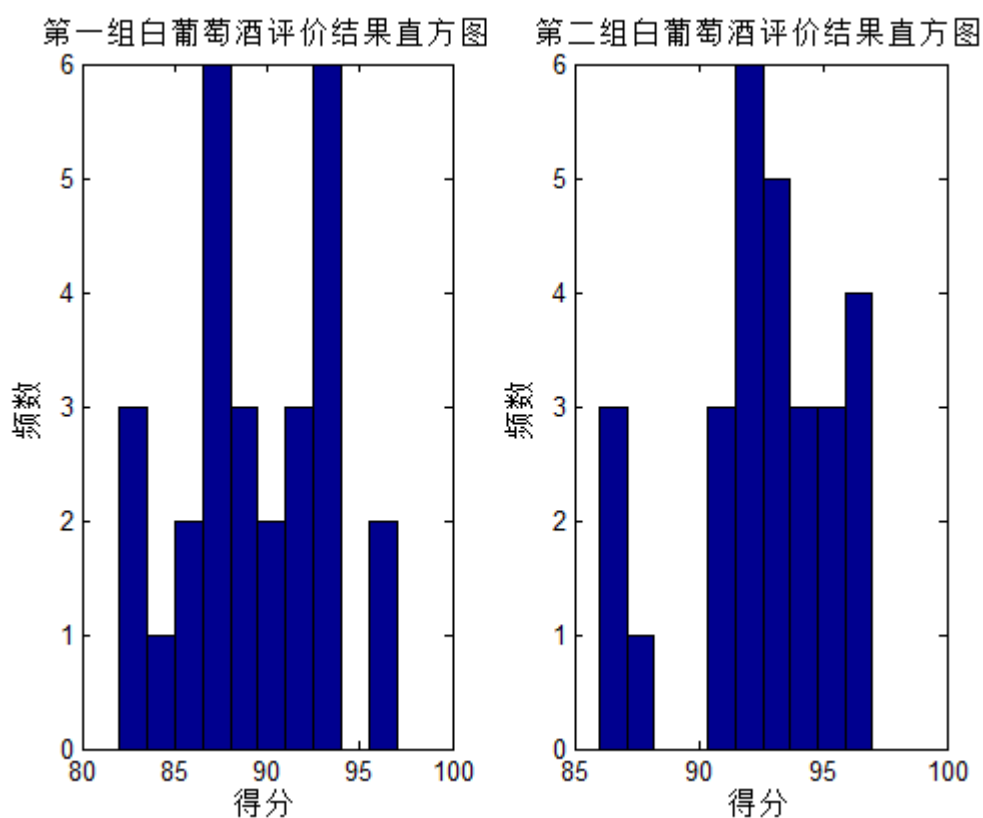


图 2 白葡萄酒评分结果直方图

计算两样本总体的方差，第一组评酒员的评价结果方差为 16.4603，第二组评酒员的评价结果方差为 8.4061。评价结果与期望越接近，其方差越小，其更可信。因此第二组评酒员的评价结果更可信。

该种模型的一般数学表达式、意义和式中各种参数的意义。注明参考文献。

5.1.2 模型 I 的建立与求解

- (1) 说明问题一适用用此模型来解决，并将模型进行改进以适应问题一。
- (2) 借助准备工作中的采样，（用拟合等方法）确定出模型中的参数。
- (3) 给出问题一的数学模型 I 的表达式和图形表达式。
- (4) 给出误差分析的理论估计。

如有多个模型，各模型间必须进行相互比较

5.1 问题二模型的建立与求解

5.1.1 数据导入：

所给的数据中包含有一项指标多次测量的情况，对其在 excel 表格中做平均处理，认为多次测量的平均值为该项指标的值。数据本身具有 63 项指标包含氨基酸总量各种氨基酸含量，单宁含量，蛋白质含量，各种有机酸含量等一系列理化指标。加上又第一问中得到的可信组的评酒员的评价结果一共构成了 64 项指标。由于葡萄样品中，红葡萄有 27 组，白葡萄有 28 组，最终导入 MATLAB 进行分析的数据为一个 27×64 的矩阵（红葡萄）和 28×64 的矩阵（白葡萄），以及一个元胞数组用于存储各个指标的名称。

设 x_1, x_2, \dots, x_n 为取自总体 X 的样本。 $x_i = \begin{bmatrix} x_{i1} & x_{i2} & \dots & x_{ip} \end{bmatrix}$

其中 $(i = 1, 2, \dots, n)$ 记样本观测矩阵为 X

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

对于问题中的数据，红葡萄 n 取得 27， p 取 64；白葡萄 n 取 28， p 取 64。

样本观测矩阵 X 的每一行对应一个葡萄样本，每一列对应一个变量。

首先以红葡萄的分级为例说明分级过程，再对白葡萄进行分级。

5.1.2 数据预处理

原始数据变量之间量纲不同，并且数量级差异较大，并且不同的变量之间可能存在共线性。所以对数据进行先方差标准化再进行主成分分析的数据预处理。

(1) 数据方差标准化

方差标准化的作用是消除变量之间的量纲效应，并使得数据数量级差异不会过大，使得不同变量之间具有可比性和相同的表现能力。

方差标准化的过程可以表示为：

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_{jj}}} \quad \text{其中 } i=1,2,\dots,n; \quad j=1,2,\dots,p,$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$\sqrt{s_{jj}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

\bar{x}_j 为变量 x_j 观测值的平均值, s_{jj} 为变量 x_{ij} 的方差, 经过标准化后的样本观测

矩阵 X 各列的均值为 0 方差为 1。

(2) 做出方差标准化后的数据的箱线图

箱线图利用图形的方式显示位置度量(中位数), 散度量(四分位数)和可能出现的离群点。同时还能表征数据分布的对称性或非对称性(相对于分位数的中位数位置)。

箱线图的构造过程如下:

1. 在中位数, 上分位数和下分位数处化三条水平线, 增加垂直线, 制作成一个箱子。
2. 从上分位数向上画一条垂直线, 知道偏离上分位数 1.5 倍距离内的最大极值点。同样从下分位数向下划一条类似的垂直线。在这些垂直线的末尾增加短的水平线。
3. 用星号或点标识出超出垂直线端点的每个数据点。

以下为红葡萄标准化后数据的箱线图

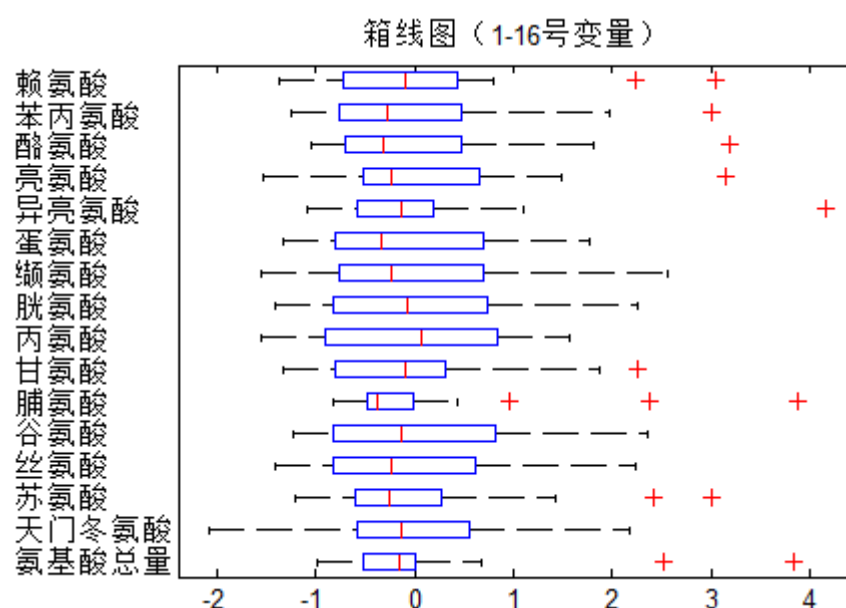


图 红葡萄 1 到 16 号变量的箱线图

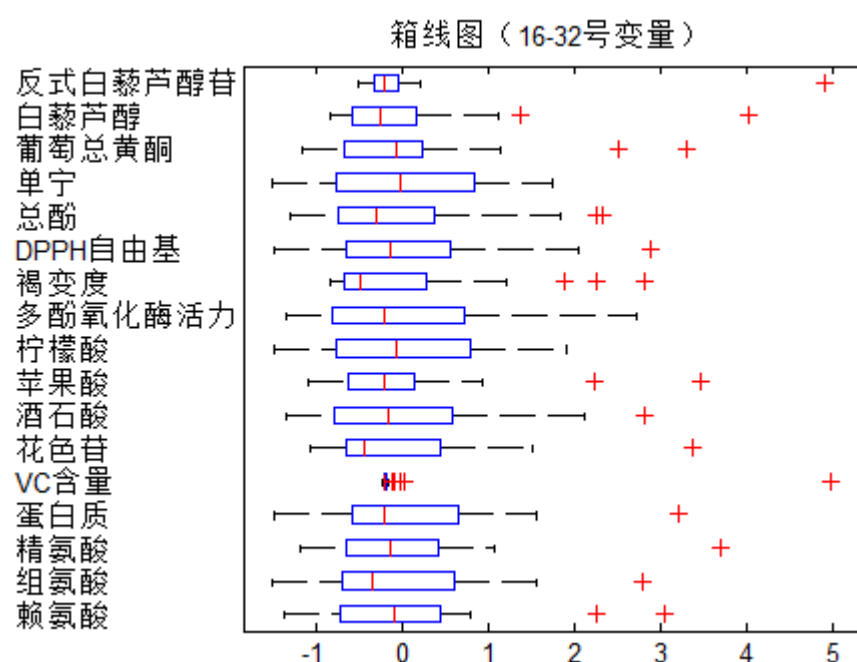


图 红葡萄 16 到 32 号变量的箱线图

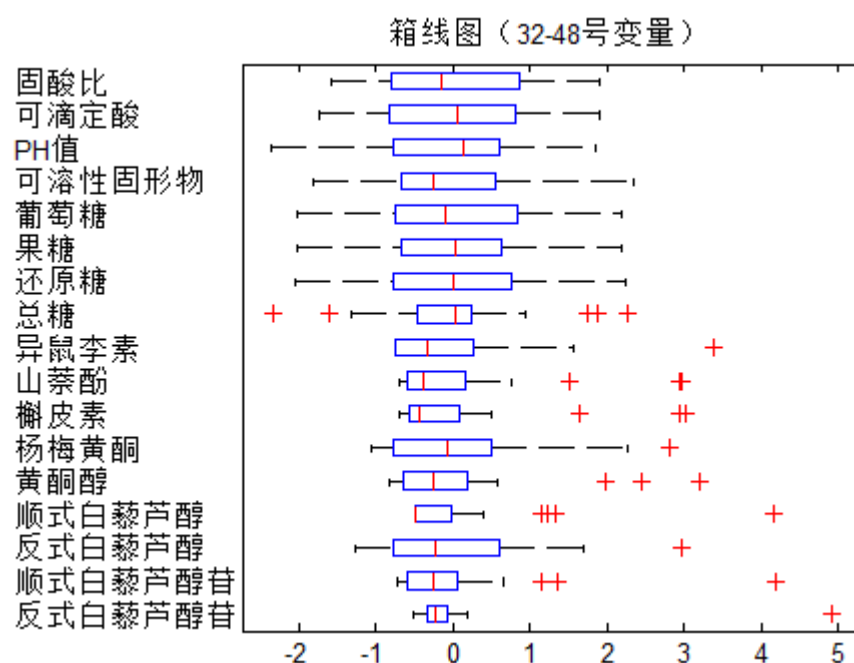


图 红葡萄 32 到 48 号变量的箱线图

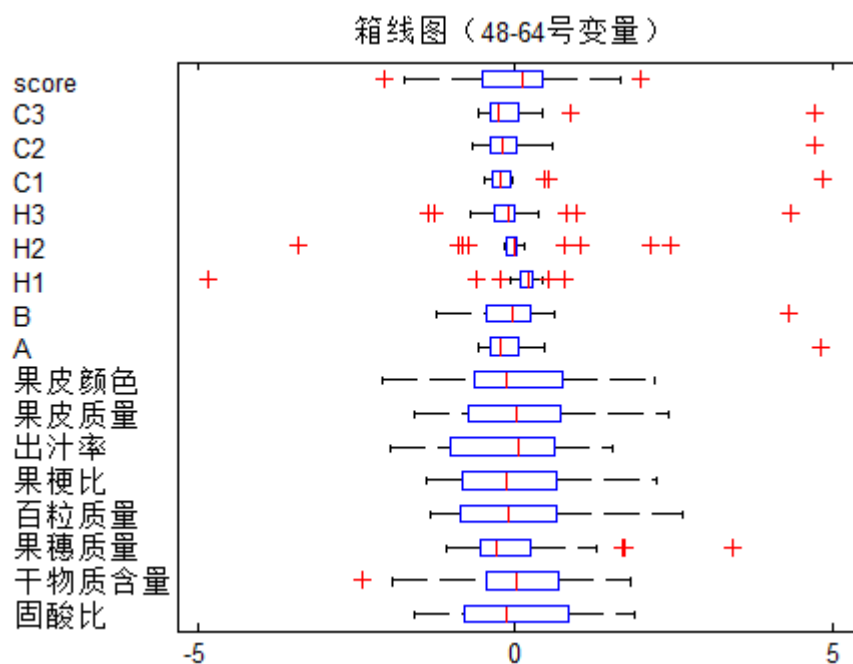


图 红葡萄 48 到 64 号变量的箱线图

很显然，箱线图可以非常有效的展示和汇总数据。对于较多组的数据箱线图可以表示出各个变量的特点。例如从箱线图中可以看出 27 个红葡萄样本的脯氨酸数据的散度在氨基酸中最小，表示脯氨酸在葡萄样本中的含量与其他氨基酸相比还是比较稳定，但是也有少数的红葡萄样本脯氨酸含量较整体的多，图中已被红色的“+”号标注。同时也可以从箱线图中直观的看到不同红葡萄样本之间的单宁含量差异较大。红葡萄葡萄样本的 VC 含量差别较小。

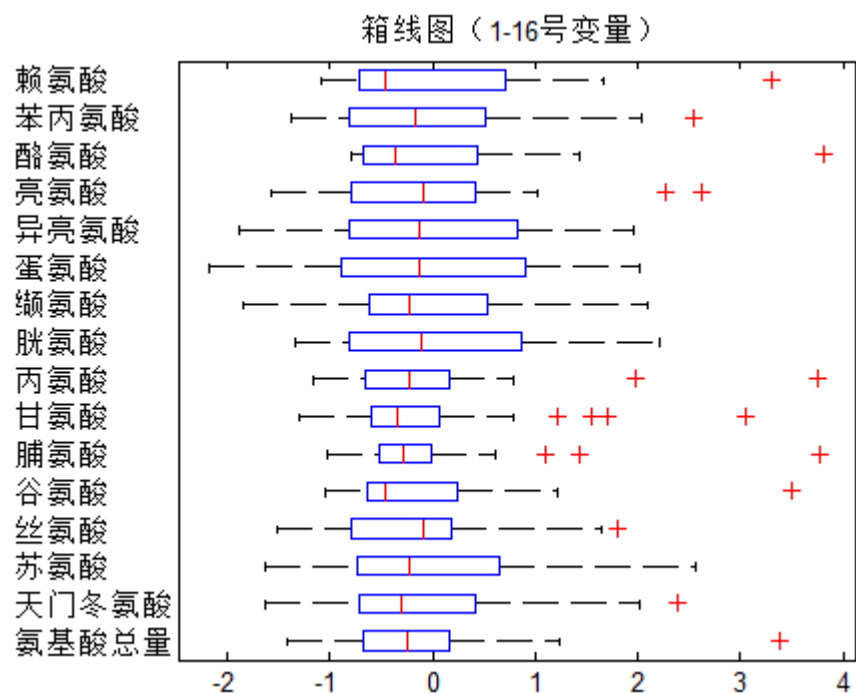


图 白葡萄 1 到 16 号变量的箱线图

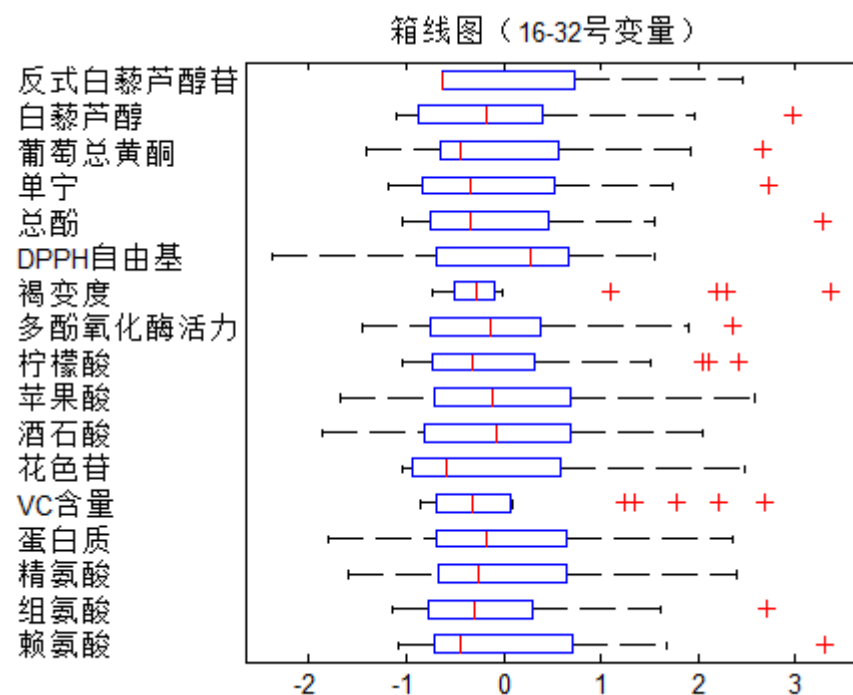


图 白葡萄 16 到 32 号变量的箱线图

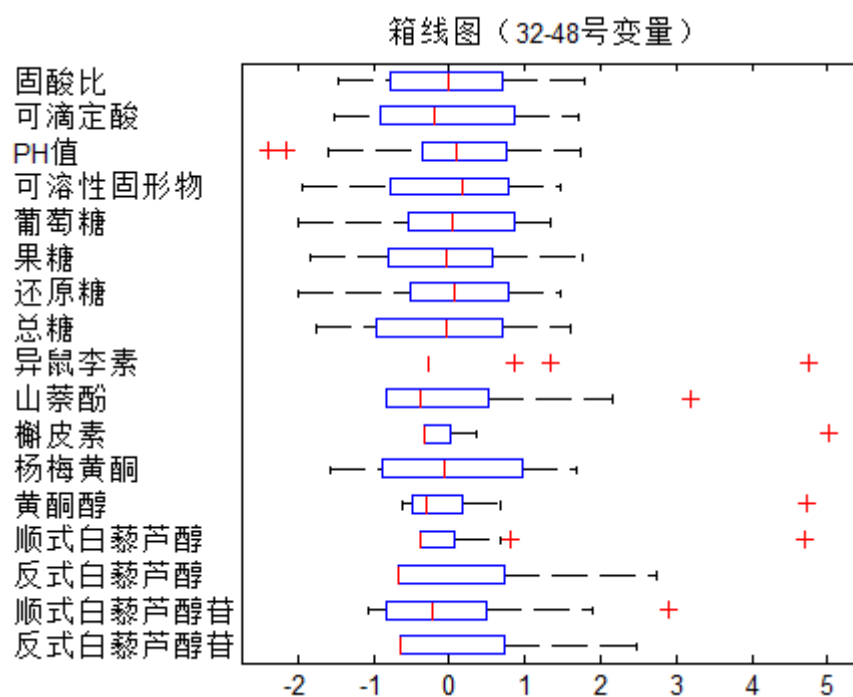


图 白葡萄 32 到 48 号变量的箱线图

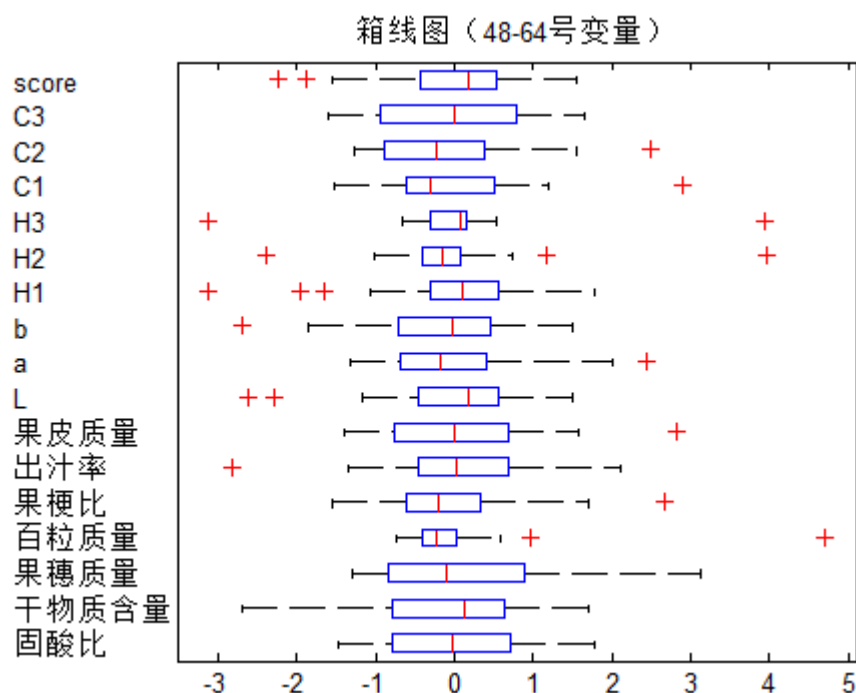


图 白葡萄 48 到 64 号变量的箱线图

从箱线图中可以看出白葡萄与红葡萄的理化性质具有相似性，脯氨酸的含量在氨基酸中也比较稳定，样本单宁的差异较大但是总体上数值较为接近，同时在某些指标上与红葡萄也有不同，比如不同的样本的 VC 的含量差异大。单宁和花色苷在不同样品间也有较大差异。

(3) 对方差标准化后的数据进行主成分分析

当研究问题涉及很多变量，并且变量之间相关性明显，使用主成分分析法可以抓住事物的主要矛盾使得问题得到简化。主成分分析的过程其实就是坐标系旋转过程，新坐标系的各个坐标轴方向是原始数据变差最大的方向，各主成分表达式就是新旧坐标转换关系式。

主成分分析过程：

(1) 计算样本的协方差矩阵和样本相关系数矩阵：

样本协方差矩阵：

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T = (s_{jj})$$

样本相关系数矩阵：

$$\hat{R} = (r_{ij})$$

$$r_{ij} = \frac{s_{jj}}{\sqrt{s_{ii}} \cdot \sqrt{s_{jj}}}$$

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 为样本均值，显然样本经过方差标准化变换后 $\bar{x} = 0$ 并且样本

协方差矩阵 S 和样本相关系数矩阵 \hat{R} 的对角线元素均为 1。

总体 X 的期望为：

$$E(x) = \mu$$

总体 X 的协方差为：

$$Var(x) = \sum$$

(2) 根据协方差矩阵进行主成分分析：

将 S 作为 \sum 的估计，从而从 \sum 出发可以求得样本的主成分。

做以下线性变换：

$$\begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_p^T x \end{bmatrix}$$

问题化为求 $a_1, a_2 \cdots a_p$ 使得 $y_1, y_2 \cdots y_p$ 方差达到最大。

设 $\lambda_1 \geq \lambda_2 \geq \lambda_p \geq 0$ 为协方差矩阵 \sum 的特征值， t_1, t_2, \cdots, t_p 为特征值对应的单位特征向量。则协方差矩阵可以表示为：

$$\sum = \sum_{i=1}^p \lambda_i t_i t_i^T = T \Lambda T^T$$

1. 当 $a_1 = t_1$ 时， $y_1 = a_1^T x$ 方差达到最大 λ_1 ， $y_1 = a_1^T x$ 为第一主成分。

2. 在 $\text{cov}(y_1, y_2) = 0$ 前提下, 当 $a_2 = t_2$ 时, $y_2 = a_2^T x$ 方差达到最大 λ_2 ,

$y_2 = a_2^T x$ 为第二主成分。

3. 在 $\text{cov}(y_k, y_i) = 0$ 前提下, 当 $a_i = t_i$ 时, 方差 $y_i = a_i^T x$ 达到最大 λ_i ,

$y_i = a_i^T x$ 为第 i 主成分。

主成分的贡献率:

总方差中第 i 个主成分 y_i 的方差所占比例 $\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$, ($i = 1, 2, \dots, p$) 称为主成分

分 y_i 的贡献率。主成分的贡献率反映了主成分综合原始变量信息的能力, 也可理解为

解释原始变量的能力。前 $m(m \leq p)$ 个主成分的贡献率之和 $\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$ 称为前 m 个主成分

的累积贡献率, 它反映了前 m 个主成分综合原始变量信息的能力。主成分的帕累托图可以形象的展示出主成分的贡献率和累积贡献率。为保留主成分的数目提供参考。

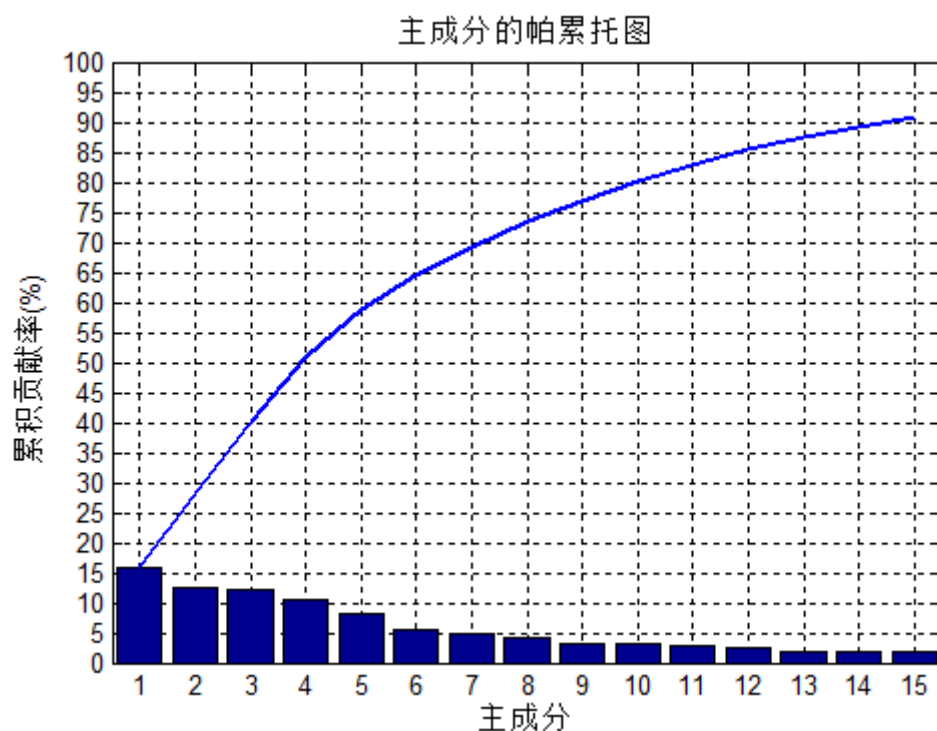


图 主成分的帕累托图（红葡萄）

从红葡萄的主成分的帕累托图中看出前5个主成分的贡献率较大，并且在解释原始变量90%变差的意义下保留前15个主成分就已经足够了。故在对红葡萄数据做系统聚类分析时用到的数据为保留前15个主成分的数据，即只有15个变量，原始变量有64个现在减为15个并且也只损失了少量的数据，达到了降低数据冗余性的目的。

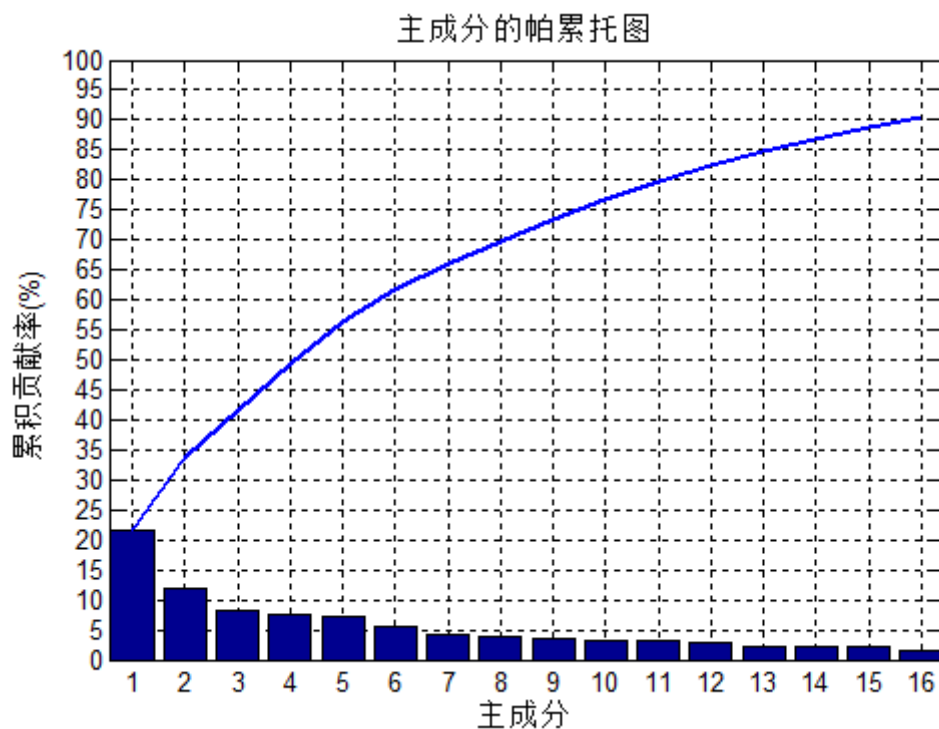


图 主成分的帕累托图（白葡萄）

分析白葡萄的主成分帕累托图可以看出第一个主成分具有突出的贡献度，前 5 个主成分贡献度较大，为了达到解释原始数据 90% 变差的目的，需要保留前 16 个主成分，也就是在对白葡萄进行聚类分析时候，有 16 个变量分别对应前 16 个主成分，通过主成分分析将白葡萄 64 个原始变量缩减到 16 个，有效的削弱了变量间的信息重叠。

原始变量对主成分的贡献

主成分 \mathcal{Y}_j 的表达式为：

$$y_j = t_j^T x = t_{1j}x_1 + t_{2j}x_2 + \cdots + t_{pj}x_p \quad (j = 1, 2, \cdots, p)$$

称 t_{ij} 为第 j 个主成分 \mathcal{Y}_j 在第 i 个原始变量上面的载荷，它反映了 x_i 对 \mathcal{Y}_j 的重要程度。通常又称 t_{ij} 为主成分载荷矩阵，即：

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \vdots & \vdots & & \vdots \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{bmatrix}$$

主成分载荷阵的每列构成一个列向量，将某一个列向量在主成分坐标系下做出可以直观的

表现出某一个指标对各个主成分 \mathcal{Y}_j 的影响的大小。若这个行向量在某一个主成分坐

标轴方向的投影大则代表了这一指标对该主成分影响大，其中坐标投影可能为正也可能负，若为负则代表了这一指标对该主成分的影响为负。在原始变量到主成分之间的线性变换过程中，主成分载荷阵的作用又起到了坐标变换作用，经过线性变换后的变量在主成分坐标系下表现为一个点，数据在主成分坐标系下的分布可以直观的展示出主成分分析的效果。主成分散点图中同时做出主成分载荷矩阵的列向量同时也做出经过主成分载荷矩阵线性变换后的数据点。数据点用红点表示，主成分载荷矩阵列向量用蓝色线条表示。以下为主成分散点图：

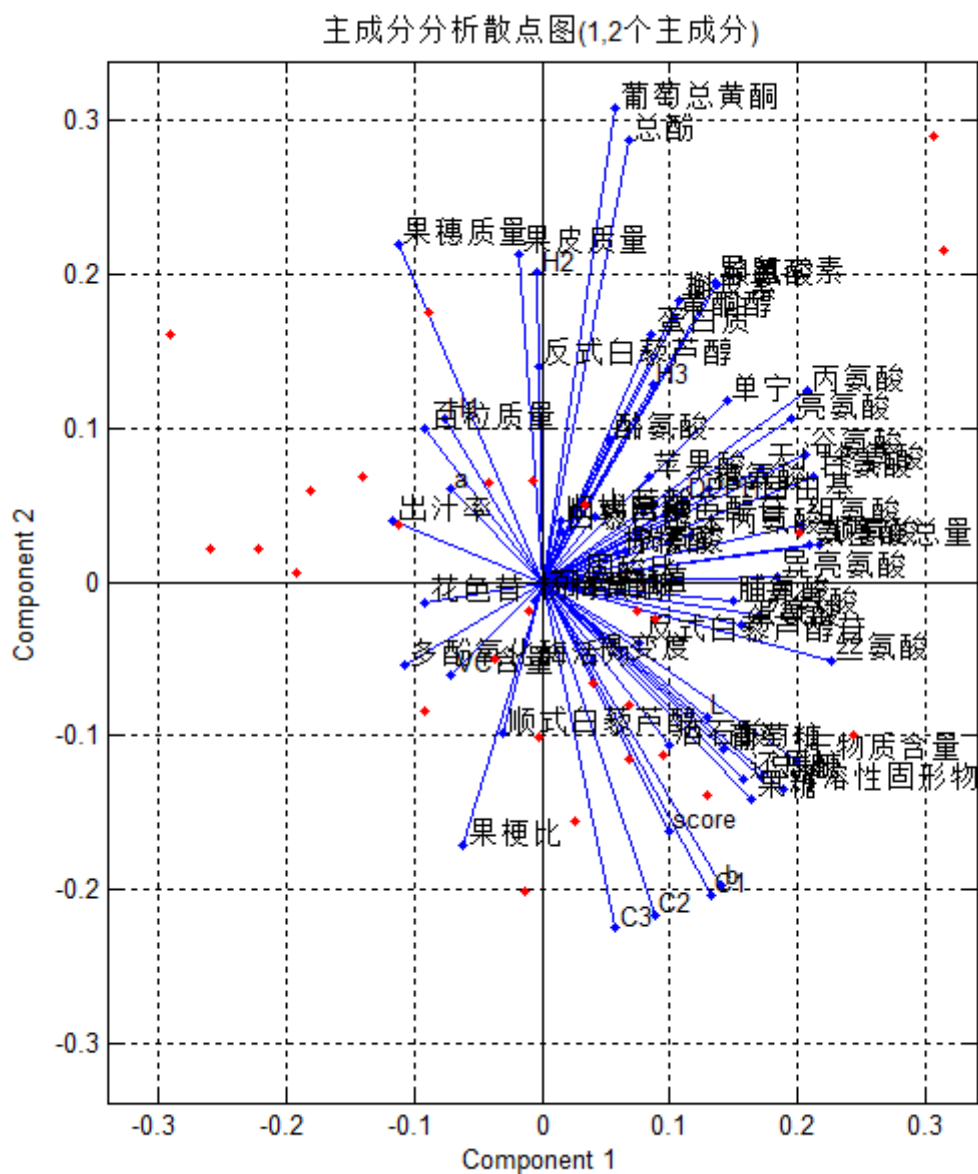


图 主成分散点图（第一二主成分）（白葡萄）

第一主成分的贡献度达到了 20%左右，多种指标对第一主成分都有影响，其中影响最大的是葡萄总黄酮和总酚，果皮颜色对第一主成分起到负的影响。氨基酸类物质，有机酸类物质，醇苷类的物质对于第二主成分正影响较大。主成分坐标下的样本数据点分布比较对称，主成分分析达到了目的。

（4）对方差标准化后和主成分分析预处理后的数据进行系统聚类分析

（1） 聚类分析中的距离：

聚类分析中常用的距离有 Minkowski 距离, Lance 距离, Mahalanobis 距离, 斜交空间距离等。在第二问系统聚类法中采用 Minkowski 距离。

第 i 个样本 x_i 与第 j 个样本 x_j 之间的 Minkowski 距离定义为:

$$d_{ij}(q) = \left[\sum_{k=1}^P (x_{ik} - x_{jk})^q \right]^{1/q}, i = 1, 2, \dots, n; j = 1, 2, \dots, n$$

其中 q 为正整数。

特别的:

当 $q = 1$ 时, $d_{ij}(1) = \sum_{k=1}^P |x_{ik} - x_{jk}|$ 称为绝对值距离。

当 $q = 2$ 时, $d_{ij}(2) = \left[\sum_{k=1}^P (x_{ik} - x_{jk})^2 \right]^{1/2}$ 称为欧氏距离。

当 $q \rightarrow \infty$ 时, 称 $d_{ij}(\infty) = \max_{1 \leq k \leq p} |x_{ik} - x_{jk}|$ 为切比雪夫距离。

在第本问的系统聚类法中采用的又是 Minkowski 距离中 $q = 2$ 的情况, 即样本间的欧式距离, Minkowski 距离不适合于变量间单位不同和测量值范围相差较大的情况, 由于预先对数据进行了预处理所以仍然可以采用 Minkowski 距离。

(2) 系统聚类的基本思想

聚类开始时将 n 个样品各自作为一类, 并规定样品之间的距离和类与类之间的距离, 然后将距离最近的两类合并成一个新类, 计算新类与其它类之间的距离, 重复进行两个最近类的合并, 每次减少一类, 直至所有的样品合并为一类。最后形成一个表示亲疏关系的动态聚类树, 从动态聚类树上面可以清晰的看出应分为几类, 以及每一类所包含的样品。以 G 表示类, 假定 G 中有 m 个元素 (即样品), 用列向量 $x_i (i = 1, 2, \dots, m)$ 来表示,

d_{ij} 表示样本 x_i 和样本 x_j 间的距离, D_{KL} 表示类 G_k 与类 G_L 之间的距离。类与类之间用不同的方法定义距离, 就产生了不同的系统聚类方法。系统聚类方法有最短距离法, 最长距离法, 中间距离法, 重心法, 类平均法, 离差平方和法。在该问题中采用离差平方和法。

离差平方和法吧方差分析的思想用在聚类上, 同一个类内的离差平方和小, 而类间的离

差平方和应当大。类中个元素到类重心的平方欧氏距离之和称为类内离差平方和。设某一步 G_k 与 G_L 聚成一个新类 G_M ，则 G_k 、 G_L 、 G_M 的类内离差平方和分别为：

$$W_k = \sum_{x_i \in G_k} (x_i - \bar{x}_k)^T (x_i - \bar{x}_k)$$

$$W_L = \sum_{x_i \in G_L} (x_i - \bar{x}_L)^T (x_i - \bar{x}_L)$$

$$W_M = \sum_{x_i \in G_M} (x_i - \bar{x}_M)^T (x_i - \bar{x}_M)$$

它们反映了类内元素的分散程度。 G_k 与 G_L 聚成一个新类 G_M ，类内的离差平方和会有所增加，即 $W_M - (W_k + W_L) > 0$ ，若 G_k 与 G_L 靠近，则增加的离差平方和较小，于是定义 G_k 与 G_L 的平方距离为：

$$D_{KL}^2 = W_M - (W_k + W_L) = \frac{n_k n_L}{n_M} (\bar{x}_k - \bar{x}_L)^T (\bar{x}_k - \bar{x}_L)$$

类间平方距离的递推公式为：

$$D_{MJ}^2 = \frac{n_J + n_K}{n_J + n_M} D_{KJ}^2 + \frac{n_J + n_L}{n_J + n_M} D_{LJ}^2 - \frac{n_J}{n_J + n_M} D_{KL}^2$$

对于红葡萄酒经过方差标准化和主成分分析后的数据采用欧式距离离差平方和法聚类结果：

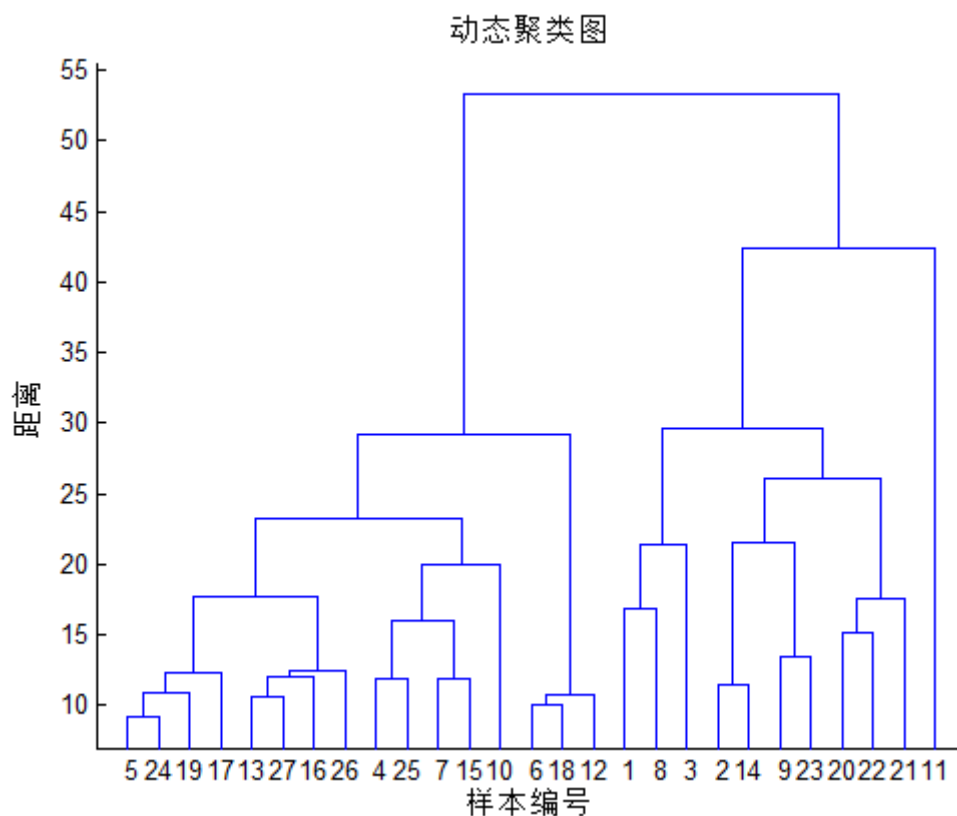


图 动态聚类树（红葡萄）

需要列表一说明分类结果：

第 1 类的有 1 3 8

第 2 类的有 2 9 14 20 21 22 23

第 3 类的有 11

第 4 类的有 4 5 6 7 10 12 13 15 16 17 18 19 24 25 26 27

结合单宁的含量，和脯氨酸（高营养价值氨基酸），和评价得分，和分类结果，不难看出 11 号是差的葡萄即第三类是差的 D；138 是中等的葡萄即第一类是中等 C；第二类是优 A，第四类是良好 B。

列表二解释这样的结果，花色苷影响酒的色泽，单宁是红酒的灵魂，影响口感，脯氨酸营养价值高。评价得分直接表示酒的感官质量。分类的结果客观可靠。

样本编号	评价得分	脯氨酸	花色苷	单宁	等级
1	84	723.8787	408.0278	22.01903	C

2	91	1560.974	224.3667	23.36131	A
3	89	7472.284	157.9394	20.37287	C
4	88	1182.233	79.68514	8.638437	B
5	89	816.0796	120.6061	14.48577	B
6	86	2932.761	46.18635	15.17299	B
7	84	1096.282	60.76652	5.619186	B
8	84	962.0112	241.3969	22.48899	C
9	94	1334.186	240.8433	24.36228	A
10	84	477.4962	44.20342	16.6877	B
11	81	1150.094	7.787321	4.54301	D
12	87	2127.908	32.34264	7.168961	B
13	86	621.2542	65.32386	9.821984	B
14	93	677.7822	140.2568	13.94054	A
15	82	817.568	52.79195	25.41701	B
16	88	679.2525	60.66019	10.08617	B
17	90	806.5634	59.42362	15.73034	B
18	88	2097.613	40.22765	5.388159	B
19	87	1438.085	115.7041	13.70002	B
20	93	754.8928	23.52304	8.114932	A
21	90	5144.812	89.28164	13.61304	A
22	89	863.9912	74.02659	12.1551	A
23	89	1341.124	172.6258	24.25694	A
24	88	797.5533	144.8807	14.41657	B
25	86	479.1744	49.64335	9.324297	B
26	88	147.7038	58.46919	3.777922	B
27	88	418.0097	34.19033	10.3096	B

运用相同的聚类步骤处理白葡萄的方差标准化后的和主成分分析后的数据：

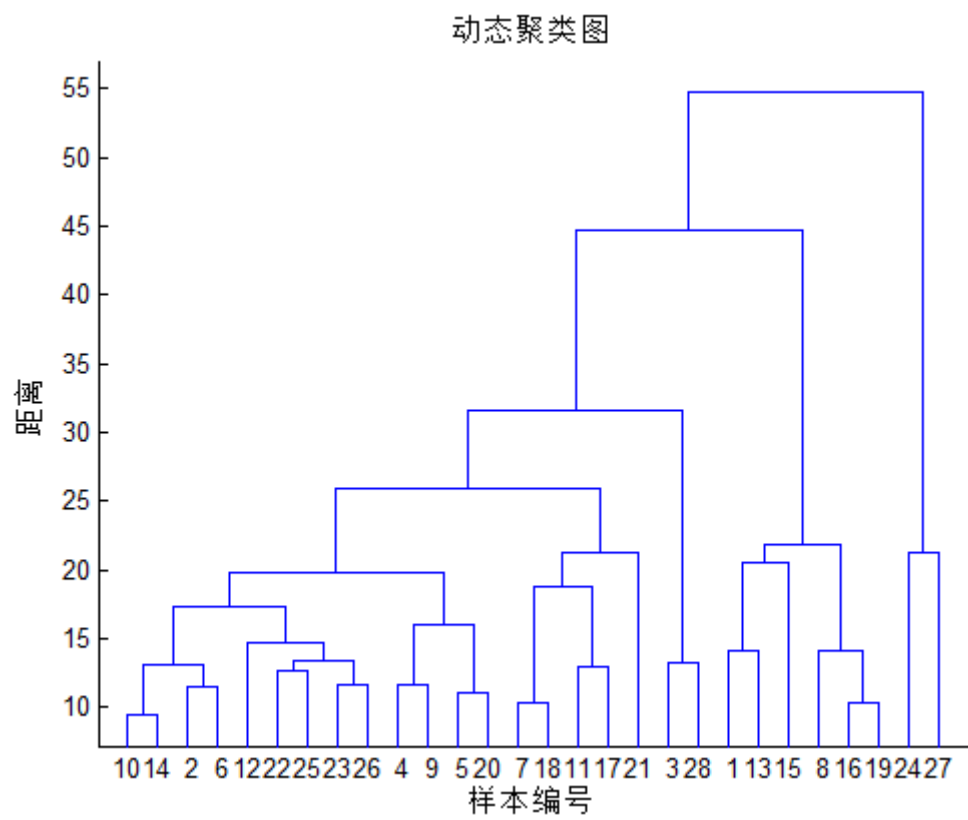


图 动态聚类树（白葡萄）

需要列表一整理聚类结果：

第 1 类的有 3 28
第 2 类的有 2 4 5 6 7 9 10 11 12 14 17 18 20 21 22 23 25 26
第 3 类的有 1 8 13 15 16 19
第 4 类的有 24 27
白葡萄好像有点模糊，分类效果不太好。但是结合得分情况可以归纳出：
第一类是差的一类即 D 类，第 4 类是优的一类即 A 类，第 2 类是良好的一类即 B 类，第 3 类是中等的一类即 C 类。

根据特征性指标类表解释聚类结果： 脯氨酸营养价值高，花色苷关系色泽，单宁也可以用影响色泽同时影响口感。

样本编号	评价得分	脯氨酸	花色苷	单宁	等级
1	94	589.1939	0.819055	2.947388	C
2	92	742.0523	0.404145	2.23914	B
3	93	3217.961	2.790123	2.990077	A
4	93	678.0074	2.031791	3.147712	B
5	96	1559.175	0.387765	2.625873	B
6	92	687.2569	0.834402	4.502007	B
7	86	701.8037	2.477212	4.728893	B
8	91	399.4266	2.864162	1.672219	C
9	94	800.5865	3.684752	4.433853	B
10	95	823.1779	0.844081	6.780594	B
11	93	702.8816	0.406982	3.311858	B
12	86	600.1397	0.408936	3.212347	B
13	88	238.4973	2.508596	2.128787	C
14	94	586.7064	0.403579	2.387905	B
15	93	1767.271	1.629477	2.750864	C
16	87	380.3584	4.10292	2.228054	C
17	97	387.6527	2.088771	2.246663	B
18	92	551.6437	0.811532	5.782615	B
19	91	287.2348	2.072079	2.217499	C
20	91	995.1164	0.829305	3.141327	B
21	95	638.1954	0.818914	1.952293	B
22	92	538.1494	1.650536	6.462577	B
23	96	702.939	1.243803	3.388694	B
24	92	812.4552	0.371996	8.506305	D
25	96	1020.181	0.670057	2.756784	B
26	92	1259.757	1.6568	5.517457	B
27	93	891.0949	0.420203	6.250788	D
28	95	2006.25	2.056614	4.583495	A

5.1 问题三模型的建立与求解

六. 模型的优缺点（模型的评价与推广）

6.1 模型的优点

- (1)
- (2)
- (3)

6.2 模型的缺点

对文本中的模型给出比较客观的评价，必须实事求是，有根据，以便阅卷人参考。推广和优化，需要挖空心思，想出合理的、甚至可以合理改变题目给小胡的条件的、不一定可行但是具有一定想象空间的准理想的方法、模型。（大胆、合理、心细。反复推敲，这段 500 字半页左右的文字，可能决定生死存亡。）

参考文献

- [1] 孙沛杰 孙丽颖,《葡萄中单宁对生产葡萄酒的影响》,《酿酒》,1998 年
- [2] 柴菊花,《昌黎产区 7 个单品种干红葡萄酒氨基酸分析》,《中国酿造》,2010 年
- [3]
- [4]
- [5]

注：5 篇以上！

附录

1.1 秩和检验差异显著性程序

difference.m
<pre>%红葡萄酒的差异显著性分析 red_1=xlsread('grade_ok.xlsx','grade_exact','A2:A28') red_2=xlsread('grade_ok.xlsx','grade_exact','B2:B28') [p,h,stats]=ranksum(red_1,red_2,0.05) figure(1) subplot(1,2,1);hist(red_1);xlabel('得分');ylabel('频数');title('第一组红葡萄酒评价结果直方图'); subplot(1,2,2);hist(red_2);xlabel('得分');ylabel('频数');title('第二组红葡萄酒评价结果直方图'); red_1ss=var(red_1) red_2ss=var(red_2) %白葡萄酒的差异显著性分析 wine_1=xlsread('grade_ok.xlsx','grade_exact','E2:E29') wine_2=xlsread('grade_ok.xlsx','grade_exact','F2:F29') [p,h,stats]=ranksum(wine_1,wine_2,0.05) figure(2)</pre>

subplot(1,2,1);hist(wine_1);xlabel('得分');ylabel('频数');title('第一组白葡萄酒评价结果直方图'); subplot(1,2,2);hist(wine_2);xlabel('得分');ylabel('频数');title('第二组白葡萄酒评价结果直方图'); wine_1ss=var(wine_1) wine_2ss=var(wine_2)	
备注	

1.2 红葡萄酒方差标准化数据箱线图绘制程序

程序名 WineRboxplot.m	
<pre> %%%%%%%%%%%%%% Box Plot Function By John Yang%%%%%%%%%%%%%% %%%%%%%%%%%%%% load data load WineR %%%%%%%%%%%%%% plot box1 figure('Color',[1 1 1]); boxplot(Y(1:27,1:16),'Orientation','horizontal','labels',names(1:16)); title('箱线图（1-16号变量）') %%%%%%%%%%%%%% plot box2 figure('Color',[1 1 1]); boxplot(Y(1:27,16:32),'Orientation','horizontal','labels',names(16:32)); title('箱线图（16-32号变量）') %%%%%%%%%%%%%% plot box3 figure('Color',[1 1 1]); boxplot(Y(1:27,32:48),'Orientation','horizontal','labels',names(32:48)); title('箱线图（32-48号变量）') %%%%%%%%%%%%%% plot box4 figure('Color',[1 1 1]); boxplot(Y(1:27,48:64),'Orientation','horizontal','labels',names(48:64)); title('箱线图（48-64号变量）') </pre>	
备注	程序用来画出方差标准化后的红葡萄酒数据的箱线图

1.3 红葡萄酒方差标准化数据主成分分析程序

程序名 WinePCA.m	
<pre> %%%%%%%%%%%%%% Principle Component Analysis Function By John Yang%%%%%%%%%%%%%% %%%%%%%%%%%%%% 导入数据并做主成分分析 load WineR [coef,score,latent,t2]=princomp(Y); percent_explain=100*latent/sum(latent); %%%%%%%%%%%%%% 绘制帕累托图 figure('Color',[1 1 1]) % pareto(percent_explain); x=1:15; kk(1)=percent_explain(1); for index=2:15 </pre>	

```

    kk(index)=kk(index-1)+percent_explain(index);
end
bar(percent_explain(1:15));
hold on
plot(x,kk,'linewidth',2);
hold off;
grid on;
title('主成分的帕累托图');
set(gca,'ytick',0:5:100);
xlabel('主成分');
ylabel('累积贡献率(%)');
axis([0.5 15.5 0 100]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第一二主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,1:2),'score',score(:,1:2),'varlabels',names);
grid on;
title('主成分分析散点图(1,2 个主成分)');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第三四主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,3:4),'score',score(:,3:4),'varlabels',names);
grid on;
title('主成分分析散点图(3,4 个主成分)');
xlabel('component 3');
ylabel('component 4');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第五六主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,5:6),'score',score(:,5:6),'varlabels',names);
grid on;
title('主成分分析散点图(5,6 个主成分)');
xlabel('component 5');
ylabel('component 6');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第七八主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,7:8),'score',score(:,7:8),'varlabels',names);
grid on;
title('主成分分析散点图(7,8 个主成分)');
xlabel('component 7');
ylabel('component 8');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第九十主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,9:10),'score',score(:,9:10),'varlabels',names);
grid on;
title('主成分分析散点图(9,10 个主成分)');
xlabel('component 9');

```


<pre> T=cluster(z,'maxclust',4); for i=1:4 tm=find(T==i); %求 dii 类的对象 tm=reshape(tm,1,length(tm)); %变成行向量 fprintf('第%d 类的有%s\n',i,int2str(tm)); %display cluster result end </pre>	
备注	程序首先作出主成分的帕累托图以根据累积贡献度保留主成分，然后作出动态聚类树和给出各个类别包含的葡萄样本。

1.5 白葡萄方差标准化数据箱线图绘制程序

程序名	WineWboxplot.m
<pre> %%%%%%%%%%%%%% Box Plot Function By John Yang%%%%%%%%%%%%%% %%%%%%%%%%%%%% load data load WineW %%%%%%%%%%%%%% plot box1 figure('Color',[1 1 1]); boxplot(Y(1:27,1:16),'Orientation','horizontal','labels',names(1:16)); title('箱线图（1-16 号变量）') %%%%%%%%%%%%%% plot box2 figure('Color',[1 1 1]); boxplot(Y(1:27,16:32),'Orientation','horizontal','labels',names(16:32)); title('箱线图（16-32 号变量）') %%%%%%%%%%%%%% plot box3 figure('Color',[1 1 1]); boxplot(Y(1:27,32:48),'Orientation','horizontal','labels',names(32:48)); title('箱线图（32-48 号变量）') %%%%%%%%%%%%%% plot box4 figure('Color',[1 1 1]); boxplot(Y(1:27,48:64),'Orientation','horizontal','labels',names(48:64)); title('箱线图（48-64 号变量）') </pre>	
备注	程序利用方程标准化后的数据画出箱线图，形象的展示出数据在各个变量上面的分布情况

1.6 白葡萄方差标准化数据主成分分析程序

程序名	WinePCA.m
<pre> %%%%%%%%%%%%%% Principle Component Analysis Function By John Yang%%%%%%%%%%%%%% %%%%%%%%%%%%%% 导入数据并做主成分分析 load WineW [coef,score,latent,t2]=princomp(Y); percent_explain=100*latent/sum(latent); %%%%%%%%%%%%%% 绘制帕累托图 figure('Color',[1 1 1]) % pareto(percent_explain); </pre>	

```

x=1:16;
kk(1)=percent_explain(1);
for index=2:16
    kk(index)=kk(index-1)+percent_explain(index);
end
bar(percent_explain(1:16));
hold on
plot(x,kk,'linewidth',2);
hold off;
grid on;
title('主成分的帕累托图');
set(gca,'ytick',0:5:100);
set(gca,'xtick',0:1:16);
xlabel('主成分');
ylabel('累积贡献率(%)');
axis([0.6 16.4 0 100]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第一二主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,1:2),'score',score(:,1:2),'varlabels',names);
grid on;
title('主成分分析散点图(1,2 个主成分)');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第三四主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,3:4),'score',score(:,3:4),'varlabels',names);
grid on;
title('主成分分析散点图(3,4 个主成分)');
xlabel('component 3');
ylabel('component 4');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第五六主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,5:6),'score',score(:,5:6),'varlabels',names);
grid on;
title('主成分分析散点图(5,6 个主成分)');
xlabel('component 5');
ylabel('component 6');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第七八主成分散点图
figure('Color',[1 1 1]);
biplot(coef(:,7:8),'score',score(:,7:8),'varlabels',names);
grid on;
title('主成分分析散点图(7,8 个主成分)');
xlabel('component 7');
ylabel('component 8');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 第九十主成分散点图
figure('Color',[1 1 1]);

```

```
biplot(coef(:,9:10),'score',score(:,9:10),'varlabels',names);
grid on;
title('主成分分析散点图(9,10 个主成分)');
xlabel('component 9');
ylabel('component 10');
pause;
close all;
```

备注	程序中画出主成分的帕累托图 and 主成分在几个维度上的散点图
----	---------------------------------

1.7 白葡萄预处理后的数据系统聚类分析程序

程序名 S Cluster.m

[illegible]

```

figure('Color',[1 1 1])
[h,t]=dendrogram(z);    %画动态聚类图
xlabel('样本编号');
ylabel('距离');
title('动态聚类图');
T=cluster(z,'maxclust',4);

for i=1:4
    tm=find(T==i);    %求 dii 类的对象
    tm=reshape(tm,1,length(tm)); %变成行向量
    fprintf('第%d 类的有%s\n',i,int2str(tm)); %display cluster result
end
pause;
close all;

```

备注	程序首先作出主成分的帕累托图以根据累积贡献度保留主成分，然后作出动态聚类树和给出各个类别包含的白葡萄样本。
----	---

1. 8.....（程序文件说明）

程序名	
程序编码	
备注	

1. 9.....（程序文件说明）

程序名	
程序编码	
备注	

2. 0.....（程序文件说明）

程序名	
程序编码	
备注	

2. 1.....（程序文件说明）

程序名	
程序编码	
备注	

2. 2.....（程序文件说明）

程序名	
程序编码	
备注	

