

全国大学生数学建模竞赛编写的 L^AT_EX 模板

摘要

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，从而确定出葡萄酒的质量。酿酒葡萄的好坏和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在一定程度上反映了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解决以下问题：

1. 分析附件 1 中两组评酒员的评价结果有无显著性差异，并指出哪一组结果更可信？
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、问题分析

三、符号定义与说明

符号定义	符号含义说明
------	--------

四、模型假设

五、模型建立与求解

5.1 问题一模型的建立与求解

数据预处理

考虑到数据录入难免出现问题，我们先利用 EXCEL 对附件一的数据进行预处理，对缺失值用该指标的其他品酒员对该项指标的打分平均值代替，异常值剔除后，同样也用其他品酒员对该项指标的打分平均值代替，得到我们所需要的数据。如下所示 **缺失值异常值处理表格**

5.2 问题二模型的建立与求解

5.2.1 主成分分析模型

step1: 数据标准化

由附件 2，酿酒葡萄的一级理化指标有 30 个，红葡萄有 27 个样本，白葡萄有 28 个样本。为消除不同量纲带来的影响，需将所有数据进行标准化处理。设第 i 个样本的第 j 个指标的观测值为 y_{ij} ，第 j 个指标的均值为 \bar{y}_j ，第 j 个指标的标准差为 s_j ，则标准化公式如下：

$$x_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j}$$

其中 x_{ij} 表示标准化后的数据，经过处理的数据符合标准正态分布，即均值为 0，标准差为 1。

step2: 求样本均值及相关矩阵

设有 n 个样本， p 个指标，经标准化处理后的原始数据矩阵为 X ，则

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = (X_1, X_2, \dots, X_p)$$

然后求解样本均值 $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)^T$ 和样本相关矩阵 R 。

step3: 计算特征值和对应的单位特征向量

计算得到 R 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_p$ ，对应的单位特征向量为

$$(\xi_1, \xi_2, \dots, \xi_p) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix}$$

则 p 个主成分的表达式为

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots\dots\dots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases}$$

step4: 选取主成分

根据累积贡献率的大小选取前 $m(m < p)$ 个主成分，选取原则为：

$$\frac{\sum_{i=1}^{m-1} \lambda_i}{\sum_{i=1}^p \lambda_i} < 80\% \sim 85\% \text{ 且 } \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} \geq 80\% \sim 85\%$$

5.2.2 相关性分析

计算所有主成分与葡萄酒评分的相关矩阵并作图

若葡萄酒评分与所有主成分均成正相关，可得到每个样品的综合得分。

5.2.3 聚类分析

将葡萄酒评分标准化后，结合综合得分可得到一个二维向量，对其进行聚类分析。

5.3 问题三模型的建立与求解

5.4 问题四模型的建立与求解

六、 模型评价与推广

6.1 模型优点

6.2 模型缺点

6.3 模型推广

参考文献

- [1] 刘海洋. \LaTeX 入门[J]. 电子工业出版社, 北京, 2013.
- [2] 全国大学生数学建模竞赛论文格式规范 (2019 年 9 月 12 日修改).