

# 红酒质量分析报告

回归分析小组

2024 年 1 月 4 日

# 目录

1	问题背景	1
2	数据说明	1
3	描述性统计	2
4	数据建模	2
5	结论及建议	2

## 1 问题背景

随着酒类产品在市场上的广泛受欢迎，对于酒的质量和特征的深入了解变得至关重要。为了更好地了解酒的品质，我们进行了一项回归分析，重点关注了酒的酸度，二氧化硫（ $SO_2$ ）含量等特征。这些特征在很大程度上影响了酒的口感、风味和保存能力。

## 2 数据说明

我们通过 UC Irvine 仓库下载了红酒数据，数据集中的特征包含酒的各类化学成分指标以及酒的品质评价。其中，成分指标，如  $pH$  值、 $SO_2$  含量、残糖量等通过物理化学检测得出；酒的品质由专业品酒师做出评价。（每个样本由三个品酒师做出评价，每个人的评分为 0（差）到 10（好）的一个整数，最终评价取三人的中位数）。红酒共 1599 条数据，其中不含缺失值，共记录了 12 个特征，特征说明如 1 所示。

表 1: 酒的特征说明

变量名	中文含义	变量类型	单位
fixed acidity	固定酸度	连续型变量	g/L
Volatile Acidity	挥发性酸度	连续型变量	g/L
Citric Acid	柠檬酸	连续型变量	g/L
Residual Sugar	残糖	连续型变量	g/L
Chlorides	氯化物	连续型变量	g/L
Free Sulfur Dioxide	游离二氧化硫	连续型变量	g/L
Total Sulfur Dioxide	总二氧化硫	连续型变量	g/L
Density	密度	连续型变量	g/mL
pH	葡萄酒的 pH 值	连续型变量	
Sulphates	硫酸盐	连续型变量	g/L
Alcohol	醇度	连续型变量	%
Quality	酒品	离散型变量	

### 3 描述性统计

对数据的初步描述如表 2 所示, 包含平均值、最小值、最大值、中位数。

表 2: 数据描述

变量名	平均值	最小值	最大值	中位数
fixed.acidity	8.32	4.60	15.90	7.90

### 4 数据建模

### 5 结论及建议