

2022 年西安明德理工学院第十三届数学建模竞赛

参赛学校： 西安明德理工学院

参赛学院： 信息工程学院

参赛题号： B 题 葡萄酒质量评价问题

参赛队号： 047

参赛队员：

1. 高帆 班级： 101082102 班 学号： 1212053

2. 张闻鑫 班级： 101082101 班 学号： 1212016

3. 韩振玮 班级： 101082101 班 学号： 1212009

日期： 2022 年 6 月 5 日

基于多元线性回归方法的葡萄酒评价问题研究

摘 要

在我们日常的生活中,对葡萄酒的分析一般是通过聘请一批有资质的评酒员进行品评,从而确定葡萄酒的质量。根据每位评酒员品尝葡萄酒后对其分类指标的打分,求其总和作为该类酒的总分,以此来确定不同葡萄酒样本的质量。酿酒葡萄的好坏一般与所酿葡萄酒的质量有直接的关联,葡萄酒的理化指标和酿酒葡萄检测的理化指标均会在一定程度上反映出葡萄酒的优劣和葡萄的质量葡萄酒的质量与酿酒葡萄的理化指标、葡萄酒的理化指标等紧密相关。本文以酿酒葡萄和葡萄酒为研究对象,利用多元线性回归方法并结合数据挖掘中的 Q-Q 图对所建立模型进行正态分布检验分析研究葡萄酒理化指标和葡萄酒质量之间的定性和定量关系。利用 matlab 对问题进行求解。

关键词: 多元线性回归方法; matlab; 正态分布; 数据挖掘; Q-Q 图

一、问题重述

葡萄酒是一种成分复杂的酒精饮料，不同种类的葡萄酒也都各具风味，葡萄酒行业评价葡萄酒质量时一般采用理化指标或者专家人工品尝的方法。而后者的方案受不同评定人员专家的嗜好、习惯、情绪、年龄、经验等因素的影响较大，评定常有一定程度的主观性和不确定性，尤其在葡萄酒质量差别不大时，更易引起打分不一致的参差和数据分析产生偏差，使品评结果不具有代表性和科学性。但是对葡萄酒质量通过量化的方法进行评价，通过不同组别的数据分析，为酿酒行业对葡萄酒进行质量分类提供理论参考成为当下时代一种趋势。葡萄酒所含有的不同成分以及不同成分的含有量与葡萄酒的质量关系密不可分，是判定葡萄酒质量的重要依据之一。

二、问题分析

该模型主要是探讨理化指标对葡萄酒质量的影响，我们不难发现，葡萄酒的质量由 11 种理化指标共同决定，即多种变量对同一因素的影响，因此，我们自然而然想到需要建立多元线性回归模型，但通过后续的计算和验证可以证明，他们的确存在某种线性关系，且拟合效果相当好，因此该模型有一定的使用价值。

三、模型假设

1. 建模仅依据赛题提供数据，不考虑其他信息；
2. 假设葡萄酒质量与各理化指标呈线性相关

四、符号说明

\bar{x}_j	均值	$x_1 \sim x_{11}$	各理化指标
S^2	方差	R^2	拟合优度
s_j	标准差	b	回归系数估计值
r	相关系数	bint	b 的置信区间
y	质量	r	残差(向量)
$Cov(X, Y)$	协方差	rint	r 的置信区间
s^2	误差方差的估计值		

五、模型建立与求解

5.1 问题一 数字特征的简单计算

问题 1 通过线性回归分析以及附件一直接计算出各数字特征的值,将附录一中的数值导入 matlab 做出质量与理化指标的数据矩阵 Physicochemical。

求解均值:

$$\bar{x}_j = \frac{\sum x_i}{n}$$

求解方差:

$$S^2 = \frac{1}{n} \left(\sum (x_i - \bar{x})^2 \right)$$

求解标准差:

$$s_j = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

在 matlab 中分别利用 mean、var、std 函数进行均值、方差、标准差的求解即可得出结果。

表 1-表 7 为不同质量下理化指标的数字特征

表 1-质量为 9 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	7.8667	0.2567	0.4333	4.9333	0.0327
S^2	1.1633	2.3333e-04	0.0044	24.0933	4.3333e-06
s_j	1.0786	0.0153	0.0666	4.9085	0.0021

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	28.6667	125.3333	0.9926	3.2833	0.4533	11.9333
S^2	4.3333	170.3333	1.4e-05	0.0072	9.3333e-04	1.8033
s_j	2.0817	13.0512	0.0038	0.0850	0.0306	1.3429

表 2-质量为 8 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	6.6393	0.2763	0.3261	5.7254	0.0384
S^2	0.6511	0.0117	0.0074	18.1312	1.7508e-04
s_j	0.8069	0.1082	0.0859	4.2581	0.0132

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	36.9480	126.5461	0.9923	3.2208	0.4882	11.6353
S^2	261.0380	1.08914728 e+03	7.8e-06	0.0231	0.0216	1.6578
s_j	16.1567	33.0022	0.0028	0.1519	0.1468	1.2875

表 3-质量为 7 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	6.7486	0.2626	0.3264	5.1856	0.0382
S^2	0.5466	0.0082	0.0062	18.5007	1.1510e-04
s_j	0.7393	0.0905	0.0787	4.3012	0.0107

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	34.1459	125.2471	0.9925	3.2125	0.5032	11.3625
S^2	175.1752	1.0766e+03	7.6e-06	0.0248	0.0168	1.5547
s_j	13.2505	32.8121	0.0028	0.1576	0.1297	1.2469

表 4-质量为 6 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	6.8265	0.2605	0.3378	6.4431	0.0452
S^2	0.6513	0.0078	0.0142	26.7054	4.1856e-04
s_j	0.8071	0.0881	0.1192	5.1677	0.0205

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	35.6615	136.9886	0.9940	3.1894	0.4910	10.5771
S^2	247.9917	1.7066e+03	9.1e-06	0.0226	0.0128	1.3188
s_j	15.7478	41.3108	0.0030	0.1502	0.1133	1.1484

表 5-质量为 5 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	6.9340	0.3020	0.3377	7.3350	0.0515
S^2	0.7030	0.0100	0.0198	28.3835	7.0202e-04
s_j	0.8384	0.1001	0.1408	5.3276	0.0262

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	36.4321	150.9046	0.9953	3.1688	0.4822	9.8088
S^2	329.2770	1.9436e+03	6.4e-06	0.0199	0.0096	0.7175
s_j	18.1460	44.0862	0.0025	0.1409	0.0982	0.8471

表 6-质量为 4 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	7.1294	0.3812	0.3042	4.6282	0.0501
S^2	1.1696	0.0301	0.0268	17.3356	6.7019e-04
s_j	1.0815	0.1735	0.1639	4.1636	0.0259

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	23.8589	125.2791	0.9943	3.1829	0.4761	10.1525
S^2	415.8071	2.7830e+03	6.0e-06	0.0268	0.0139	1.0064
s_j	20.3913	52.7538	0.0025	0.1637	0.1180	1.0032

表 7-质量为 3 时的理化指标的数字特征

	非挥发性酸含量	挥发性酸含量	柠檬酸含量	残余糖分含量	氯化钠含量
\bar{x}_j	7.5938	0.3284	0.3356	5.9719	0.0549
S^2	3.3313	0.0198	0.0065	25.7083	0.0026
s_j	1.8252	0.1405	0.0805	5.0703	0.0514

	游离二氧化硫含量	总二氧化硫含量	密度	酸碱度	硫酸钾含量	酒精浓度
\bar{x}_j	60.6563	178.6250	0.9946	3.1706	0.4862	10.3375
S^2	5.7635e+03	1.3602e+03	8.3e-06	0.0455	0.0169	1.4398
s_j	75.9174	116.6283	0.0029	0.2131	0.1300	1.1999

进一步的，我们可以计算出不同质量与理化指标的样本 Pearson 相关系数^[1]（即使相关系数接近于 1 或-1，也不一定就相关，受异常值的影响，这里在成线性关系的前提下检验才有意义），进而方便我们对于相关性的分析，通过：

$$r = \frac{Cov(X,Y)}{S_x S_y}, \text{其中样本协方差: } Cov(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

以及 matlab 中 corrcoef 函数，即可计算出 r 值。并进一步得到相关系数矩阵：

`[-0.1130 -0.1975 -0.0078 -0.0954 -0.02084 0.0075 -0.1723 -0.3050 0.0995 0.0543 0.04352]`

下文中，我们还会再次使用该相关系数矩阵。

5.1.1 相关性描述

我们已经假设葡萄酒的质量与各理化指标存在某种线性关系。我们不妨假设葡萄酒质量与各理化指标有如下关系^[2]：

$$f(x_1, x_2, \dots, x_{11}) = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_{11} x_{11}$$

我们根据原始数据矩阵绘制出质量关于各理化指标的散点图，如下图：

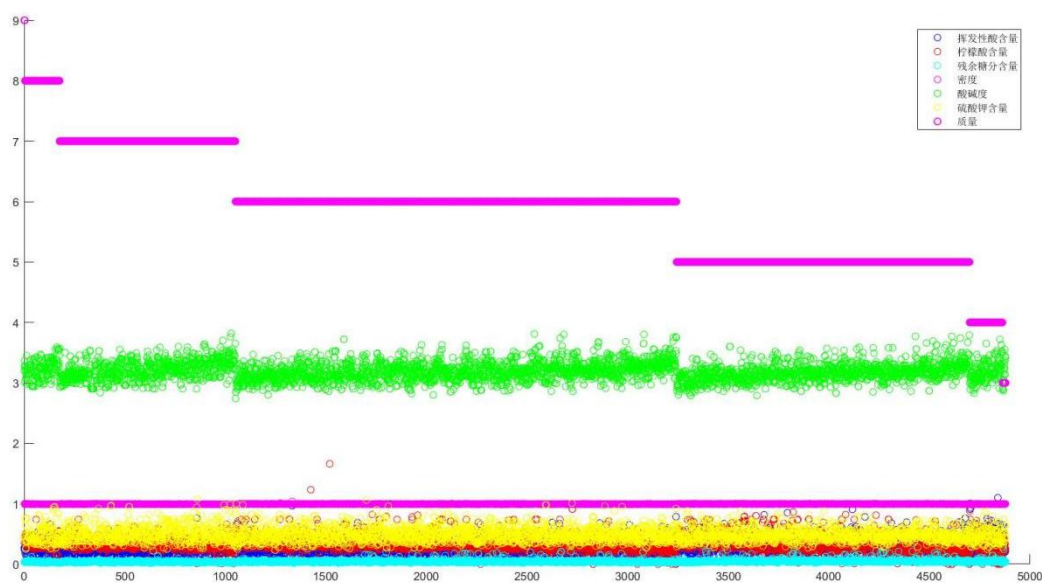


图 5.1-质量与各理化指标的散点图（1）

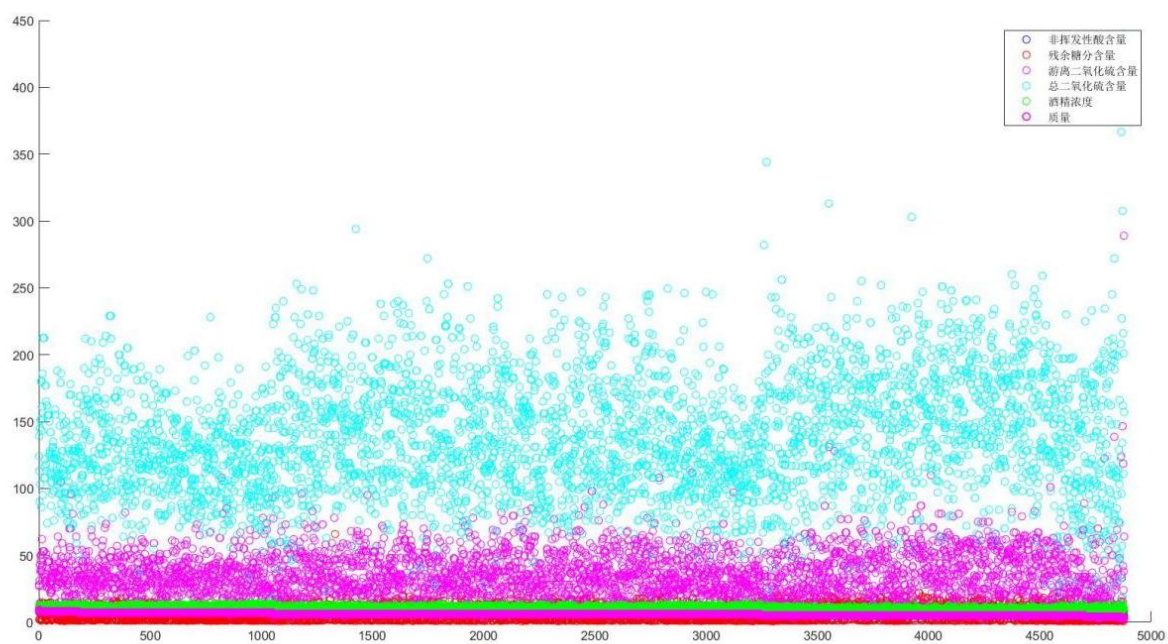


图 5.2-质量与各理化指标的散点图（2）

由于样本数量过多，我们很难从上图中看出质量与理化指标的线性关系，如果我们尝试性的进行直接进行拟合，那么会得到以下线性回归拟合曲线：

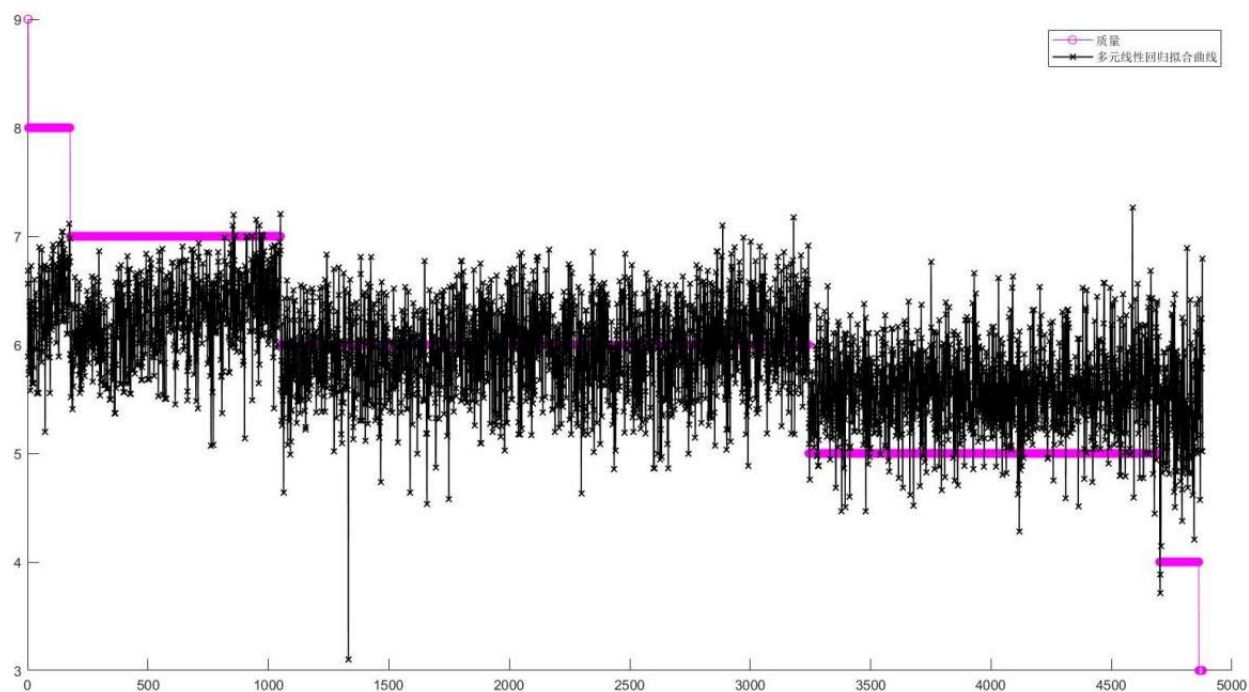


图 5.3-质量的线性回归拟合曲线

我们会得到一条混乱不堪的曲线，并可以计算出他的拟合优度为：0.28406，说明此模型的建立并不成功，我们猜测是由于各理化指标样本数量过多，因此无法很好的表现出相关性所导致的，因此我们进行的二次尝试，我们想到平均值可以作为一组数据的代表值，因此这次仅仅使用不同质量下各理化指标的平均值进行求解。根据之前计算的各指标的数量特征，可以重新创建一个各指标的平均值矩阵

Physicochemical2:

7.8667	0.2567	0.4333	4.9333	0.0327	28.6667	125.3333	0.9926	3.2833	0.4533	11.9333	9
6.6393	0.2763	0.3261	5.7254	0.0384	36.948	126.5461	0.9923	3.2208	0.4882	11.6353	8
6.7486	0.2626	0.3264	5.1856	0.0382	34.1459	125.2471	0.9925	3.2125	0.5032	11.3625	7
6.8265	0.2605	0.3378	6.4431	0.0452	35.6615	136.9886	0.994	3.1894	0.491	10.5771	6
6.934	0.302	0.3377	7.335	0.0515	36.4321	150.9046	0.9953	3.1688	0.4822	9.8088	5
7.1294	0.3812	0.3042	4.6282	0.0501	23.8589	125.2791	0.9943	3.1829	0.4761	10.1525	4
7.5938	0.3284	0.3356	5.9719	0.0549	60.6563	178.625	0.9946	3.1706	0.4862	10.3375	3

对这个矩阵，再次进行多元线性回归拟合画出质量关于各理化指标的散点图，如下图：

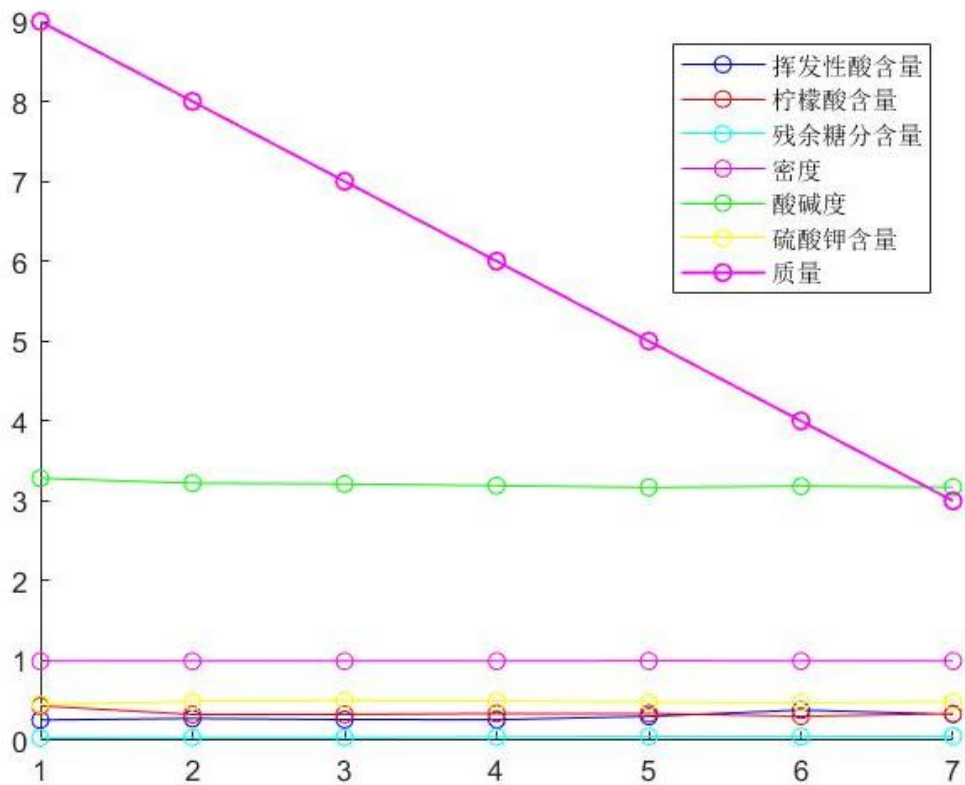


图 5.4-质量与各平均理化指标的散点图（1）

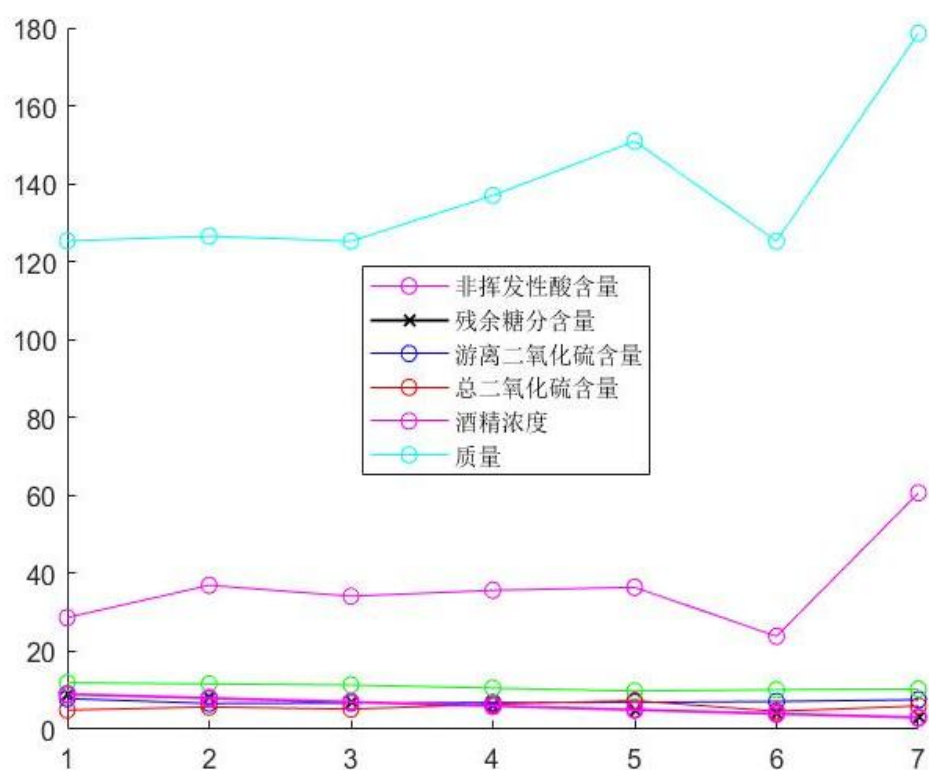


图 5.5-质量与各平均理化指标的散点图 (2)

从散点图中我们很容易就可以观察到葡萄酒的质量与各理化指标有很强的相关性，所以我们有理由推断，这十一项指标能线性合成最后的质量，于是利用最小二乘法进行拟合，得到如下曲线：

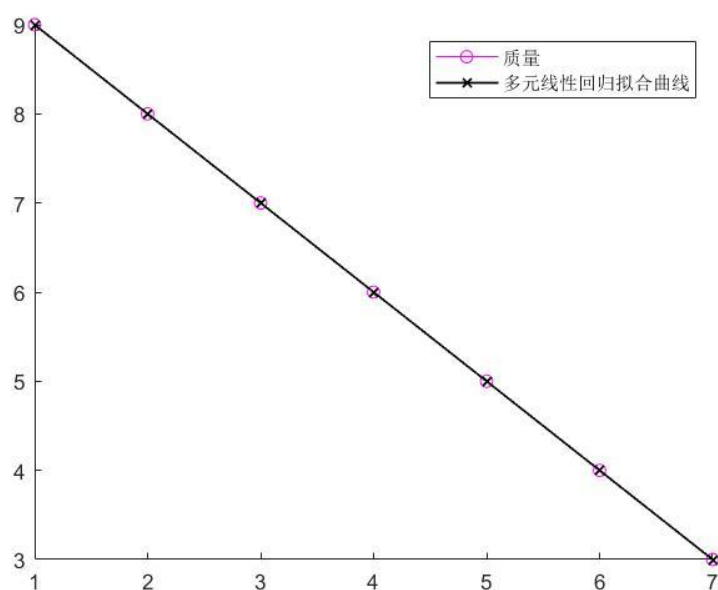


图 5.6-质量的线性回归拟合曲线

在上图中，黑色的就是多元线性回归拟合曲线。可以看出，拟合结果和附录数据完全贴近，误差不大。参数运行结果如下图：

```

b =
    0
 -0.2324
 -1.1546
    0
  0.9038
    0
 -0.2138
  0.0896
    0
-14.2222
    0
  4.0448

bint =
    0    0
 NaN NaN
 NaN NaN
    0    0
 NaN NaN
    0    0
 NaN NaN
 NaN NaN
    0    0
 NaN NaN
    0    0
 NaN NaN

r =
 1.0e-14 *
    0.7105
    0
  0.7105
    0
    0
    0
    0

rint =
 NaN NaN
 NaN NaN
 NaN NaN
 NaN NaN
 NaN NaN
 NaN NaN
 NaN NaN

stats =
    1 NaN NaN NaN

```

图 5.7-质量的线性回归拟合曲线的参数运行结果

由上图可知， $R^2=1$ ^[3]表示因变量 y(质量) 的 100%可由模型确定，F 值^[4]远远超过 F 检验的临界值，p^[5]远小于显著性水平(缺省时默认为 0.05)，因而该回归模型从整体来看是可用的。上图中列向量 b 中存放的是拟合的系数和常数项。 $b(1)$ 为常数项， b_2, b_3, \dots, b_{12} 分别是 x_1, x_2, \dots, x_{11} 的系数。得到各项理化指标与质量的最终表达式如下：

$$f(x_1, x_2, x_3, \dots, x_{11}) = 0 - 0.2324x_1 - 1.1546x_2 + 0x_3 + 0.9038x_4 + 0x_5 - 0.2138x_6 + 0.0896x_7 + 0x_8 - 14.2222x_9 + 0x_{10} + 4.0448x_{11}$$

5.1.2 探讨理化指标对葡萄酒质量的影响

从图 5.7 只能够说明所有指标的共同影响与葡萄酒质量呈很强的相关性，但无法看出单个指标的对于质量的影响，但是在问题 5.1 中，我们已经得到了各指标对质量影响的相关系数矩阵：

```
[-0.1130 -0.1975 -0.0078 -0.0954 -0.02084 0.0075 -0.1723 -0.3050 0.0995 0.0543 0.04352]
```

通过比较各相关系数的相对大小，就可以得出影响红酒质量的主要为酒精浓度，挥发性酸与柠檬酸。

5.2 问题二 利用数据挖掘的方法论证能否用理化指标来评价葡萄酒的质量

在这里，我们采用数据挖掘中的 Q-Q 图^[6]来验证前面检验相关系数时，前提除了要成线性关系，还要服从正态分布。由于样本量过于庞大，我们选择 Q-Q 图对样本进行正态分布检验，得到如下结果：

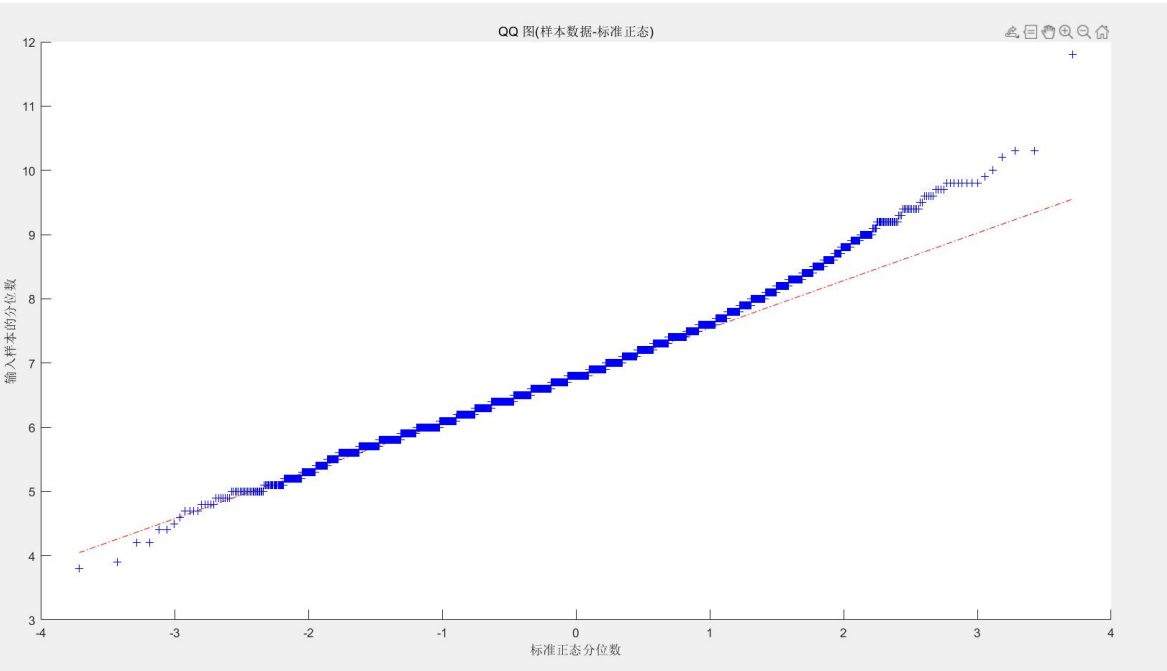


图 5.8-样本 Q-Q 图

发现图像几乎在一条直线上，则近似于正态分布，因此我们的拟合是合理的，说明的确可以用理化指标来评价葡萄酒的质量。

因此通过最终得出的表达式，很容易可以通过附录 2 估算出不同理化指标下的葡萄酒质量。这里采用 Excel 的函数模块可以更加快速高效的得出我们想要的结果

M2													fx		=B2*(-0.2324)-1.1516*C2+0.9838*E2-0.2138*G2+0.0896*H2-14.2222*I2+4.0448*L2	
A	B	C	D	E	F	G	H	I	J	K	L	M	N			
序号	非挥发性酸	挥发性酸	柠檬酸含量	残余糖分	氯化钠含量	游离二氧化硫	总二氧化硫	密度	酸碱度	硫酸钾含量	酒精浓度					
1	8.5	0.26	0.21	16.2	0.074	41	197	0.998	3.02	0.5	9.8	19.23614				
2	5.8	0.24	0.44	3.5	0.029	5	109	0.9913	3.53	0.43	11.7	7.63619				
3	9.1	0.59	0.38	1.6	0.066	34	182	0.9968	3.23	0.38	8.5	6.26089				
4	7.1	0.32	0.32	11	0.038	16	66	0.9937	3.24	0.4	11.5	8.73132				
5	6.9	0.36	0.34	4.2	0.018	57	119	0.9898	3.28	0.36	12.7	5.309768				
6	6.6	0.36	0.29	1.6	0.021	24	85	0.98965	3.41	0.61	12.4	3.768282				
7	8.2	0.37	0.36	1	0.034	17	93	0.9906	3.04	0.32	11.7	7.4389				
8	8.2	0.37	0.36	1	0.034	17	93	0.9906	3.04	0.32	11.7	7.4389				
9	4.8	0.21	0.21	10.2	0.037	17	112	0.99324	3.66	0.48	12.2	9.371312				
10	4.8	0.17	0.28	2.9	0.03	22	111	0.9902	3.38	0.34	11.3	4.418932				
11	4.8	0.26	0.23	10.6	0.034	23	111	0.99274	3.46	0.28	11.5	8.347932				
12	4.8	0.13	0.32	1.2	0.042	40	98	0.9898	3.42	0.64	11.8	9.232848				
13	4.7	0.455	0.18	1.9	0.036	33	106	0.98746	3.21	0.83	14	5.6691				
14	4.4	0.54	0.09	5.1	0.038	52	97	0.99022	3.41	0.4	12.2	6.795414				
15	14.2	0.27	0.49	1.1	0.037	33	156	0.992	3.15	0.54	11.1	4.490718				
16	10.7	0.22	0.56	8.2	0.044	37	181	0.998	2.87	0.68	9.5	5.242014				
17	10.7	0.22	0.56	8.2	0.044	37	181	0.998	2.87	0.68	9.5	5.242014				
18	10	0.2	0.39	1.4	0.05	19	152	0.994	3	0.42	10.4	7.77932				
19	10	0.23	0.27	14.1	0.033	45	166	0.9988	2.72	0.43	9.7	7.085488				
20	9.9	0.49	0.23	2.4	0.087	19	115	0.9948	2.77	0.44	9.4	4.363502				

图 5.9-附录二样本的估算结果

六、模型评价与推广

本文仍然存在很多不足，可能还有很多可以改进的地方。一方面，对于多元线性回归模型，数据使用过于模糊化，直接使用平均值使样本数量大幅度减少，导致产生一定误差，对模型的估计结果会产生一定影响，从而影响模型的正判率。另一方面，本文将拟合优度作为模型效果评价指标存在不足。因为我们所拟合出来的表达式所用样本数量过少，数据本身存在误判率，本文用拟合优度对模型进行评价建立在原始数据的分类是正确的假设前提之下。若该假设不成立，则用拟合优度作为模型效果评价指标是不准确的，应该使用一种更好的指标来评价模型。

通过预览数据我们知道红酒各个性质的数据类型，且无空值，意味着我们不必进行空值处理。得到了数据的简单的描述性统计，让我们对数据有了一个大体上的印象。

此次我们分析的目的是找出与红酒质量相关性最大的因素，即红酒的那些性质最有可能最终影响红酒的质量评估。

七、参考文献

- [1] mango660.,《相关系数》, <https://blog.csdn.net/mango660/article/details/>, 2022. 6. 4
- [2] 韩中庚, 数学建模方法及其应用[M]. 北京: 高等教育出版社, 2005, 148-150
- [3] 荔菲骞泽 Dj, 《如何用 stata 求误差方差 s^2 的估计值》, <https://zhidao.baidu.com/question/397143810094044805.html>, 2022. 6. 4
- [4] 尚轶伦, 《F 检验》, <https://baike.baidu.com/item/F%E6%A3%80%E9%AA%8C/9910842?>, 2022. 6. 4
- [5] 杜强, 《P 值》, <https://baike.baidu.com/item/P%E5%80%BC>, 2022. 6. 4
- [6] 王学仁, 实用多元统计分析[M], 上海: 上海科技出版社, 1990: 98-103

附录

附录 1 问题一 数字特征的简单计算 Matlab 程序

```
clear all
clc
%不同质量下理化指标的数字特征
Physicochemical=readmatrix('附件 1 葡萄酒数据.csv');

P1=Physicochemical(1:3,1:11);
P1_mean_col_9=mean(P1); %为列向量的均值
P1_std_col_9=std(P1); %求列向量标准差
P1_var_col_9=var(P1); %求列向量方差

P2=Physicochemical(4:176,1:11);
P2_mean_col_8=mean(P2);
P2_std_col_8=std(P2);
P2_var_col_8=var(P2);

P3=Physicochemical(177:1050,1:11);
P3_mean_col_7=mean(P3);
P3_std_col_7=std(P3);
P3_var_col_7=var(P3);

P4=Physicochemical(1051:3242,1:11);
P4_mean_col_6=mean(P4);
P4_std_col_6=std(P4);
P4_var_col_6=var(P4);

P5=Physicochemical(3243:4699,1:11);
P5_mean_col_5=mean(P5);
P5_std_col_5=std(P5);
P5_var_col_5=var(P5);

P6=Physicochemical(4700:4862,1:11);
P6_mean_col_4=mean(P6);
P6_std_col_4=std(P6);
P6_var_col_4=var(P6);

P7=Physicochemical(4863:4878,1:11);
P7_mean_col_3=mean(P7);
P7_std_col_3=std(P7);
```



```
P7_var_col_3=var(P7);
```

附录 2 问题一 – 相关系数矩阵

%相关系数的计算

```
y=Physicochemical(1:4878,12:12);y=y(:)';  
X=Physicochemical(1:4878,1:11);  
X1=X(1:4878,1:1);x1=X1(:)';  
X2=X(1:4878,2:2);x2=X2(:)';  
X3=X(1:4878,3:3);x3=X3(:)';  
X4=X(1:4878,4:4);x4=X4(:)';  
X5=X(1:4878,5:5);x5=X5(:)';  
X6=X(1:4878,6:6);x6=X6(:)';  
X7=X(1:4878,7:7);x7=X7(:)';  
X8=X(1:4878,8:8);x8=X8(:)';  
X9=X(1:4878,9:9);x9=X9(:)';  
X10=X(1:4878,10:10);x10=X10(:)';  
X11=X(1:4878,11:11);x11=X11(:)';  
[R,P]=corrcoef(x1,y);  
[R,P]=corrcoef(x2,y);  
[R,P]=corrcoef(x3,y);  
[R,P]=corrcoef(x4,y);  
[R,P]=corrcoef(x5,y);  
[R,P]=corrcoef(x6,y);  
[R,P]=corrcoef(x7,y);  
[R,P]=corrcoef(x8,y);  
[R,P]=corrcoef(x9,y);  
[R,P]=corrcoef(x10,y);  
[R,P]=corrcoef(x11,y);
```

附录 3 问题一

图 5.1-5.2-质量与各理化指标的散点图 (1) (2)

图 5.4-5.5-质量与各平均理化指标的散点图 (1) (2)

```
figure(1);  
hold on;  
plot(x1,'bo');  
plot(x4,'ro');  
plot(x6,'mo');  
plot(x7,'co');  
plot(x11,'go');  
plot(y,'mo','LineWidth',1);  
legend('非挥发性酸含量','残余糖分含量','游离二氧化硫含量','总二氧化硫含量','酒精浓度','质量')
```

```

figure(2);
hold on;
plot(x2, 'bo');
plot(x3, 'ro');
plot(x5, 'co');
plot(x8, 'mo');
plot(x9, 'go');
plot(x10, 'yo');
plot(y, 'mo', 'LineWidth', 1);
legend('挥发性酸含量', '柠檬酸含量', '残余糖分含量', '密度', '酸碱度', '硫酸钾含量', '质量')

```

附录 4

图 5.3-质量的线性回归拟合曲线

图 5.6-质量的线性回归拟合曲线

$$y_m = b(1) + b(2) \cdot X1 + b(3) \cdot X2 + b(4) \cdot X3 + b(5) \cdot X4 + \dots \\ b(6) \cdot X5 + b(7) \cdot X6 + b(8) \cdot X7 + b(9) \cdot X8 + b(10) \cdot X9 + \dots \\ b(11) \cdot X10 + b(12) \cdot X11$$

```

figure(3);
hold on;
plot(y, 'mo-');
plot(y_m, 'kx-', 'LineWidth', 1);
legend('质量', '多元线性回归拟合曲线')
R_2 = 1 - sum( (y_m - y).^2 ) ./ sum( (y - mean(y)).^2 );
str = num2str(R_2);
disp(['拟合优度为: ', str])

```

附录 5 图 5.7-质量的线性回归拟合曲线的参数运行结果

```

y=[y'];
x=[ones(size(x1')), x1', x2', x3', x4', x5', x6', x7', x8', x9', x10', x11'];
[b, bint, r, rint, stats]=regress(y, x)
len = length(y)
pelta = ones(len, 1);
x = [pelta, X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11];
y_m = b(1) + b(2) \cdot X1 + b(3) \cdot X2 + b(4) \cdot X3 + b(5) \cdot X4 + \dots \\ b(6) \cdot X5 + b(7) \cdot X6 + b(8) \cdot X7 + b(9) \cdot X8 + b(10) \cdot X9 + \dots \\ b(11) \cdot X10 + b(12) \cdot X11;

```

附录 6 图 5.8-样本 Q-Q 图

```
qqplot(X(:, 1))
```