基于 **XXXXX**XX 的葡萄酒评价

August 24, 2022

摘要

摘要部分.....

关键字：xxxxxxx

1 问题背景与重述

1.1 问题背景

葡萄酒是当今世界上最畅销的酒类之一，在各种场合都有葡萄酒的身影。然而葡萄酒的酿造是取决于多种因素，各种因素的叠加会导致葡萄酒的品质差异明显。在不同的原材料已经酿制方法的差别下葡萄酒会继续细分，例如红葡萄酒和白葡萄酒等。对各种葡萄酒的鉴别是必不可少的一个步骤，而采用人工品尝打分和采用仪器进行理化指标的检验已成为最为科学的鉴别方法。最后经过安全检查、筛选分级的葡萄酒方可上市成为饮用酒。

1.2 题目所给信息及参数

此次比赛是根据 10 位品酒员为 27 款红酒和 28 款白葡萄酒的打分，以及上述葡萄酒的指标情况和芳香物质为基础进行数学分析和建模，并探讨品酒员的打分是否合理以及论证是否科学分级。红葡萄酒和白葡萄酒的市场在国际上的价值非常之高，葡萄酒依旧是未来的主力酒类，对此分析依旧存在价值。现在根据三个数据文件，并对三个数据进行分析处理后描述统计，完成数学建模和预测。

数据一：葡萄酒品尝评分表；数据二：指标总表；数据三：芳香物质。

1.3 所需解决的问题

1) 根据附件所提供的两组品酒员对 27 款红葡萄酒和 28 款白葡萄酒的打分判断两组结果是否有显著性差异，并判断哪一组的更加可信。

2)

3) 分析酿酒葡萄和葡萄酒的理化指标之间是否具有相关性，以及其之间具有什么样的联系。

4)

2 问题分析

针对不同的国家，地区和相对应的医疗水平进行对应的数据指标分析。主要分析感染率，病人接触率，治愈率，以及传染期接触数。模型构建还需要考虑到新冠肺炎的无症状感染者这一特殊的情况，根据这些指标进行相轨线分析，合理的进行疫情的分析和未来疫情走向以及各地区、国家的防疫政策研究。

2.1 问题一的分析

第一，根据附件 1 中给出两组品酒员的打分情况判断两组的打分是否有显著性差异。对于此问题，分析附件一所提供的数据，研究发现两组对红葡萄酒和白葡萄酒的打分情况是两两相互比较和配对，适合于先进行单样本 K-S 检验判断数据是否符合正态分布，再进行两配对样本 T 检验进行显著性差异判断的办法。

第二，判断两组品酒员的打分情况哪一组更加可信。对于此问题，分析两组品酒员的打分情况，检验两组中打分的更稳定的一方，越稳定的分数即代表品酒员偏好更少，更加可信。提取两组品酒员对于白葡萄酒和红葡萄酒的分数的标准差，根据标准差的大小进行可信性的判断。

2.2 问题二的分析

2.3 问题三的分析

首先要参照附件所给的数据来进行分析酿酒葡萄和葡萄酒之间的相关性，附件的信息内容过多，需要进行合理的过滤和筛选数据，但要尽可能保证其数据的完整性和真实性。我们采取相关性分析，依据相关性皮尔森系数来判断葡萄酒的数据和酿酒葡萄的理化指标之间的相关性显著程度。

在进行了相关性分析后，可以确定下一些具有显著相关的数据流，要进一步解决其葡萄酒某指标与该些理化指标的关系，则要进行其关系的拟合，从而得出实质性的结论来判断酿酒葡萄和葡萄酒之间的关系，以及其关系的可靠性。

2.4 问题四的分析

3 符号说明

4 模型假设

5 模型建立与求解

5.1 问题一的求解

问题一分析两组评酒员的评价结果有无显著性差异，并判断两组结果哪一组更加可信。采用三个步骤完成分析，步骤如下：

1) 判断数据是否符合正态分布，以选择合适模型；

2) 使用两配对 T 检验方法完成显著性检验的判断；

3) 计算标准差的大小后进行比较，较小的表示稳定性更高，更加可信。

5.1.1 数据的预处理

因为数据较大，指标较多，所以我们对各项分数相加得到总分，接着取平均值进行比较。

均值计算如下:

$$x = \sum_{x_{mn}}^{10} (m = 1, 2, 3, , , 10 n = 1, 2, 3, , , 10) \tag{1}$$

5.1.2 各葡萄酒样本评分数据概率分布的确定

对两组品酒员差异性评价的假设检验一般要求数据符合正态分布，因为两配对样本 T 检验的前提要求为数据符合正态分布，才可以使用 T 检验的数学模型。利用 SPSS 统计软件中单样本 K-S 检验，对数据集两组品酒员分别对红、白葡萄酒品尝得到的四组评价结果进行了正态分布检验。

从图 1 和图 2 可以看出两组的双边检验结果。因此可以认为品酒员对葡萄酒的评分服从正态分布。

5.1.3 两组评价结果的显著性差异评价

上述检验显示各类葡萄酒得分情况属于正态总体，为了进一步说明品酒员评分的科学性以及两个评分组评分的可信度，需要检查两组给出的评分是否有显著性差异，即对数据进行显著性检验。

两配对样本非参数检验一般用于同一研究对象分别给予两种不同处理的效果比较。因为两组品酒员分别对同一样本组进行评分，故两组数据为配对数据。

$$z_{li} = w_{li} - w_{2i} (i = 1, 2, , , , n) \tag{2}$$

$z_l i$ 来自正态分布, 用假设检验的方法, 假设 $H_0 : u_1 = 0$ 成立;

$$\begin{cases} \bar{z} = \frac{1}{n} \sum_{i=1}^{n} (Z_l i) \\ s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_l i - z_1)^2 \\ t = \frac{\bar{z_1}}{\frac{s_1}{\sqrt{n}}} \\ w = \mid t \mid \geq t_{1-\frac{a}{2}} (n-1) \end{cases}$$

对于统计量 t, 在给定显著性水平 a 下, 该检验问题的拒绝域是 w, 若$\mid t \mid \geq t_{1-\frac{a}{2}} (n-1)$,

组数	样本数	平均值	标准差	T_1 值	p
一	27	73.056	3.9780	2.458	0.0104
二	27	70.515	7.3426	2.458	0.0104

上表给出了两组红葡萄酒评分均值的 t 检验结果，通过查表当 x =0.05，n=27 时，$t_{1-\frac{a}{2}}(n-1) = 2.0555 < 2.491$ 且方差齐性检验的 p 值为 0.0104<0.05，所以拒绝原假设，对于红葡萄酒的评价，两组评酒员的评价结果有显著性差异。因为第二组评酒员对红葡萄酒样品评分的标准差大于第一组的，第二组各评酒员得评分差异小，稳定性高，比较可信。

对于白葡萄酒采用同样的方法，得到了如下的表格：

组数	样本数	平均值	标准差	T_1 值	p
一	28	74.261	5.2012	-2.184	0.01892
二	28	76.532	3.1709	-2.184	0.01892

上表给出了两组红葡萄酒评分均值的 t 检验结果，通过查表当 $x = 0.05, n = 28$ 时，$t_{1-\frac{\alpha}{2}}(n-1) = 2.0555 < 2.491$ 且方差齐性检验的 p 值为 0. 01892<0.05，所以拒绝原假设，对于白葡萄酒的评价，两组评酒员的评价结果有显著性差异。因为第一组评酒员对红葡萄酒样品评分的标准差大于第二组，第二组各评酒员得评分差异小，稳定性高，比较可信。

综上分别对两组葡萄酒进行 t 检验，在显著性水平为 0.05 时，得出两组评酒员的评价结果有显著性差异, 第二组评酒员的评分更可信。

5.2 问题二的求解

5.3 问题三的求解

5.3.1 数据筛选

所给出酿酒葡萄和葡萄酒的数据十分多，但不能保证没一项指标都有一项对应的指标与他有着显著的相关性，所以需要进行数据指标的筛选，使得其相关性分析，更加可信。在本问中，抽选葡萄酒的花色苷、DPPH 半抑制体积、酒总黄酮、色泽这几项数据来进行数据相关性分析。

5.4 相关性分析

相关分析是描述两个变量间关系的密切程度，由相关系数和显著性程度值表示，当相关系数的绝对值越接近于 1，则表示两个变量间的相关性越显著，或者显著性 *p<0.05,**p<0.01 具有上述的效果。双变量系数测量的主要指标有卡方类测量、Spearman 相关系数、pearson 相关系数等，由于酿酒葡萄和葡萄酒的数据为定距数据，则在进行两者间的相关性检验时用 pearson 相关系数来判断，其公式为：

$$r = \frac{\sum(x_i - \bar{x})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{3}$$

相关系数 r 的取值范围为：-1≤ $r \leq 1$

$$\begin{cases} r > 0 \quad r < 0 \\ \mid r \mid = 0 \\ \mid r \mid = 1 \end{cases}$$

其中皮尔森简单相关系数检验统计为:

$$t = \frac{rsqrtn - 1}{\sqrt{1 - r^2}} \tag{4}$$

其中的 t 服从 n-2 个自由度的 t 分布

5.4.1 相关性检验

通过将筛选出的数据通过 spss 来进行相关性的检验分析，来发现针对某些特定的葡萄酒指标，有的酿酒葡萄在某个理化指标较为突出的情况下，可以根据需求来进行指定行的处理来满足要求。如下将一次选取花色苷、DPPH 半抑制体积、酒总黄酮、色泽这几项数据来进行数据相关性分析。通过对数据进行平均值和标准差统计，成对排除个案缺失值，采用 pearson 相关系数的双侧显著性检验来获取结果。如下为结果图。

相关性

		花色苷(mg/L)	多酚氧化酶活力	褐变度	DPPH自由基1/IC50 (g/L)	酒石酸 (g/L)	苹果酸 (g/L)	柠檬酸 (g/L)	黄酮醇(mg/kg)	单宁(mmol/L)	氨基酸总量	聚酚总类酶(mmol/kg)	白藜芦醇(mg/kg)	总糖g/L	PH值	可滴定酸(g/l)	固酸比	干物质含量g/100g	果糖质量/g	百粒质量/g	果树比(%)	出汁率(%)	聚点质量(g)
花色苷(mg/L)	皮尔逊相关性	1	.486**	.769**	.567**	.034	.692**	.380	.408*	.745**	.106	.442*	-.035	.005	-.039	-.215	.333	.230	-.105	-.263	.496**	.320	-.055
	显著性（双尾）		.010	< .001	.002	.867	< .001	.051	.034	< .001	.598	.021	.863	.979	.852	.282	.089	.249	.602	.186	.008	.103	.785
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
多酚氧化酶活力	皮尔逊相关性	.486*	1	.554**	.098	-.127	.344	.194	.303	.156	-.103	-.038	-.218	-.014	-.311	-.058	.127	.262	.010	-.161	.279	.038	.195
	显著性（双尾）	.010		.003	.628	.528	.079	.332	.125	.438	.608	.851	.274	.944	.114	.774	.529	.187	.960	.422	.159	.849	.329
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
褐变度	皮尔逊相关性	.769**	.554**	1	.300	-.060	.645**	.394*	.423*	.449*	-.154	.240	.033	-.141	-.248	-.288	.400*	.082	-.269	-.391*	.502**	.202	-.068
	显著性（双尾）	< .001	.003		.128	.768	< .001	.042	.028	.019	.442	.229	.871	.483	.213	.146	.039	.684	.175	.044	.008	.311	.735
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
DPPH自由基1/IC50（g/L）	皮尔逊相关性	.567**	.098	.300	1	.276	.052	-.003	.428*	.753**	.106	.837**	.211	-.020	.376	-.344	.170	-.091	.006	-.136	.467**	.465*	-.061
	显著性（双尾）	.002	.628	.128		.163	.795	.987	.026	< .001	.600	< .001	.290	.922	.053	.079	.397	.653	.975	.499	.010	.015	.763
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
酒石酸 (g/L)	皮尔逊相关性	.034	-.127	-.060	.276	1	-.111	.515**	.143	.281	.364	.247	.166	.232	.275	-.244	.260	.241	-.025	-.390*	.141	.084	-.268
	显著性（双尾）	.867	.528	.768	.163		.581	.006	.478	.155	.062	.214	.408	.245	.165	.220	.191	.226	.900	.044	.483	.676	.177
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
苹果酸 (g/L)	皮尔逊相关性	.692**	.344	.645**	.052	-.111	1	.423*	.056	.299	-.072	.052	.110	.093	-.366	-.061	.307	.299	-.257	-.308	.235	.097	-.263
	显著性（双尾）	< .001	.079	< .001	.795	.581		.028	.782	.130	.720	.796	.586	.644	.060	.761	.119	.130	.195	.118	.237	.629	.185
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
柠檬酸 (g/L)	皮尔逊相关性	.380	.194	.394*	-.003	.515**	.423*	1	.223	.145	.083	-.059	.200	-.002	-.199	-.279	.219	.239	-.061	-.383*	.135	-.053	-.132
	显著性（双尾）	.051	.332	.042	.987	.006	.028		.264	.469	.681	.772	.318	.993	.320	.158	.272	.230	.761	.049	.503	.793	.510
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
黄酮醇(mg/kg)	皮尔逊相关性	.408*	.303	.423*	.428*	.143	.056	.223	1	.580**	.465*	.263	.183	.055	.052	-.139	.134	.215	-.201	-.293	.633**	.045	.118
	显著性（双尾）	.034	.125	.028	.026	.478	.782	.264		.002	.014	.185	.361	.786	.796	.490	.507	.281	.314	.138	< .001	.825	.558
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
单宁(mmol/L)	皮尔逊相关性	.745**	.156	.449*	.753**	.281	.299	.145	.580**	1	.498**	.685**	.049	.363	.222	-.066	.261	.414*	-.271	-.330	.474*	.352	-.124
	显著性（双尾）	< .001	.438	.019	< .001	.155	.130	.469	.002		.008	< .001	.810	.063	.265	.743	.189	.032	.171	.093	.013	.072	.539
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
氨基酸总量	皮尔逊相关性	.106	-.103	-.154	.106	.364	-.072	.083	.465*	.498**	1	.129	.051	.507**	.234	.206	-.012	.601**	-.383*	-.366	.264	-.009	-.322
	显著性（双尾）	.598	.608	.442	.600	.062	.720	.681	.014	.008		.521	.800	.007	.241	.303	.954	< .001	.048	.061	.184	.965	.101
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
儿茶素总类酶（mmol/kg）	皮尔逊相关性	.442*	-.038	.240	.837**	.247	.052	-.059	.263	.685**	.129	1	.020	.210	.239	-.178	.231	.067	-.090	-.071	.259	.568**	-.016
	显著性（双尾）	.021	.851	.229	< .001	.214	.796	.772	.185	< .001	.521		.920	.293	.230	.374	.247	.739	.656	.726	.191	.002	.936
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
白藜芦醇(mg/kg)	皮尔逊相关性	-.035	-.218	.033	.211	.166	.110	.200	.183	.049	.051	.020	1	-.074	.173	-.410*	.093	-.142	-.192	-.346	.220	-.085	-.392*
	显著性（双尾）	.863	.274	.871	.290	.408	.586	.318	.361	.810	.800	.920		.713	.387	.033	.644	.479	.337	.077	.269	.674	.043
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
总糖g/L	皮尔逊相关性	.005	-.014	-.141	-.020	.232	.093	-.002	.055	.363	.507**	.210	-.074	1	.166	.196	.154	.804**	-.274	-.259	-.212	.211	-.159
	显著性（双尾）	.979	.944	.483	.922	.245	.644	.993	.786	.063	.007	.293	.713		.407	.326	.444	< .001	.167	.191	.288	.291	.427
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
PH值	皮尔逊相关性	-.039	-.311	-.248	.376	.275	-.366	-.199	.052	.222	.234	.239	.173	.166	1	-.505**	.351	-.056	.121	-.036	.018	.283	-.184
	显著性（双尾）	.852	.114	.213	.053	.165	.060	.320	.796	.265	.241	.230	.387	.407		.007	.072	.783	.549	.857	.929	.152	.358
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
可滴定酸（g/l）	皮尔逊相关性	-.215	-.058	-.288	-.344	-.244	-.061	.279	-.139	-.066	.206	-.178	-.410*	.196	-.505**	1	-.650**	.341	-.075	.227	-.285	-.056	.276
	显著性（双尾）	.282	.774	.146	.079	.220	.761	.158	.490	.743	.303	.374	.033	.326	.007		< .001	.082	.709	.255	.150	.780	.163
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
固酸比	皮尔逊相关性	.333	.127	.400*	.170	.260	.307	.219	.134	.261	-.012	.231	.093	.154	.351	-.650**	1	.023	.035	-.211	.061	.112	-.168
	显著性（双尾）	.089	.529	.039	.397	.191	.119	.272	.507	.189	.954	.247	.644	.444	.072	< .001		.908	.861	.292	.763	.577	.402
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
干物质含量g/100g	皮尔逊相关性	.230	.262	.082	-.091	.241	.299	.239	.215	.414*	.601**	.067	-.142	.804**	-.056	.341	.023	1	-.401*	-.368	-.055	.119	-.159
	显著性（双尾）	.249	.187	.684	.653	.226	.130	.230	.281	.032	< .001	.739	.479	< .001	.783	.082	.908		.038	.059	.784	.555	.427
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果糖质量/g	皮尔逊相关性	-.105	.010	-.269	.006	-.025	-.257	-.061	-.201	-.271	-.383*	-.090	-.192	-.274	.121	-.075	.035	-.401*	1	.678**	-.350	.085	.633**
	显著性（双尾）	.602	.960	.175	.975	.900	.195	.761	.314	.171	.048	.656	.337	.167	.549	.709	.861	.038		< .001	.074	.674	< .001
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
百粒质量/g	皮尔逊相关性	-.263	-.161	-.391*	-.136	-.390*	-.308	-.383*	-.293	-.330	-.366	-.071	-.346	-.259	-.036	.227	-.211	-.368	.678**	1	-.386*	-.101	.709**
	显著性（双尾）	.186	.422	.044	.499	.044	.118	.049	.138	.093	.061	.726	.077	.191	.857	.255	.292	.059	< .001		.047	.616	< .001
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果树比(%)	皮尔逊相关性	.496**	.279	.502**	.467*	.141	.235	.135	.633**	.474*	.264	.259	.220	-.212	.018	-.285	.061	-.055	-.350	-.386*	1	.128	-.247
	显著性（双尾）	.008	.159	.008	.010	.483	.237	.503	< .001	.013	.184	.191	.269	.288	.929	.150	.763	.784	.074	.047		.524	.215
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
出汁率(%)	皮尔逊相关性	.320	.038	.202	.465*	.084	.097	-.053	.045	.352	-.009	.568**	-.085	.211	.283	-.056	.112	.119	.085	-.101	.128	1	-.039
	显著性（双尾）	.103	.849	.311	.015	.676	.629	.793	.825	.072	.965	.002	.674	.291	.152	.780	.577	.555	.674	.616	.524		.848
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果皮质量（g）	皮尔逊相关性	-.055	.195	-.068	-.061	-.268	-.263	-.132	.118	-.124	-.322	-.016	-.392*	-.159	-.194	.276	-.168	-.159	.633**	.709**	-.247	-.039	1
	显著性（双尾）	.785	.329	.735	.763	.177	.185	.510	.558	.539	.101	.936	.043	.427	.358	.163	.402	.427	< .001	< .001	.215	.848	
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

*. 在 0.05 级别（双尾），相关性显著。

**. 在 0.01 级别（双尾），相关性显著。

图 1: 花色苷相关性数据

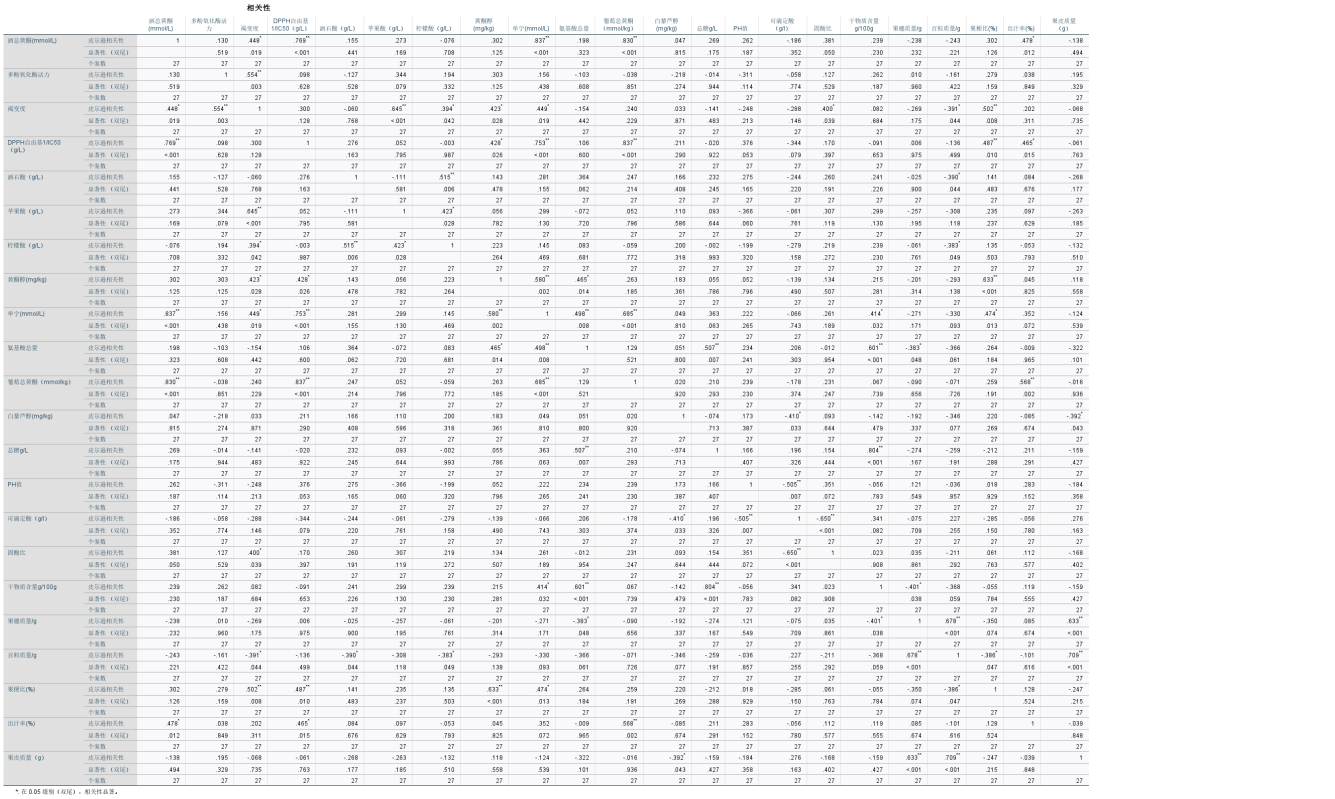
相关性

		DPPH半抑制体积（1/V50）1/V50(μL)	多酚氧化酶活力	褐变度	DPPH自由基1/IC50 (g/L)	酒石酸 (g/L)	苹果酸 (g/L)	柠檬酸 (g/L)	黄酮醇(mg/kg)	单宁(mmol/L)	氨基酸总量	聚酚总类酶(mmol/kg)	白藜芦醇(mg/kg)	总糖g/L	PH值	可滴定酸(g/l)	固酸比	干物质含量g/100g	果糖质量/g	百粒质量/g	果树比(%)	出汁率(%)	聚点质量(g)
DPPH半抑制体积（1/V50）1/V50(μL)	皮尔逊相关性	1	.092	.383*	.778**	.237	.246	.017	.422*	.915**	.400*	.814**	.073	.319	.208	-.061	.233	.329	-.201	-.229	.327	.417*	-.078
	显著性（双尾）		.685	.048	< .001	.234	.217	.931	.028	< .001	.038	< .001	.717	.105	.297	.763	.242	.094	.314	.251	.095	.031	.700
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
多酚氧化酶活力	皮尔逊相关性	.092	1	.554**	.098	-.127	.344	.194	.303	.156	-.103	-.038	-.218	-.014	-.311	-.058	.127	.262	.010	-.161	.279	.038	.195
	显著性（双尾）	.685		.003	.628	.528	.079	.332	.125	.438	.608	.851	.274	.944	.114	.774	.529	.187	.960	.422	.159	.849	.329
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
褐变度	皮尔逊相关性	.383*	.554**	1	.300	-.060	.645**	.394*	.423*	.449*	-.154	.240	.033	-.141	-.248	-.288	.400*	.082	-.269	-.391*	.502**	.202	-.068
	显著性（双尾）	.048	.003		.128	.768	< .001	.042	.028	.019	.442	.229	.871	.483	.213	.146	.039	.684	.175	.044	.008	.311	.735
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
DPPH自由基1/IC50（g/L）	皮尔逊相关性	.778**	.098	.300	1	.276	.052	-.003	.428*	.753**	.106	.837**	.211	-.020	.376	-.344	.170	-.091	.006	-.136	.467**	.465*	-.061
	显著性（双尾）	< .001	.628	.128		.163	.795	.987	.026	< .001	.600	< .001	.290	.922	.053	.079	.397	.653	.975	.499	.010	.015	.763
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
酒石酸 (g/L)	皮尔逊相关性	.237	-.127	-.060	.276	1	-.111	.515**	.143	.281	.364	.247	.166	.232	.275	-.244	.260	.241	-.025	-.390*	.141	.084	-.268
	显著性（双尾）	.234	.528	.768	.163		.581	.006	.478	.155	.062	.214	.408	.245	.165	.220	.191	.226	.900	.044	.483	.676	.177
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
苹果酸 (g/L)	皮尔逊相关性	.246	.344	.645**	.052	-.111	1	.423*	.056	.299	-.072	.052	.110	.093	-.366	-.061	.307	.299	-.257	-.308	.235	.097	-.263
	显著性（双尾）	.217	.079	< .001	.795	.581		.028	.782	.130	.720	.796	.586	.644	.060	.761	.119	.130	.195	.118	.237	.629	.185
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
柠檬酸 (g/L)	皮尔逊相关性	.017	.194	.394*	-.003	.515**	.423*	1	.223	.145	.083	-.059	.200	-.002	-.199	-.279	.219	.239	-.061	-.383*	.135	-.053	-.132
	显著性（双尾）	.931	.332	.042	.987	.006	.028		.264	.469	.681	.772	.318	.993	.320	.158	.272	.230	.761	.049	.503	.793	.510
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
黄酮醇(mg/kg)	皮尔逊相关性	.422*	.303	.423*	.428*	.143	.056	.223	1	.580**	.465*	.263	.183	.055	.052	-.139	.134	.215	-.201	-.293	.633**	.045	.118
	显著性（双尾）	.028	.125	.028	.026	.478	.782	.264		.002	.014	.185	.361	.786	.796	.490	.507	.281	.314	.138	< .001	.825	.558
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
单宁(mmol/L)	皮尔逊相关性	.915**	.156	.449*	.753**	.281	.299	.145	.580**	1	.498**	.685**	.049	.363	.222	-.066	.261	.414*	-.271	-.330	.474*	.352	-.124
	显著性（双尾）	< .001	.438	.019	< .001	.155	.130	.469	.002		.008	< .001	.810	.063	.265	.743	.189	.032	.171	.093	.013	.072	.539
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
氨基酸总量	皮尔逊相关性	.400*	-.103	-.154	.106	.364	-.072	.083	.465*	.498**	1	.129	.051	.507**	.234	.206	-.012	.601**	-.383*	-.366	.264	-.009	-.322
	显著性（双尾）	.038	.608	.442	.600	.062	.720	.681	.014	.008		.521	.800	.007	.241	.303	.954	< .001	.048	.061	.184	.965	.101
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
儿茶素总类酶（mmol/kg）	皮尔逊相关性	.814**	-.038	.240	.837**	.247	.052	-.059	.263	.685**	.129	1	.020	.210	.239	-.178	.231	.067	-.090	-.071	.259	.568**	-.016
	显著性（双尾）	< .001	.851	.229	< .001	.214	.796	.772	.185	< .001	.521		.920	.293	.230	.374	.247	.739	.656	.726	.191	.002	.936
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
白藜芦醇(mg/kg)	皮尔逊相关性	.073	-.218	.033	.211	.166	.110	.200	.183	.049	.051	.020	1	-.074	.173	-.410*	.093	-.142	-.192	-.346	.220	-.085	-.392*
	显著性（双尾）	.717	.274	.871	.290	.408	.586	.318	.361	.810	.800	.920		.713	.387	.033	.644	.479	.337	.077	.269	.674	.043
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
总糖g/L	皮尔逊相关性	.319	-.014	-.141	-.020	.232	.093	-.002	.055	.363	.507**	.210	-.074	1	.166	.196	.154	.804**	-.274	-.259	-.212	.211	-.159
	显著性（双尾）	.105	.944	.483	.922	.245	.644	.993	.786	.063	.007	.293	.713		.407	.326	.444	< .001	.167	.191	.288	.291	.427
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
PH值	皮尔逊相关性	.208	-.311	-.248	.376	.275	-.366	-.199	.052	.222	.234	.239	.173	.166	1	-.505**	.351	-.056	.121	-.036	.018	.283	-.184
	显著性（双尾）	.297	.114	.213	.053	.165	.060	.320	.796	.265	.241	.230	.387	.407		.007	.072	.783	.549	.857	.929	.152	.358
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
可滴定酸（g/l）	皮尔逊相关性	-.061	-.058	-.288	-.344	-.244	-.061	-.279	-.139	-.066	.206	-.178	-.410*	.196	-.505**	1	-.650**	.341	-.075	.227	-.285	-.056	.276
	显著性（双尾）	.763	.774	.146	.079	.220	.761	.158	.490	.743	.303	.374	.033	.326	.007		< .001	.082	.709	.255	.150	.780	.163
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
固酸比	皮尔逊相关性	.233	.127	.400*	.170	.260	.307	.219	.134	.261	-.012	.231	.093	.154	.351	-.650**	1	.023	.035	-.211	.061	.112	-.168
	显著性（双尾）	.242	.529	.039	.397	.191	.119	.272	.507	.189	.954	.247	.644	.444	.072	< .001		.908	.861	.292	.763	.577	.402
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
干物质含量g/100g	皮尔逊相关性	.329	.262	.082	-.091	.241	.299	.239	.215	.414*	.601**	.067	-.142	.804**	-.056	.341	.023	1	-.401*	-.368	-.055	.119	-.159
	显著性（双尾）	.094	.187	.684	.653	.226	.130	.230	.281	.032	< .001	.739	.479	< .001	.783	.082	.908		.038	.059	.784	.555	.427
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果糖质量/g	皮尔逊相关性	-.201	.010	-.269	.006	-.025	-.257	-.061	-.201	-.271	-.383*	-.090	-.192	-.274	.121	-.075	.035	-.401*	1	.678**	-.350	.085	.633**
	显著性（双尾）	.314	.960	.175	.975	.900	.195	.761	.314	.171	.048	.656	.337	.167	.549	.709	.861	.038		< .001	.074	.674	< .001
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
百粒质量/g	皮尔逊相关性	-.229	-.161	-.391*	-.136	-.390*	-.308	-.383*	-.293	-.330	-.366	-.071	-.346	-.259	-.036	.227	-.211	-.368	.678**	1	-.386*	-.101	.709**
	显著性（双尾）	.251	.422	.044	.499	.044	.118	.049	.138	.093	.061	.726	.077	.191	.857	.255	.292	.059	< .001		.047	.616	< .001
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果树比(%)	皮尔逊相关性	.327	.279	.502**	.467*	.141	.235	.135	.633**	.474*	.264	.259	.220	-.212	.018	-.285	.061	-.055	-.350	-.386*	1	.128	-.247
	显著性（双尾）	.095	.159	.008	.010	.483	.237	.503	< .001	.013	.184	.191	.269	.288	.929	.150	.763	.784	.074	.047		.524	.215
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
出汁率(%)	皮尔逊相关性	.417*	.038	.202	.465*	.084	.097	-.053	.045	.352	-.009	.568**	-.085	.211	.283	-.056	.112	.119	.085	-.101	.128	1	-.039
	显著性（双尾）	.031	.849	.311	.015	.676	.629	.793	.825	.072	.965	.002	.674	.291	.152	.780	.577	.555	.674	.616	.524		.848
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
果皮质量（g）	皮尔逊相关性	-.078	.195	-.068	-.061	-.268	-.263	-.132	.118	-.124	-.322	-.016	-.392*	-.159	-.194	.276	-.168	-.159	.633**	.709**	-.247	-.039	1
	显著性（双尾）	.700	.329	.735	.763	.177	.185	.510	.558	.539	.101	.936	.043	.427	.358	.163	.402	.427	< .001	< .001	.215	.848	
	个案数	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

*. 在 0.05 级别（双尾），相关性显著。

**. 在 0.01 级别（双尾），相关性显著。

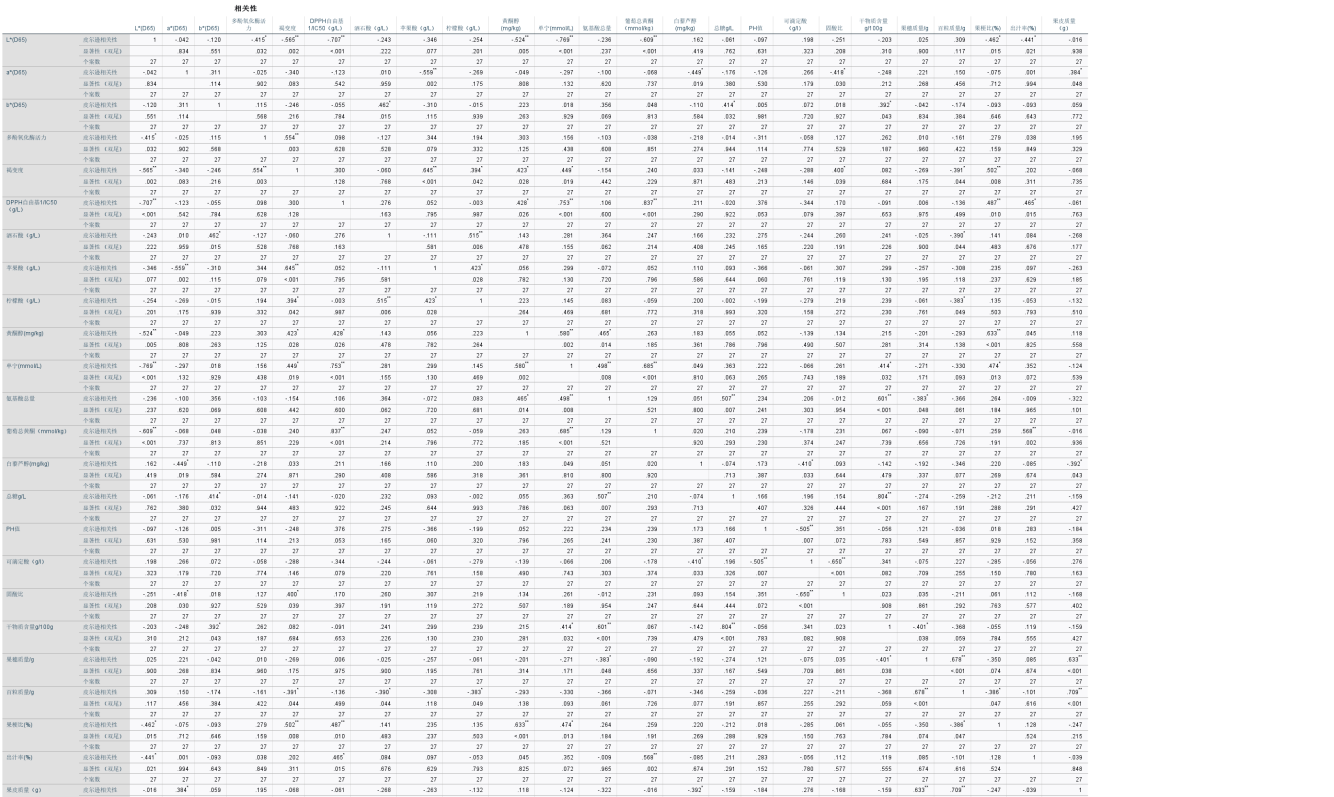
图 2: DPPH 半抑制体积相关性数据



* 在 0.05 级别（双尾），相关性显著。

** 在 0.01 级别（双尾），相关性显著。

图 3: 酒总黄酮相关性数据



* 在 0.05 级别（双尾），相关性显著。

** 在 0.01 级别（双尾），相关性显著。

图 4: 色泽相关性数据

5.4.2 结果分析

对相关性分析的结果进行分析，其重点就是观察其 pearson 和双侧显著性的值，pearson 相关系数的值大，则相关性高，同理观察算观测双侧显著性则是判断其值的范围，*p<0.05 或者 **p<0.01 都证明其有较好的相关显著性。通过对结果进行分析，得出如下的一些较为明显的相关性结论:

1) 花色苷的指标与褐变度、苹果酸、单宁、果梗比等数据的相关性较为显著

2) DPPH 半抑制体积与多酚氧化酶活力、DPPH 自由基、单宁、葡萄总黄酮、白藜芦醇相关性高，与果皮质量和果穗质量等指标都成负相关

3) 酒总黄酮则跟 DPPH 自由基、单宁、葡萄总黄酮，白藜芦醇等指标相关性高

4) 色泽跟可滴定酸、果穗质量的略微相关，总体数据的相关性不大

按照分析标准来分析，无论是相关性系数和显著程度都选取较为明显的那组数据来进行后面的关系分析。本小问中，在花色苷则这项数据所给出的相关性指标的相关性都较高，在多元回归中则考虑采取这项数据来进行拟合。

5.5 多元线性回归模型的求解

在建立模型则需要对模型进行拟合度检验，多元回归方程的显著性检验就是检验样本回归方程的变量的线性关系是否显著，需要根据样本来判断方程中的多个回归系数中至少有一个不等于 0，主要是说明样本回归方程的显著性。检验的方法用方差分析，这时因变量的总体为回归平方和与误差平方和，即表示为：

$$L_x x = Q + U \tag{5}$$

其中该公式又可以表示为：

$$L_x x = \sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{6}$$

$$Q = \sum_{i=1}^{N} (y_i - \hat{y})^2 \tag{7}$$

$$U = \sum_{i=1}^{N} (\hat{y_i} - \bar{y})^2 \tag{8}$$

对花色苷和其相关性较高的指标进行拟合，依据德宾沃森残差，离群值为 3 标准差进行模型拟合，来根据其 R 方及其德宾沃森残差来观察其拟合的契合度。另外进行单因素方差分析 (ANOVA) 来观测 F 检验对整个回归进行显著性检验，考虑的 k 个变量自变量是否有显著性线性关系 F 检测通过与 F 边界值来进行比对判断其水平显著性

$$\begin{cases} F_0.05(k, n-k-1) \le F \le F_0.01(k, n-k-1) & 0.05 \\ F_0.1(k, n-k-1) \le F \le F_0.05(k, n-k-1) & 0.01 \\ F < F_0.1(k, n-k-1) \end{cases}$$

5.5.1 单因素方差分析

对花色苷等多项相关性数据进行 ANOVA 分析，观察 F 检测值和显著性，通过三组不同数据的模型，对数据进行共线性诊断，依据 VIF 值来确定较为合理的自变量，来进行试验观察。得出如下的结果：

ANOVA[a]

模型		平方和	自由度	均方	F	显著性
1	回归	1220392.299	9	135599.144	14.906	<.001[b]
	残差	154643.029	17	9096.649		
	总计	1375035.327	26			
2	回归	1332224.511	15	88814.967	22.821	<.001[c]
	残差	42810.817	11	3891.892		
	总计	1375035.327	26			
3	回归	1333687.803	16	83355.488	20.160	<.001[d]
	残差	41347.524	10	4134.752		
	总计	1375035.327	26			

a. 因变量：花色苷(mg/L)

b. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）

c. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）, 白藜芦醇(mg/kg), PH值, 多酚氧化酶活力, 酒石酸（g/L）, 总糖g/L, 氨基酸总量

d. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）, 白藜芦醇(mg/kg), PH值, 多酚氧化酶活力, 酒石酸（g/L）, 总糖g/L, 氨基酸总量, 可滴定酸（g/l）

图 5: ANOVA

分析观察三组回归的数据，发现三个模型的显著性都是 <0.01，其 F 检测值分别为：14.906、22.821、20.160。根据该结果和显著性 p 值可以拒绝原假设，认为被解释变量个解释变量间存在显著的线性关系，可建立线性回归模型。

如下为残差统计图：

残差统计[a]

	最小值	最大值	平均值	标准偏差	个案数
预测值	41.21715546	1000.019653	263.3166736	226.4855067	27
标准预测值	-.981	3.253	.000	1.000	27
预测值的标准误差	35.139	61.535	50.582	6.824	27
调整后预测值	62.05002975	1292.442383	274.1753193	256.8458629	27
残差	-74.4163361	109.3441620	.0000000000	39.87843273	27
标准残差	-1.157	1.700	.000	.620	27
学生化残差	-1.919	2.182	-.038	1.048	27
剔除残差	-319.314911	233.3688507	-10.8586457	135.3274944	27
学生化剔除残差	-2.290	2.861	-.036	1.177	27
马氏距离	6.801	22.847	15.407	4.248	27
库克距离	.002	1.328	.190	.315	27
居中杠杆值	.262	.879	.593	.163	27

a. 因变量：花色苷(mg/L)

图 6: 残差统计

在经过处理后的数据是比较合理的。

5.5.2 多元回归拟合

花色苷作为因变量，褐变度、苹果酸、单宁、果梗比、DPPH 自由基、葡萄总黄酮、多酚氧化酶活力、总酚等相关性较为显著等数据作为自变量来进行多元回归拟合。在进行比对后发现并无需要严格剔除的数据，则进行多元线性回归变量筛选结果及系数的拟合求解。观察系数表和其显著性指标，通过 R^2 来判断其回归拟合的契合度，往往 R^2 越贴近于 1，契合度越高。确定酿酒葡萄与葡萄酒理化指标的联系则将系数组合成回归方程即可。

系数[a]

模型		未标准化系数 B	未标准化系数 标准错误	标准化系数 Beta	t	显著性
1	(常量)	-420.632	212.575		-1.979	.064
	苹果酸（g/L）	19.012	7.680	.315	2.475	.024
	果梗比(%)	-2.420	25.326	-.012	-.096	.925
	柠檬酸（g/L）	20.007	29.573	.064	.677	.508
	DPPH自由基1/IC50（g/L）	701.641	385.535	.341	1.820	.086
	黄酮醇(mg/kg)	-.569	.756	-.100	-.752	.462
	出汁率(%)	1.669	3.176	.053	.525	.606
	葡萄总黄酮（mmol/kg）	-11.366	8.164	-.241	-1.392	.182
	褐变度	.229	.089	.334	2.564	.020
	单宁(mmol/L)	35.187	12.618	.444	2.789	.013
2	(常量)	-452.439	264.020		-1.714	.115
	苹果酸（g/L）	21.635	6.040	.358	3.582	.004
	果梗比(%)	7.296	19.974	.036	.365	.722
	柠檬酸（g/L）	79.312	30.080	.254	2.637	.023
	DPPH自由基1/IC50（g/L）	-133.202	598.655	-.065	-.223	.828
	黄酮醇(mg/kg)	-.740	.578	-.130	-1.281	.227
	出汁率(%)	1.381	2.284	.044	.604	.558
	葡萄总黄酮（mmol/kg）	3.247	7.805	.069	.416	.685
	褐变度	.002	.105	.003	.018	.986
	单宁(mmol/L)	57.834	18.069	.730	3.201	.008
	多酚氧化酶活力	5.585	2.247	.240	2.485	.030
	酒石酸（g/L）	-13.320	6.718	-.186	-1.983	.073
	氨基酸总量	.000	.019	.003	.021	.983
	白藜芦醇(mg/kg)	-4.099	3.303	-.098	-1.241	.240
	总糖g/L	-2.760	1.061	-.294	-2.601	.025
	PH值	161.159	77.717	.175	2.074	.062
3	(常量)	-180.363	532.189		-.339	.742
	苹果酸（g/L）	21.524	6.228	.356	3.456	.006
	果梗比(%)	2.492	22.116	.012	.113	.913
	柠檬酸（g/L）	70.579	34.305	.226	2.057	.067
	DPPH自由基1/IC50（g/L）	-123.926	617.248	-.060	-.201	.845
	黄酮醇(mg/kg)	-.727	.596	-.128	-1.220	.250
	出汁率(%)	2.169	2.702	.069	.803	.441
	葡萄总黄酮（mmol/kg）	1.564	8.528	.033	.183	.858
	褐变度	-.013	.111	-.018	-.113	.912
	单宁(mmol/L)	60.830	19.293	.768	3.153	.010
	多酚氧化酶活力	5.367	2.345	.231	2.288	.045
	酒石酸（g/L）	-12.987	6.948	-.182	-1.869	.091
	氨基酸总量	.004	.020	.024	.175	.865
	白藜芦醇(mg/kg)	-4.647	3.527	-.111	-1.318	.217
	总糖g/L	-2.826	1.099	-.301	-2.570	.028
	PH值	102.473	127.078	.111	.806	.439
	可滴定酸（g/l）	-12.635	21.239	-.075	-.595	.565

a. 因变量：花色苷(mg/L)

图 7: 系数

通过如下图表判断 R^2

模型摘要[d]

模型	R	R 方	调整后 R 方	标准估算的错误	德宾-沃森
1	.942[a]	.888	.828	95.37635321	
2	.984[b]	.969	.926	62.38503369	
3	.985[c]	.970	.922	64.30204065	1.568

a. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）

b. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）, 白藜芦醇(mg/kg), PH值, 多酚氧化酶活力, 酒石酸（g/L）, 总糖g/L, 氨基酸总量

c. 预测变量：(常量), 单宁(mmol/L), 柠檬酸（g/L）, 出汁率(%), 果梗比(%), 苹果酸（g/L）, 褐变度, 黄酮醇(mg/kg), 葡萄总黄酮（mmol/kg）, DPPH自由基1/IC50（g/L）, 白藜芦醇(mg/kg), PH值, 多酚氧化酶活力, 酒石酸（g/L）, 总糖g/L, 氨基酸总量, 可滴定酸（g/l）

d. 因变量：花色苷(mg/L)

图 8: 摘要

通过上述分析得出，三组数据的 R^2 都相对趋近于 1，具有较高的契合度。比对之后选择模型 3 的数据来进行方程的建立。由如下的系数表来得到系数关系。选取苹果酸、果梗比、柠檬酸、DPPH 自由基、黄酮醇、出汁率..... 等指标作为自变量 $x_1, x_2, , , , , x_n$ 构建如下方程：

$$y=21.524x_1+2.49x_2+70.579x_3-123.926x_4-7.27x_5+2.169x_6+1.564x_7-0.013x_8+60.83x_9$$
$$+5.367x_{10}-12.987x_{11}+0.004x_{12}-4.647x_{13}-2.826x_{14}+102.473x_{15}-12.635x_{16}-420.632$$

其中 y 为因变量花色苷，$x_1, x_2, , , , , x_n$ 则为自变量苹果酸、果梗比、柠檬酸、DPPH 自由基、黄酮醇、出汁率..... 通过该方程可以反映出葡萄酒的某些指标与酿酒葡萄理化指标之间的关系。