

2012年全国大学生数学建模竞赛A题：葡萄酒的评价

葡萄酒和酿酒葡萄质量的一些分析方法
—— 2012年A题解答综述

周义仓

西安交通大学数学与统计学院

zhouyc@mail.xjtu.edu.cn

029-82668741

主要内容

问题与背景

赛题解答举例

一些问题和建议

说明：根据命题人给出的求解思路和阅卷中看到的学生解答情况整理和述评，也包括自己的一些理解。所讲述的内容不是标准答案，也不是最好的解法，仅供大家参考。不妥之处请大家指正。

问题与背景：2012A题 葡萄酒的评价

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？

附件1 葡萄酒品尝得分表

总分: 100		干白品种	品酒员1	品酒员2	品酒员3	品酒员4	品酒员5	品酒员6	品酒员7	品酒员8	品酒员9	品酒员10
26项目满分		酒样品26										
外观分析	5	澄清度	4	3	4	4	4	5	4	4	5	4
15	10	色调	6	8	8	8	8	8	8	10	10	8
香气分析	6	纯正度	5	5	5	5	5	6	5	5	5	5
30	8	浓度	7	6	8	8	8	8	7	6	8	7
	16	质量	14	10	14	12	16	14	14	12	14	14
口感分析	6	纯正度	4	4	5	4	6	5	4	3	5	4
44	8	浓度	7	6	6	6	8	8	7	6	8	6
	8	持久性	7	6	6	6	8	8	7	5	6	7
	22	质量	13	10	16	13	19	19	16	16	19	19
平衡/整体评价	11		8	8	10	9	11	10	9	9	10	10
5		酒样品5										
外观分析	5	澄清度	4	1	2	3	2	1	2	2	1	2
	10	色调	6	4	6	6	6	4	6	6	6	4
香气分析	6	纯正度	6	3	5	4	5	3	6	5	5	5
	8	浓度	7	4	7	6	7	6	7	7	6	7
	16	质量	14	10	14	12	14	12	12	14	12	14
口感分析	6	纯正度	4	2	4	2	5	3	4	4	6	4
	8	浓度	7	2	7	4	6	6	7	6	8	6
	8	持久性	7	4	6	5	6	6	6	5	6	6
	22	质量	19	10	16	10	19	13	16	16	19	16
平衡/整体评价	11		10	7	10	8	9	8	8	9	10	10

附件2 葡萄酒和酿酒葡萄的理化指标

葡萄酒：花色苷(mg/L) ,单宁(mmol/L),总酚(mmol/L),酒总黄酮(mmol/L),白藜芦醇(mg/L), DPPH半抑制体积,色泽;

酿酒葡萄：氨基酸总量，蛋白质， VC含量，花色苷mg/100g鲜重,酒石酸，苹果酸，柠檬酸，多酚氧化酶活力，褐变度， DPPH自由基1/IC50，总酚，单宁，葡萄总黄酮，黄酮醇，白藜芦醇,总糖，还原糖，可溶性固形物， PH值，可滴定酸，固酸比，干物质含量，果穗质量，百粒质量，果梗比，出汁率，果皮质量，果皮颜色

品种编号	花色苷(mg/L)			单宁 (mmol/L)		
	1	2	3	1	2	3
红葡萄酒						
酒样品1	973.128	974.380	974.128	11.049	11.030	11.010
酒样品2	516.830	518.083	517.830	11.030	11.146	11.059
酒样品3	398.686	399.938	397.686	13.289	13.250	13.240
酒样品4	182.435	183.688	184.435	6.463	6.511	6.458
酒样品5	279.706	280.959	279.906	5.869	5.849	5.830
酒样品6	116.475	117.727	116.875	7.349	7.378	7.334
酒样品7	90.174	91.426	90.874	4.028	4.019	3.994
酒样品8	917.604	918.856	919.604	12.023	12.052	12.008
酒样品9	387.414	388.667	387.214	12.928	12.957	12.914
酒样品10	138.601	139.853	137.687	5.567	5.586	5.547
酒样品11	11.272	12.524	11.717	4.574	4.622	4.569
酒样品12	83.494	84.747	83.994	6.473	6.463	6.438
酒样品13	199.552	200.804	199.885	6.414	6.375	6.365
酒样品14	251.318	252.571	250.822	6.083	6.083	6.054
酒样品15	121.902	123.154	122.719	3.951	4.038	3.965
酒样品16	170.746	171.999	171.761	4.846	4.837	4.812
酒样品17	233.784	235.037	234.439	9.160	9.199	9.150
酒样品18	71.388	72.640	71.677	4.457	4.457	4.428

附件3-芳香物质

葡萄酒：70多种，葡萄：50多种

中文名称	分子量	化学式	葡萄样品 11	葡萄样品 14	葡萄样品 1	葡萄样品 25	葡萄样品 20	葡萄样品 18
乙醛	44	C2H4O	1.155	0.344	1.836	1.087	1.403	4.165
乙酸乙酯	88	C4H8O2	5.026	4.135	12.113	7.683	15.955	12.490
乙醇	46	C2H6O	53.304	41.144	89.657	47.006	68.427	77.416
丙酸乙酯	102	C5H10O2		0.235		0.118	0.599	0.825
乙酸正丙酯	102	C5H10O2	0.096	0.172			0.701	
乙酸-2-甲基丙基酯	116	C6H12O2	0.279	2.786	0.498	0.276	0.524	0.275
丁酸乙酯	116	C6H12O2	1.148	0.674	2.655	1.340	2.360	1.178
1-丙醇	60	C3H8O		0.258				
2,3-乙酰基丙酮	100	C5H8O2	0.193	0.086	3.059			
正十一烷	156	C11H24		0.344	3.059			
2-甲基-1-丙醇	74	C4H10O	0.776	0.381	2.046	0.670	1.049	1.964
3-甲基-1-丁醇-乙 酸酯	130	C7H14O2	9.184	8.182	39.104	18.125	43.558	19.992
乙酸戊酯	130	C7H14O2	0.096	0.086	0.306		0.701	0.771
柠檬烯	136	C10H16	0.096				0.421	0.154
3-甲基-1-丁醇	88	C5H12O	13.993			21.119	37.341	71.367
己酸乙酯	144	C8H16O2	17.499	32.933	73.285	23.719	12.846	10.526

对题目的反映和几个数据错误

- 今年的A题来源于实际，保持原貌；
 - 容易理解，学生选择多；
 - 风格不同，许多统计方法的综合使用；
 - 有许多资料和研究结果，但无与竞赛题完全类似的论文；
 - 个别叙述不是很明确，个别数据不齐。
-
- 第一组红葡萄酒样20，评酒员4号中 缺色调数据， 6；
 - 第一组白葡萄酒样3, 7号评酒员的持久性超上限， 77--7；
 - 第一组白葡萄酒样8, 9号评酒员的持久性超上限， 16--6；
 - 酿酒白葡萄的百粒质量第三组数据太大， 2226.1--226.1。

葡萄酒的质量

- “七分原料、三分酿造”；
- 优质葡萄酒是酸度、香气、风味等的平衡，葡萄果实中糖、酸决定葡萄酒的酒度、味感，丹宁、总酚、芳香物质、花色素苷等决定葡萄酒的结构、香气、外观等质量特征；
- 单宁和色素对红葡萄酒的特色和风味作用也是显著的。单宁是很好的抗氧化物质，它的涩味和收敛感又造就了葡萄酒丰富的厚重品质。葡萄酒的颜色来源于葡萄中的色素。葡萄的色素则决定着红葡萄酒的颜色气质。
- 芳香物质是造就葡萄酒风味的物质之一，芳香物质越多，葡萄酒的风味就浓厚。

GB15037-2006 《葡萄酒》

-----葡萄酒感观分类评述描述

- 优级品：90分以上 具有该产品应有的色泽，自然、悦目、澄清（透明）、有光泽；具有纯正、浓郁、优雅和谐的果香（酒香），诸香协调，口感细腻、舒顺、酒体丰满、完整、回味绵长、具该产品应有的怡人的风格。
- 优良品：80～89分 具有该产品的色泽；澄清透明，无明显悬浮物，具有纯正和谐的果香（酒香），无异香，浓郁度稍差，口感纯正，较舒顺，完整，欠优雅，回味较长，具良好的风格。
- 合格品：70～79分 与该产品应有的色泽略有不同，缺少自然感，允许有少量沉淀，具有该产品应有的气味，无异味，口感尚平衡，欠协调、完整，无明显缺陷。
- 不合格品：65～69分 与该产品应有的色泽明显不符，严重失光或浑浊，有明显异香、异味，酒体寡淡、不协调，或有其他明显的缺陷。（除色泽外，只要有其中一条，则判为不合格品。）
- 劣质品：55～64分 不具备应有的特征。

问题1

- 问题：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？
- 评阅要点：附件1中给出的是评酒员对27种红葡萄酒和28种白葡萄酒的两组品评结果。这两组评酒员各不相同，两组中的每个酒样都取自相同葡萄酒厂家的同一批次的产品。要求学生给出判断这两组评价结果好坏的原理、模型和方法，给出具体的结果，并对结果进行说明。好的品评结果应该是对同一酒样评价时这些评酒员之间的差距小、且这些酒样之间的区分度明确（注：一些学生的模型和方法仅考虑评酒员的打分差距）。参考：红酒中样品23是好酒，样品12是较差的酒。
- 参考结论：两组评价结果有显著性差异；红酒第一组评价结果好，白酒第二组评价结果好。

问题1 解答综述

- 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？
- 第一问：方差分析，**t**检验，**F**检验，秩和检验，**Wilcoxon** 符号秩检验，构造一些统计量，计算它们的值，是否通过检验，给出结果。
- 第二问：比较方差大小，**Cronbach**可信度系数，比较酒样**F**值和评酒员**F**值的相对大小，**Spearman**秩相关系数，肯德尔和谐系数法。
- 应该注意所用统计方法的条件是否满足、了解其适用范围，对所得到的结果应该给出一些分析和说明。
- 应用统计方法时应该说明检验什么。
- 利用方差分析时的数据是否为正态，检验均值时是否方差是否相等
- **A441**方差分析比较明确；

是否有显著性差异的判断

双因素方差分析

因子 A 共有 r 个水平，记作 A_1, A_2, \dots, A_r ，每个水平下，进行 t 次试验，因子 B 共有 k 个水平。一个典型的双因子方差分析的数据结构如下表所示。

因子 A	因子 B			
	B_1	B_2	\dots	B_k
A_1	x_{11}	x_{12}	\dots	x_{1k}
	\vdots	\vdots	\vdots	\vdots
	x_{t1}	x_{t2}		x_{tk}
\vdots	\dots	\dots		\dots
A_r	x_{r1}	x_{r2}	\dots	x_{rk}
	\vdots	\vdots	\vdots	\vdots
	x_{t1}	x_{t2}	\dots	x_{tk}

在本题中， $r=2,k=10,t=10$ 。下表给出因子 B 所对应的各个指标：

B	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
指标	外观澄 清度	外观色 调	香气纯 正度	香气浓 度	香气质 量	口感纯 正度	口感浓 度	口感持 久性	口感质 量	整体得 分

给出双因子可重复方差分析的原假设和备择假设：

H_{01} : 两组评酒员的评价结果不存在差异. $\Leftrightarrow H_{02}$: 两组评酒员的评价结果存在着差异.

逐个进行方差分析后得到的结果如下

红葡萄酒 p_i 值以及 Y_i 值，得到 $\beta = 0.703$														
p_i	0.18971	0.00001	0.00040	0.00212	0.16314	0.00138	0.00486	0.00334	0.02476	0.00000	0.00002	0.00011	0.36479	0.21870
Y_i	0	1	1	1	0	1	1	1	1	1	1	1	0	0
p_i	0.00046	0.80100	0.00021	0.56414	0.17544	1.00000	0.00002	0.04686	0.01131	0.00017	0.00086	0.00112	0.00045	
Y_i	1	0	1	0	0	0	1	1	1	1	1	1	1	
白葡萄酒 p_i 值以及 Y_i 值，得到 $\beta = 0.535$														
p_i	0.00103	0.00001	0.10777	0.31115	0.50613	0.01060	0.34940	0.67936	0.00329	0.00460	0.00008	0.08585	0.00011	0.20310
Y_i	1	1	0	0	0	1	0	0	1	1	1	0	1	0
p_i	0.01714	0.03333	0.01381	0.19476	0.00339	0.44078	0.00034	0.00005	0.68334	0.46710	0.00031	0.16632	0.13648	0.00001
Y_i	1	1	1	0	1	0	1	1	0	0	1	0	0	1

检验结果

- 对于红葡萄酒来说，对27个葡萄酒样品评分检验中，有70.3%的评价结果中，两组评酒员的评价结果存在着显著性差异（置信水平为95%）。
- 对于白葡萄酒的28个葡萄样品评分的检验，有53%的评价结果中，两组评酒员的评价结果存在显著性检验（置信水平为95%）。
- 两组评酒员对红葡萄的评分结果更具有显著性差异
- 相比较而言，红葡萄酒两组评酒员的差异更显著一些，白葡萄酒两组评酒员的评价差异性较不明显

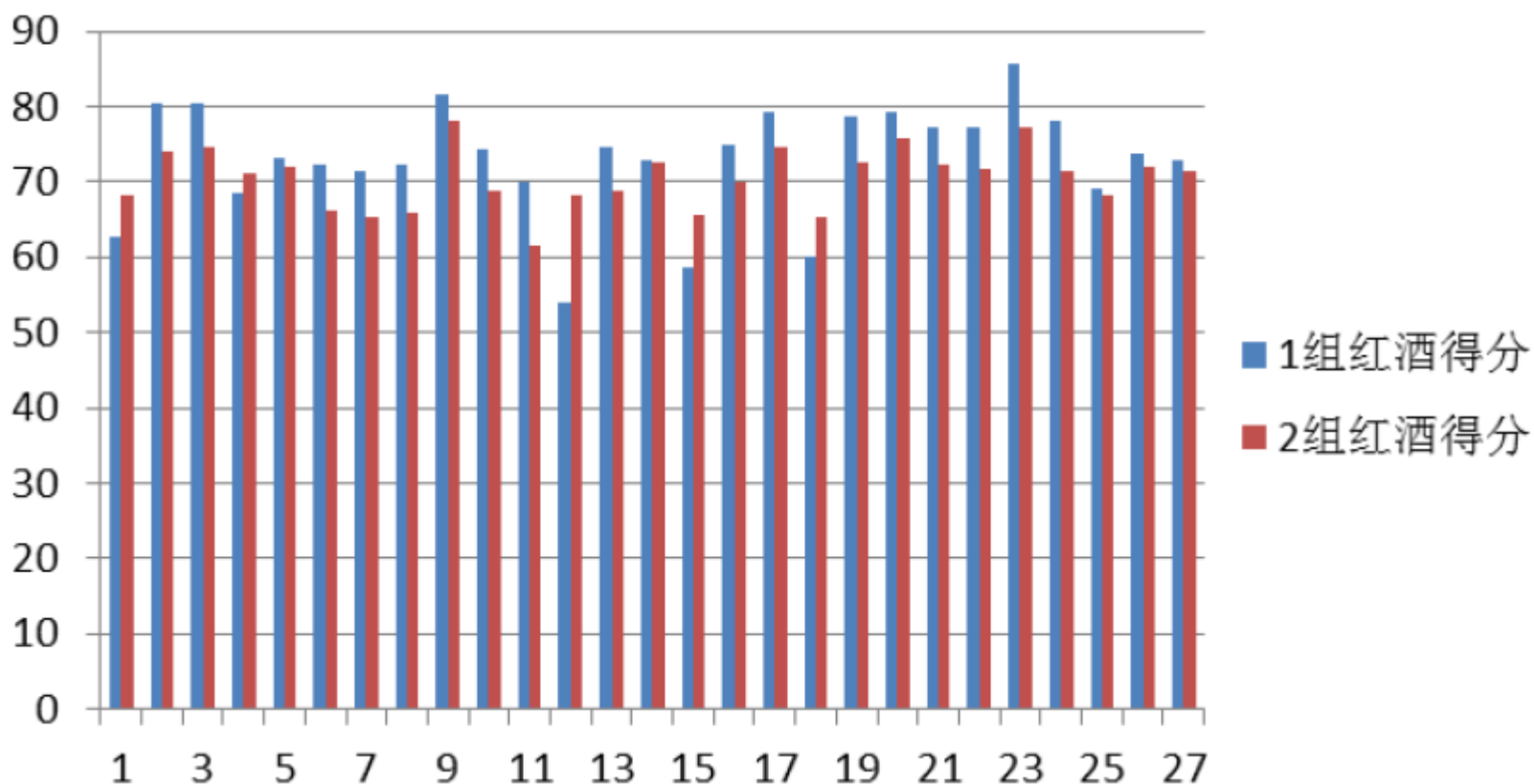
评价：逐个比较两组评酒员对每一个酒样的评分是否存在明显差异，检验的目的明确，而且考虑到了每一个酒样的10个小指标，最后计算比例，实际意义明确。

两组红葡萄酒的统计学特征

1 组红葡萄酒		2 组红葡萄酒	
平均	73.08519	平均	70.51481
中位数	73.8	中位数	71.5
标准差	7.34725	标准差	3.977988
方差	53.98208	方差	15.82439
区域	31.7	区域	16.6
最小值	53.9	最小值	61.6
最大值	85.6	最大值	78.2

先进行一些数据的统计分析

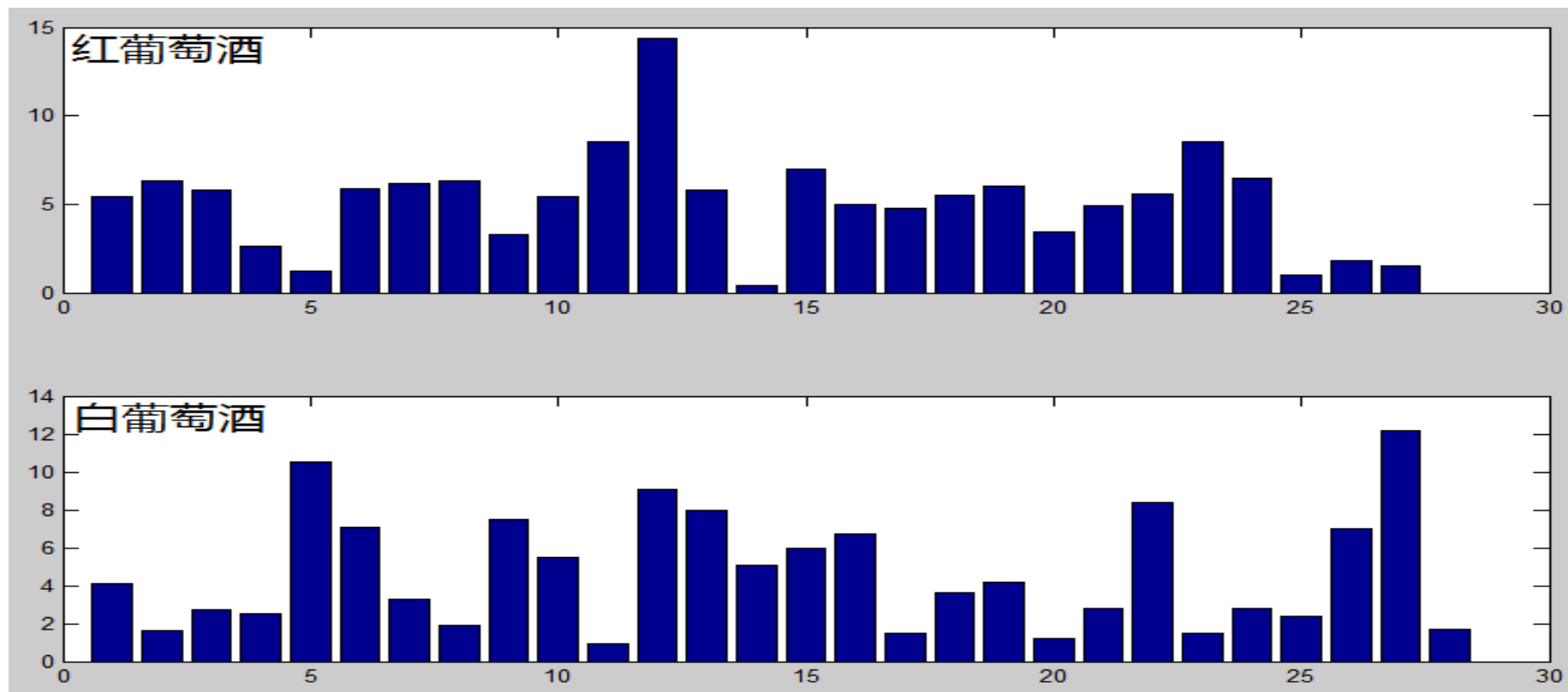
两组评酒员对**27**份红葡萄酒样品的评分结果



第一组评分较高，高低差距大；

第二组评酒员给**27**种酒样的得分差距相对较少

红白葡萄酒平均分之差的绝对值



- 27种红葡萄酒评分中，14、25、5号酒样的评分差距较小，12、11、23号酒样的评分差距很大；
- 28种白葡萄酒评分中，11、20、23号酒样的评分差距较小，27、5、12号酒样的评分差距很大。

差异显著性和评分可信度的直观判别

- 两组的平均分：

第一组红葡萄酒 73.1 第二组红葡萄酒 70.5

第一组白葡萄酒 74.0 第二组白葡萄酒 76.5

平均分没有显著性差异（差别小于5%）

- 两组对各个酒样的评分

两组红葡萄酒平均分差别大于5的数量 16

两组白葡萄酒平均分差别大于5的数量 12

两组评酒员对这些酒样的平均分有显著性差异

- 两组的排名差异

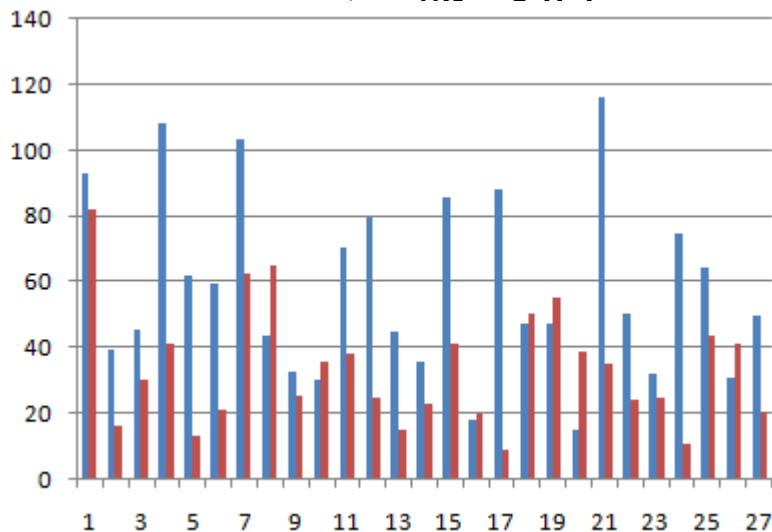
两组红酒评分排名与平均分的排名方差 4.6, 13.0

两组白酒评分排名与平均分的排名方差 19.5, 47.7

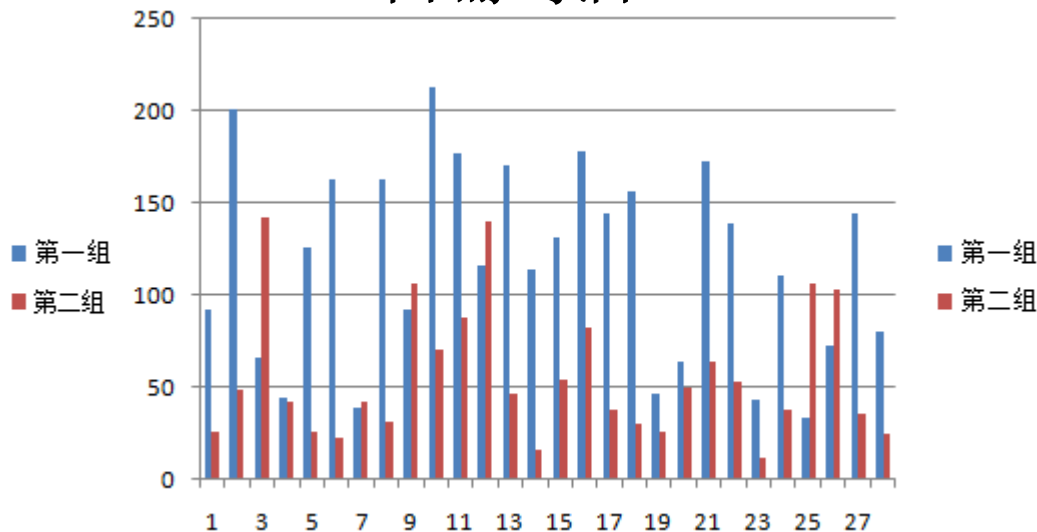
两组排名有显著性差异，红、白酒葡萄酒的排名第一组好

两组各10个评酒员对每个酒样评分的方差

红葡萄酒



白葡萄酒



红葡萄酒和白葡萄酒都是第一组的方差大；

红葡萄酒第一组评酒员对**20**、**16**号酒样评价评价相对较一致，对**4**、**7**、**21**号酒样的评价分歧较大；第二组评酒员对**17**、**24**、**5**号酒样评价较一致，对**1**、**7**、**8**号酒样的评价分歧较大；

白葡萄酒第一组评酒员对**25**、**7**号酒样评价较一致，对**2**、**10**号酒样的评价分歧较大；第二组评酒员对**23**、**14**号酒样评价较一致，对**3**、**12**号酒样的评价分歧较大；

通过排序来判断

- A031等参赛队利用排序的思想来分析和判断，一种简化的处理如下
- 分别按照各组的平均分排名；
- 比较两者的差异，看看是否超过某一个临界水平；
- 按照20人的平均分排名；
- 比较哪一组更接近20人的排名

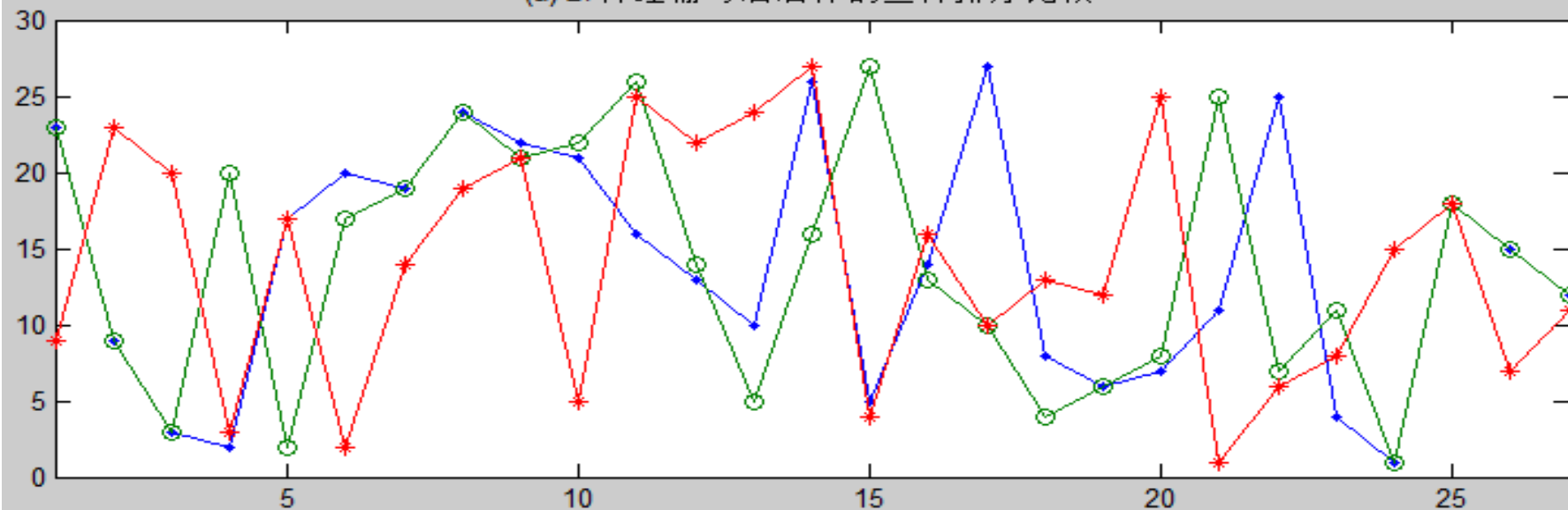
红葡萄酒排名

第一组得分		第二组得分		两组平均得分		第一组排名		第二组排名		两组平均排名	
62.7	A1	68.1	B1	65.4	C1	85.6	A23	78.2	B9	81.35	C23
80.3	A2	74	B2	77.15	C2	81.5	A9	77.1	B23	79.85	C9
80.4	A3	74.6	B3	77.5	C3	80.4	A3	75.8	B20	77.5	C3
68.6	A4	71.2	B4	69.9	C4	80.3	A2	74.6	B3	77.5	C20
73.3	A5	72.1	B5	72.7	C5	79.3	A17	74.5	B17	77.15	C2
72.2	A6	66.3	B6	69.25	C6	79.2	A20	74	B2	76.9	C17
71.5	A7	65.3	B7	68.4	C7	78.6	A19	72.6	B14	75.6	C19
72.3	A8	66	B8	69.15	C8	78	A24	72.6	B19	74.75	C24
81.5	A9	78.2	B9	79.85	C9	77.2	A22	72.2	B21	74.65	C21
74.2	A10	68.8	B10	71.5	C10	77.1	A21	72.1	B5	74.4	C22
70.1	A11	61.6	B11	65.85	C11	74.9	A16	72	B25	72.9	C26
53.9	A12	68.3	B12	61.1	C12	74.6	A13	71.6	B22	72.8	C14
74.6	A13	68.8	B13	71.7	C13	74.2	A10	71.5	B24	72.7	C5
73	A14	72.6	B14	72.8	C14	73.8	A26	71.5	B27	72.4	C16
58.7	A15	65.7	B15	62.2	C15	73.3	A5	71.2	B4	72.25	C27
74.9	A16	69.9	B16	72.4	C16	73	A14	69.9	B16	71.7	C13
79.3	A17	74.5	B17	76.9	C17	73	A27	68.8	B10	71.5	C10
59.9	A18	65.4	B18	62.65	C18	72.3	A8	68.8	B13	69.9	C4
78.6	A19	72.6	B19	75.6	C19	72.2	A6	68.3	B12	69.25	C6
79.2	A20	75.8	B20	77.5	C20	71.5	A7	68.2	B25	69.15	C8
77.1	A21	72.2	B21	74.65	C21	70.1	A11	68.1	B1	68.7	C25
77.2	A22	71.6	B22	74.4	C22	69.2	A25	66.3	B6	68.4	C7
85.6	A23	77.1	B23	81.35	C23	68.6	A4	66	B8	65.85	C11
78	A24	71.5	B24	74.75	C24	62.7	A1	65.7	B15	65.4	C1
69.2	A25	68.2	B25	68.7	C25	59.9	A18	65.4	B18	62.65	C18
73.8	A26	72	B25	72.9	C26	58.7	A15	65.3	B7	62.2	C15
73	A27	71.5	B27	72.25	C27	53.9	A12	61.6	B11	61.1	C12

白葡萄酒排名

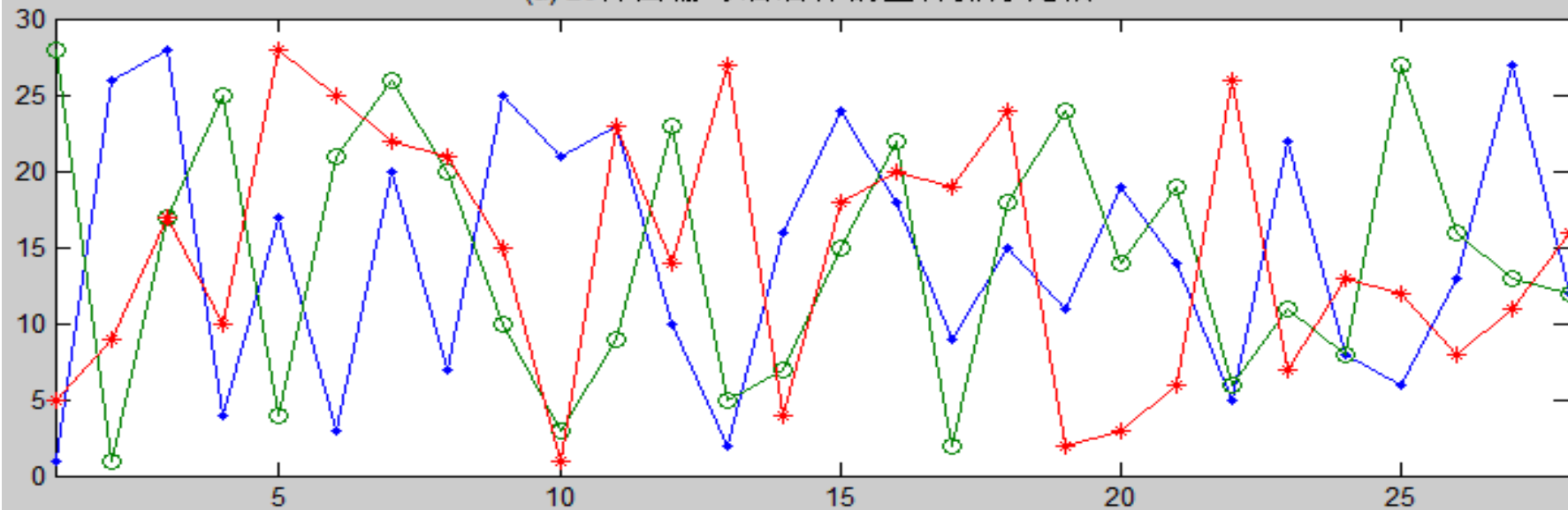
第一组得分		第二组得分		两组平均得分		第一组排名		第二组排名		两组平均排名	
82	A1	77.9	B1	79.95	C1	82	A1	81.5	B5	80.45	C28
74.2	A2	75.8	B2	75	C2	81.3	A26	80.4	B9	79.95	C1
78.3	A3	75.6	B3	76.95	C3	81.3	A28	80.3	B17	79.55	C17
79.4	A4	76.9	B4	78.15	C4	79.4	A4	79.8	B10	78.3	C25
71	A5	81.5	B5	76.25	C5	78.8	A17	79.6	B28	78.15	C4
68.4	A6	75.5	B6	71.95	C6	78.3	A3	79.5	B25	77.8	C21
77.5	A7	74.2	B7	75.85	C7	77.8	A20	79.4	B22	77.8	C26
70.4	A8	72.3	B8	71.35	C8	77.5	A7	79.2	B21	77.2	C20
72.9	A9	80.4	B9	76.65	C9	77.1	A25	78.4	B15	77.05	C10
74.3	A10	79.8	B10	77.05	C10	76.4	A21	77.9	B1	76.95	C3
72.3	A11	71.4	B11	71.85	C11	75.9	A23	77.4	B23	76.65	C9
63.3	A12	72.4	B12	67.85	C12	74.3	A10	77.1	B14	76.65	C23
65.9	A13	73.9	B13	69.9	C13	74.2	A2	77	B27	76.25	C5
72	A14	77.1	B14	74.55	C14	74	A16	76.9	B4	75.85	C7
72.4	A15	78.4	B15	75.4	C15	73.3	A24	76.7	B18	75.4	C15
74	A16	67.3	B16	70.65	C16	73.1	A18	76.6	B20	75.2	C22
78.8	A17	80.3	B17	79.55	C17	72.9	A9	76.4	B19	75	C2
73.1	A18	76.7	B18	74.9	C18	72.4	A15	76.1	B24	74.9	C18
72.2	A19	76.4	B19	74.3	C19	72.3	A11	75.8	B2	74.7	C24
77.8	A20	76.6	B20	77.2	C20	72.2	A19	75.6	B3	74.55	C14
76.4	A21	79.2	B21	77.8	C21	72	A14	75.5	B6	74.3	C19
71	A22	79.4	B22	75.2	C22	71	A5	74.3	B25	71.95	C6
75.9	A23	77.4	B23	76.65	C23	71	A22	74.2	B7	71.85	C11
73.3	A24	76.1	B24	74.7	C24	70.4	A8	73.9	B13	71.35	C8
77.1	A25	79.5	B25	78.3	C25	68.4	A6	72.4	B12	70.9	C27
81.3	A26	74.3	B25	77.8	C26	65.9	A13	72.3	B8	70.65	C16
64.8	A27	77	B27	70.9	C27	64.8	A27	71.4	B11	69.9	C13
81.3	A28	79.6	B28	80.45	C28	63.3	A12	67.3	B16	67.85	C12

(a) 27种红葡萄酒酒样的三种排序比较

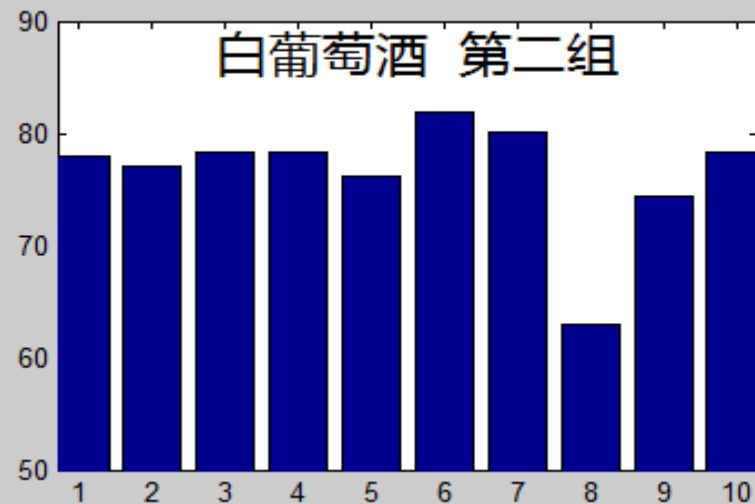
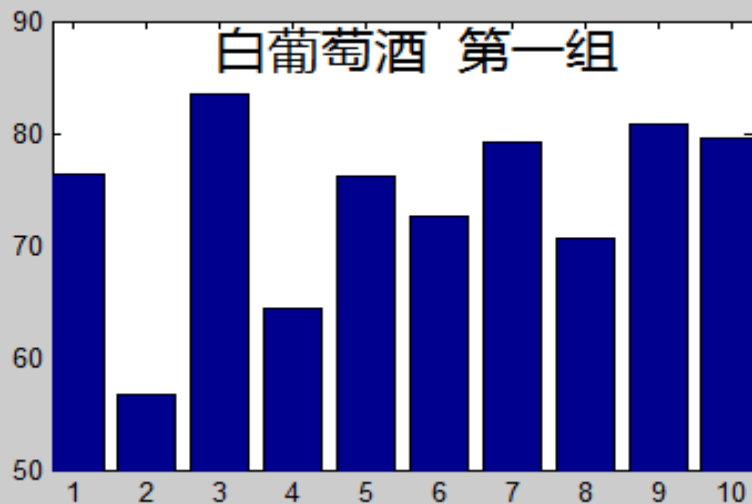
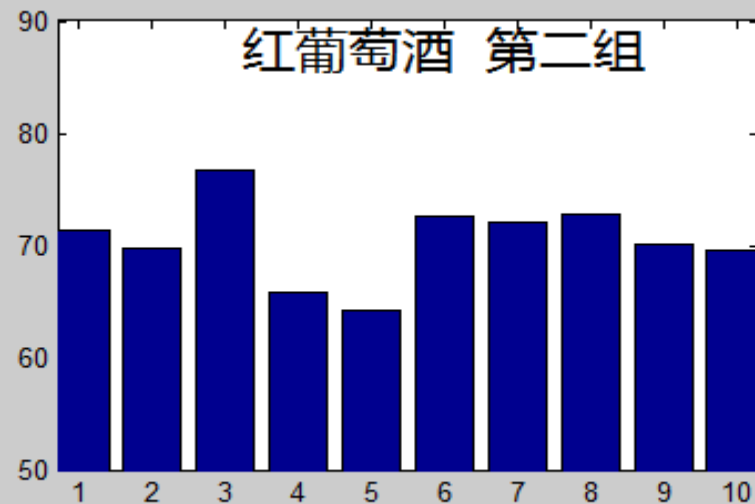
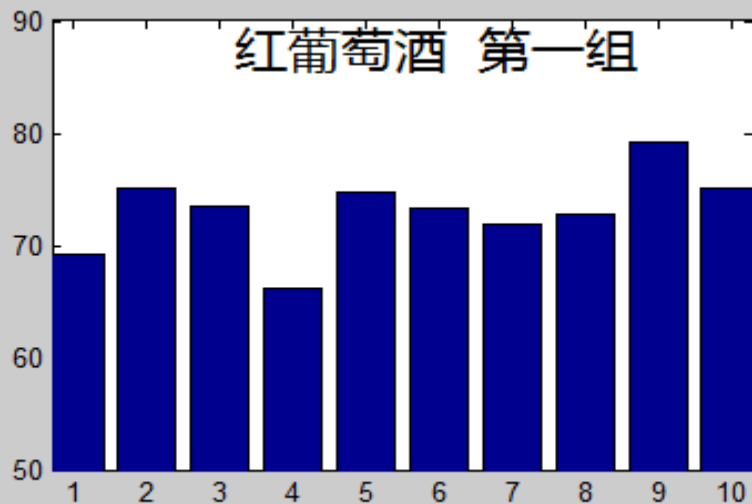


带点的线：第一组排序；带圈的线：两组平均排序；带星号的线：第二组排名

(a) 28种白葡萄酒酒样的三种排序比较

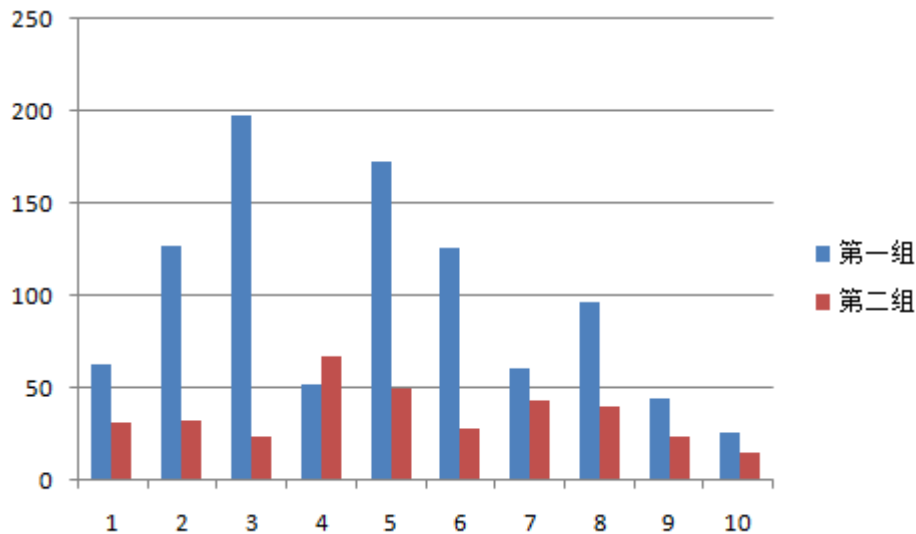


红白葡萄酒两组评酒员的平均分

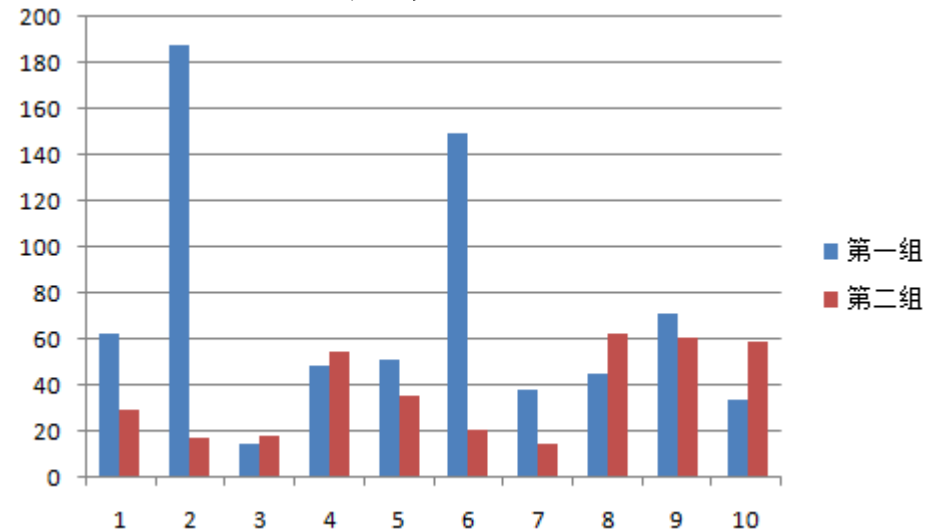


两组每个评酒员对27/28个酒样评分的方差

红葡萄酒



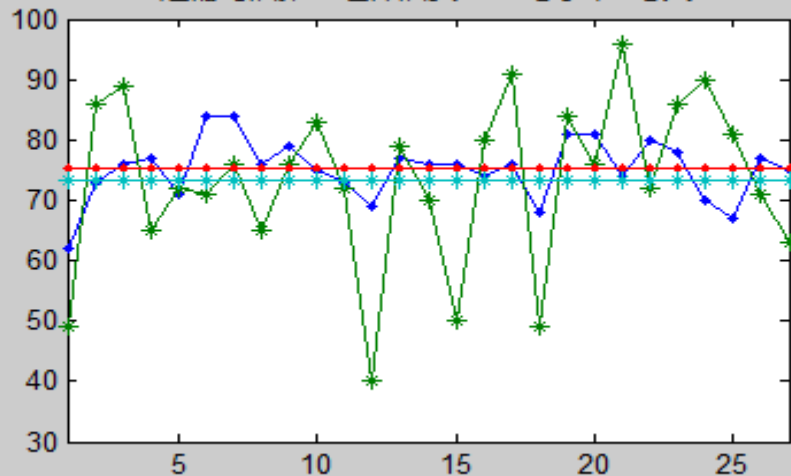
白葡萄酒



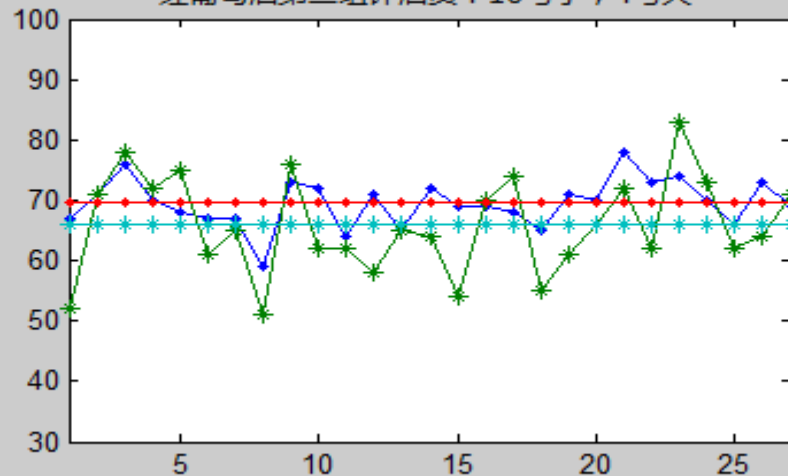
- 红葡萄酒和白葡萄酒都是第一组评酒员的方差大；
- 红葡萄酒第一组评酒员中**10号**评酒员的方差最小； **3号**评酒员的方差最大； 第二组评酒员中**10号**评酒员的方差最小； **4号**评酒员的方差最大；
- 白葡萄酒第一组评酒员中**3号**评酒员的方差最小； **2号**评酒员的方差最大； 第二组评酒员中**7号**评酒员的方差最小； **8号**评酒员的方差最大；

两组评酒员对红、白葡萄酒评分方差最小和最大的评分

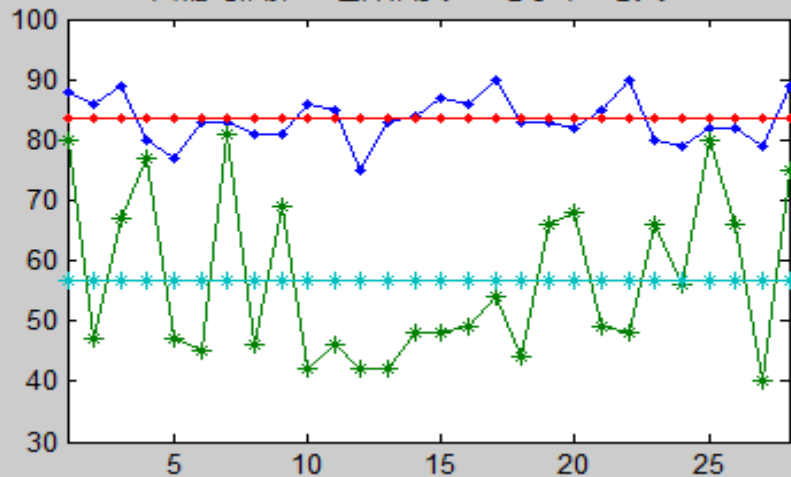
红葡萄酒第一组评酒员：10号小，3号大



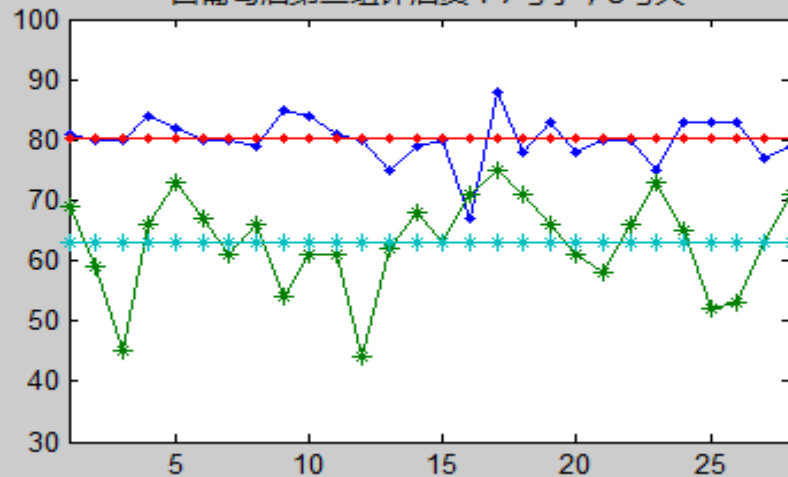
红葡萄酒第二组评酒员：10号小，4号大



白葡萄酒第一组评酒员：3号小，2号大

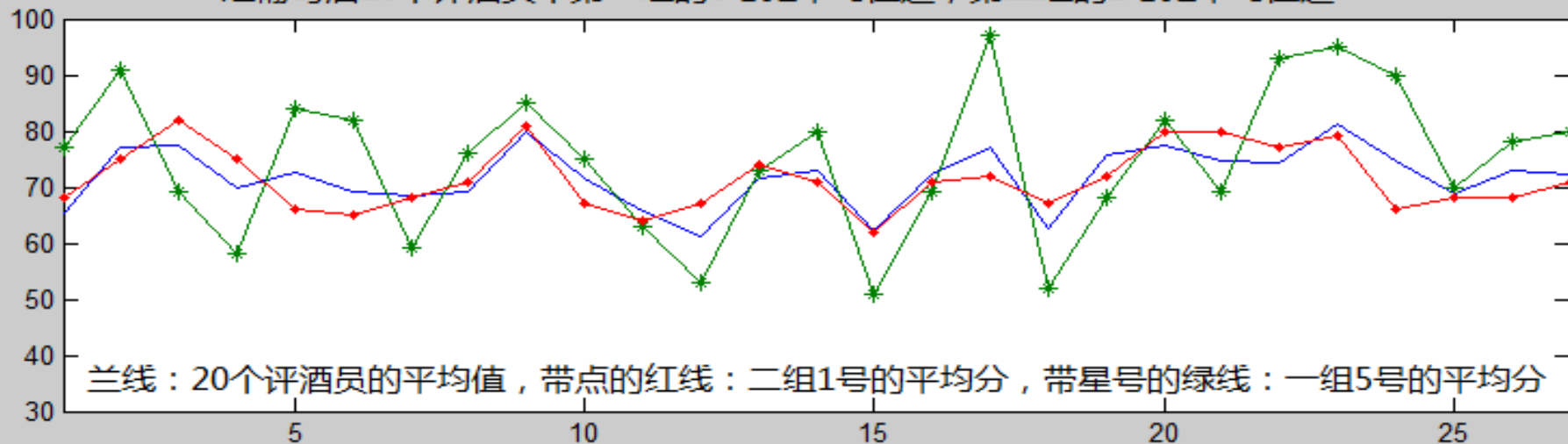


白葡萄酒第二组评酒员：7号小，8号大

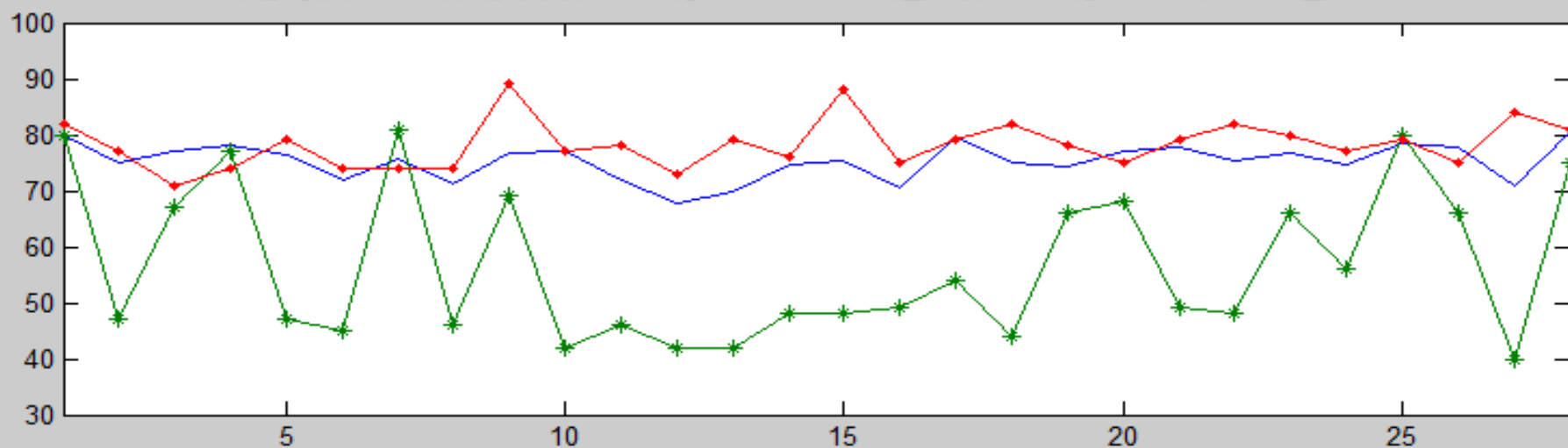


20名评酒员对红、白葡萄酒评分均值及比较

红葡萄酒20个评酒员中第一组的5号距平均值远，第二组的1号距平均值近

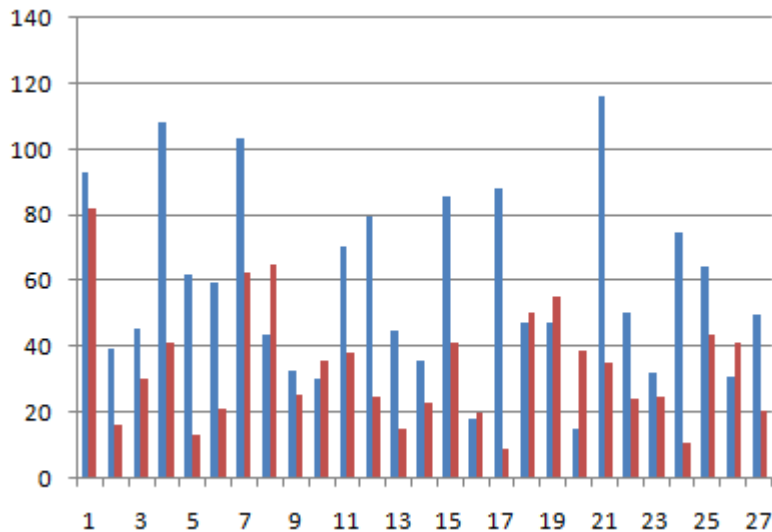


白葡萄酒20个评酒员中第一组的2号距平均值远，第二组的3号距平均值近

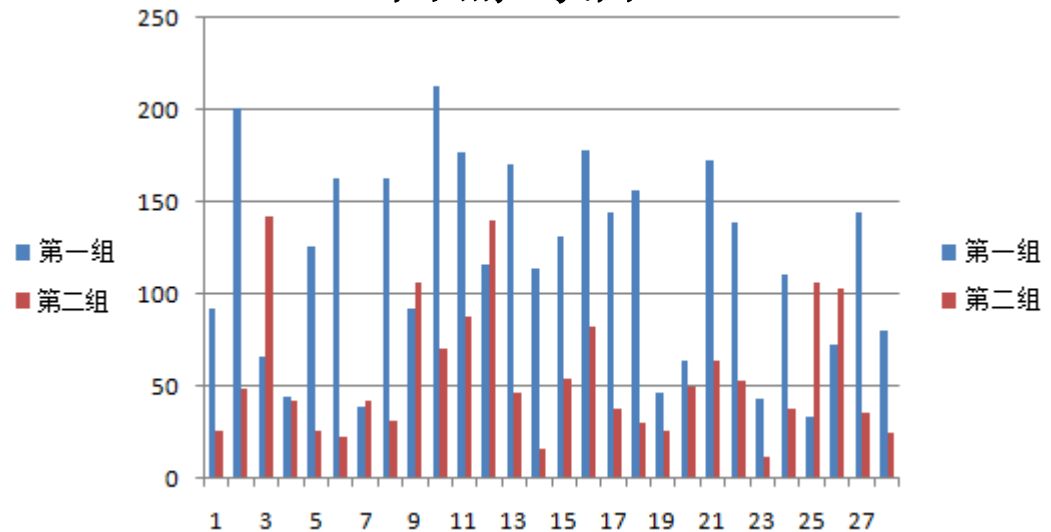


两组各10个评酒员对每个酒样评分的方差

红葡萄酒



白葡萄酒

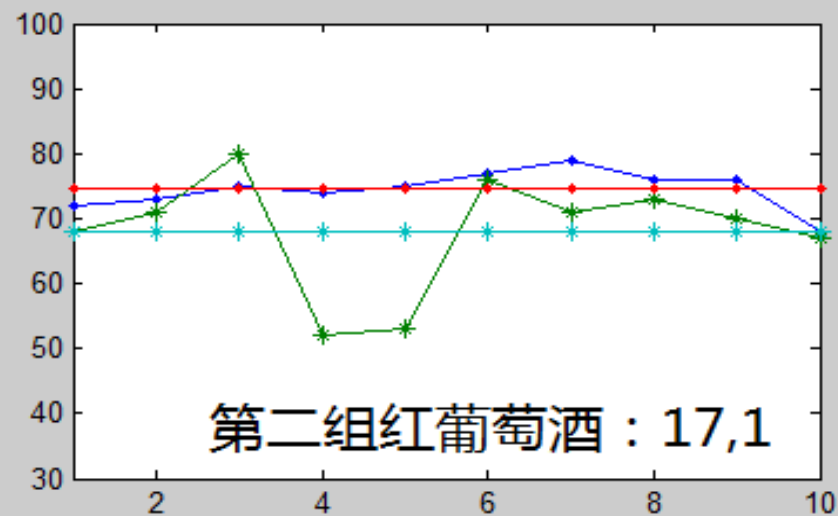
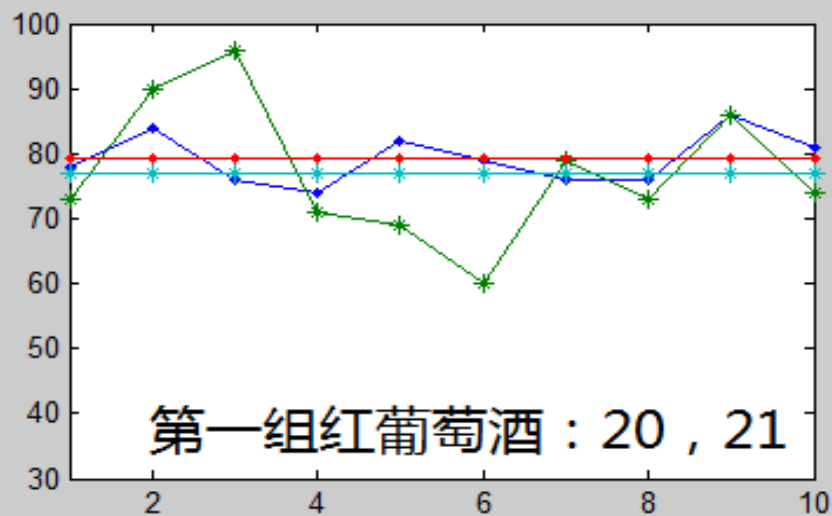


红葡萄酒和白葡萄酒都是第一组的方差大；

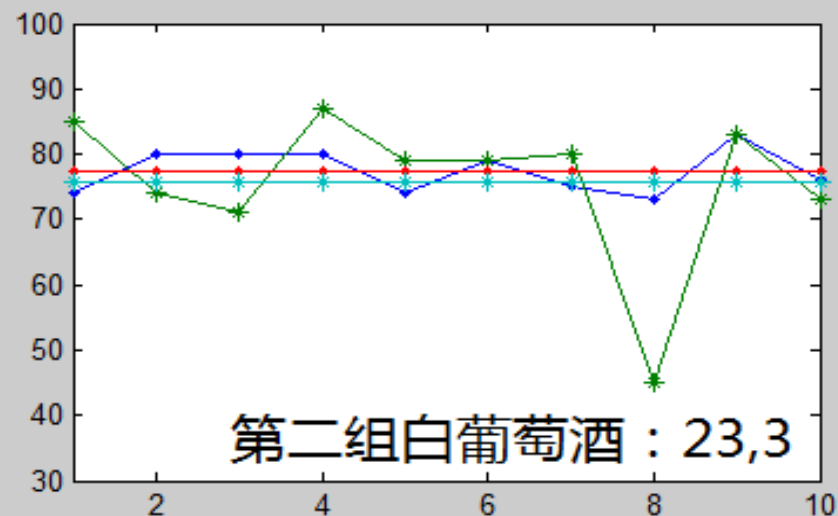
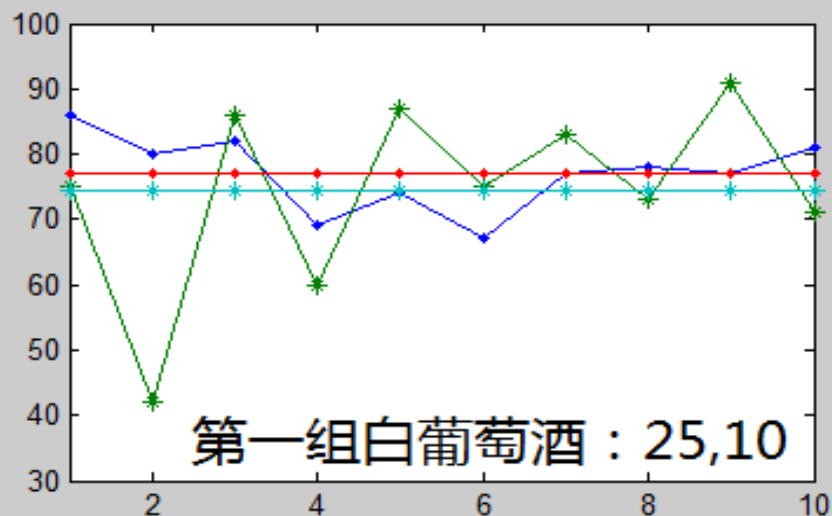
红葡萄酒第一组评酒员对**20**、**16**号酒样评价评价相对较一致，对**4**、**7**、**21**号酒样的评价分歧较大；第二组评酒员对**17**、**24**、**5**号酒样评价较一致，对**1**、**7**、**8**号酒样的评价分歧较大；

白葡萄酒第一组评酒员对**25**、**7**号酒样评价较一致，对**2**、**10**号酒样的评价分歧较大；第二组评酒员对**23**、**14**号酒样评价较一致，对**3**、**12**号酒样的评价分歧较大；

每组平均分方差最小和最大的葡萄酒样品的评分



带点的线是方差小的酒样的评分，带星号的线是方差大的酒样的评分



A335: 方差分析给出哪一组更可信

- 考虑两组评酒员对酒样评分的差异是否显著，可分别构造以评酒员为组别、元素为单个酒样的评分方差数据序列和以酒样为组别、元素为单个评酒员对全部酒样的评分方差数据序列，基于两类数据序列进行双向显著性检验，通过比较两者的检验量给出结论。
- 数据处理的方法有多种，如标准化处理、聚类处理和区间收敛处理等，可

方差分析的基本思路

构造 F 统计量

$$F = \frac{MS_R}{MS_E} = \frac{SS_R / (m-1)}{SS_E / (n-m)} \sim F(f_R, f_E) = F(m-1, n-m)$$

取一个显著性水平 α (0.01 或 0.05)，可查表得到 $F_\alpha(m-1, n-m)$ ，计算 $F(m-1, n-m)$ 与 $F_\alpha(m-1, n-m)$ 比较：当 $F(m-1, n-m) > F_\alpha(m-1, n-m)$ 时，显著

用置信区间法等对数据进行处理，居然后进行F值的比较

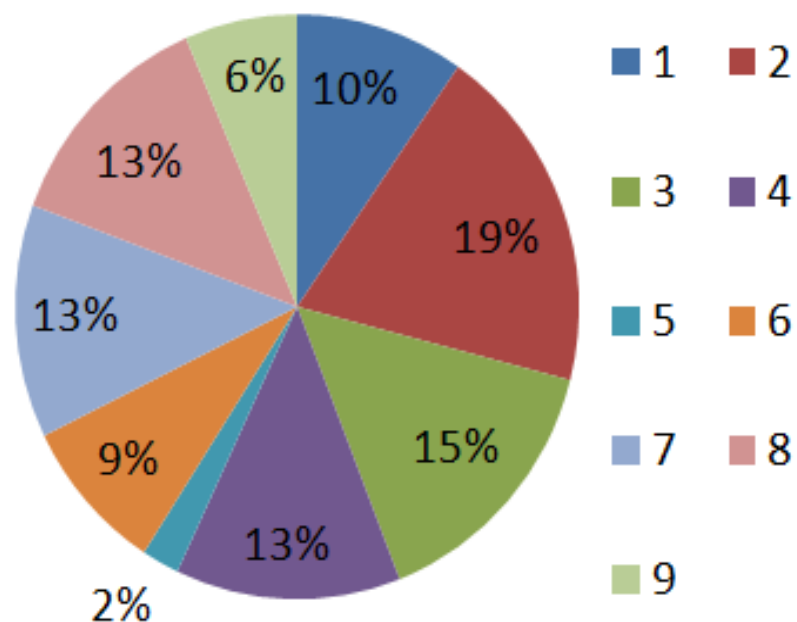
第一组 红葡萄酒	差异源	<i>F</i>	<i>F-crit</i>	第一组 白葡萄酒	差异源	<i>F</i>	<i>F-crit</i>
	行	7.108117	2.483848		行	17.95678	2.480975
	列	39.31581	1.837216		列	10.99991	1.81848
第二组 红葡萄酒	差异源	<i>F</i>	<i>F-crit</i>	第二组 白葡萄酒	差异源	<i>F</i>	<i>F-crit</i>
	行	8.477539	2.483848		行	6.361323	2.480975
	列	21.6988	1.837216		列	8.758393	1.81848

- 比较红葡萄酒评价结果的“行”区组的F 统计量，第一组**F=7.1**，小于第二组的值**F= 8.5**，表示第二组评酒员因主观因素造成的酒样评价差异，相较于第一组评酒员更明显。再比较红葡萄酒评价结果的“列”区组F 统计量，第一组的值为**F = 39.3**，大于第二组的值**F = 21.7**，表示第一组评酒员因酒样质量造成的评价差异比第二组评酒员更明显。所以，第一组评酒员对红葡萄酒的质量评价结果更可信。
- 白葡萄酒的两组数据，由于第一组白葡萄酒的“行”区组F 统计量**F =18.0**显著高于“列”区组F 统计量**F=11.0** 表示该组评价结果中评酒员因主观因素造成的酒样评价差异高于因酒样质量造成的评价差异；第二组白葡萄酒的“列”区组F 统计量大于“行”区组，优于第一组的评价结果。因此，第二组评酒员对白葡萄酒的质量评价结果更可信。

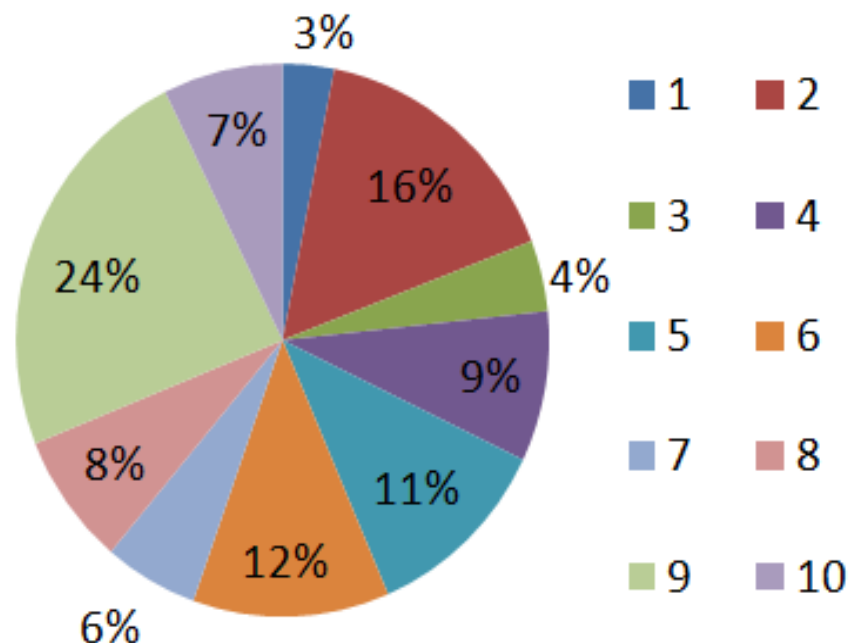
一个好的建议

- 分析比较每个评酒员的信度；
- 根据信度比重，量化每个品酒员在今后评酒中所占的权重

红葡萄酒各品酒员打分权重



白葡萄酒各品酒员打分权重



问题2

- 问题：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
- 评阅要点：给出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的原则、模型、算法和结果。确定酿酒葡萄质量好坏的主要依据是问题1中评酒员对酒的质量的评价结果，根据这个评价结果和酿酒葡萄的各种理化指标给出确定葡萄质量的模型，由此给出这些酿酒葡萄的分级结果。参考：分级结果中好的红葡萄应包含样品**23**，差的应该包含样品**12**。
- 说明：葡萄分级模型中应该根据第一问中葡萄酒质量的评价结果和葡萄的理化指标两个方面进行，分级的理化指标不要太多；确定红葡萄分级的指标中应该包含花色苷、单宁、总酚，白葡萄分级指标中应该包括总酚，而不包括花色苷，也可以给出这些指标的分级区间；分级不要太多，一个级别中包含的葡萄数量不宜太少。

问题2的解法

- 回归模型；
- 相关性分析；
- 聚类分析法，定义一些距离（欧氏距离，Minkowski距离），利用聚类分析法按不同的原则进行聚类，根据评分决定类别的高低；
- 主成分分析法；
- 因子分析降维，综合评判，隶属度函数进行分级；
- 根据酿酒葡萄的理化指标分类，根据葡萄酒的评分高低分类，对两者进行加权处理
- SOM神经网络的聚类；
- 层次分析法

将酿酒葡萄分为3~4个级别比较合适，有些对分类太多5,6，7,8个级别的都有，有些级就一个品种；没有给出分类的合理性说明：葡萄酒蛤蟆行业一般分几类，评分相近的是否在同一类，有没有差的和好的同类现象？

A441

- 根据第一问中的结果，建立葡萄的理化指标与葡萄酒得分之间的对应关系，根据得分的多少对葡萄进行分级
- 优点：既考虑了葡萄的理化指标，又考虑了葡萄酒的评分对葡萄分级的影响
- 具体实施方案：先对葡萄的理化指标进行筛选，得到一些主要理化指标的关系，再进行分级
- 筛选的方法：利用多元统计分析中逐步回归的思想，将葡萄酒的评分作为应变量，葡萄的理化指标作为自变量。把酿酒葡萄的理化指标逐个加入到函数中进行拟合，通过统计量的检验判断其影响是否显著，显著时保留，否则删除
- 筛选的好处：能保留对酒的质量有较大影响的理化指标，又能减少模型的复杂性

筛选出的理化指标

红葡萄:

蛋白质、VC、花色苷、苹果酸、多酚氧化酶活力、DPPH 自由基、酒总黄酮、白藜芦醇、还原糖、固酸比、干物质含量、百粒质量、出汁率、果皮质量、色泽(a);

白葡萄:

氨基酸总量、蛋白质、柠檬酸、褐变度、DPPH自由基、总酚、单宁、酒总黄酮、白藜芦醇、黄酮醇、还原糖、固酸比、干物质含量、果穗质量、百粒质量、色泽(a)。

得到的回归方程分别为

$$y = -0.097x_2 + 0.054x_3 + 0.153x_4 - 0.302x_6 - 0.236x_8 + 0.319x_{10} + 0.212x_{13} \\ - 0.150x_{14} + 0.266x_{17} + 0.287x_{21} + 0.148x_{22} - 0.091x_{24} - 0.137x_{26} + 0.161x_{27} - 0.144x_{29}$$

$$y = 0.131x_1 - 0.296x_2 + 0.192x_7 + 0.439x_9 + 0.164x_{10} + 0.914x_{11} + 0.156x_{12} - 0.922x_{13} \\ + 0.117x_{14} - 0.406x_{15} - 0.641x_{17} - 0.240x_{21} + 0.732x_{22} - 0.184x_{23} + 0.373x_{24} + 0.177x_{29}$$

两类葡萄的分级结果

红葡萄样本分级

葡萄样本	1	2	3	4	5	6	7	8	9	10	11	12	13	14
分级	3	1	1	2	2	3	3	3	1	3	4	3	3	2
葡萄样本	15	16	17	18	19	20	21	22	23	24	25	26	27	
分级	3	3	2	3	2	1	2	2	1	2	3	2	3	

白葡萄样本分级

葡萄样本	1	2	3	4	5	6	7	8	9	10	11	12	13	14
分级	2	2	2	1	1	3	2	3	1	1	3	3	2	2
葡萄样本	15	16	17	18	19	20	21	22	23	24	25	26	27	28
分级	1	4	1	2	2	2	1	1	2	2	1	2	2	1

根据分级情况也可以提出一些建议

酿酒葡萄加工适宜性评价标准⁴

分	总酚 (mg/g) ⁴		果皮 L* ⁴		羧基类 (g/L) ⁴		花色苷 (mg/g) ⁴		果梗比 (%) ⁴	
类	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴
I ⁴	≥4.5 ⁴	19 ⁴	≥27.5 ⁴	4 ⁴	≥100 ⁴	11 ⁴	≤0.5 ⁴	21 ⁴	≥4.7 ⁴	11 ⁴
II ⁴	4.5-2.6 ⁴	17 ⁴	27.5-26.5 ⁴	2 ⁴	100-50 ⁴	8 ⁴	0.5-0.6 ⁴	20 ⁴	4.7-3.5 ⁴	10.5 ⁴
III ⁴	2.6-2.05 ⁴	13 ⁴	26.5-25.5 ⁴	1 ⁴	50-20 ⁴	5 ⁴	0.6-4 ⁴	18 ⁴	3.5-2.6 ⁴	9 ⁴
IV ⁴	≤2.05 ⁴	8 ⁴	≤25.5 ⁴	0.5 ⁴	≤20 ⁴	4 ⁴	≥4 ⁴	4 ⁴	≤2.6 ⁴	1 ⁴
分	酯类 (g/L) ⁴		固酸比 ⁴		出汁率 ⁴		百粒质量 (g) ⁴		酒石酸 (g/L) ⁴	
类	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴	含量范围 ⁴	分值 ⁴
I ⁴	≤35 ⁴	19 ⁴	≥39 ⁴	5 ⁴	≥75 ⁴	4 ⁴	≥300 ⁴	3 ⁴	≥9.5 ⁴	3 ⁴
II ⁴	35-1000 ⁴	18 ⁴	39-35 ⁴	4 ⁴	75-70 ⁴	3 ⁴	300-230 ⁴	2 ⁴	9.5-5.8 ⁴	2 ⁴
III ⁴	1000-4000 ⁴	10 ⁴	35-27.5 ⁴	2 ⁴	70-66 ⁴	2 ⁴	230-150 ⁴	1.5 ⁴	5.8-3.5 ⁴	1 ⁴
IV ⁴	≥4000 ⁴	5 ⁴	≤27.5 ⁴	1 ⁴	≤66 ⁴	0.5 ⁴	≤150 ⁴	0.5 ⁴	≤3.5 ⁴	0.5 ⁴

问题3

问题：分析酿酒葡萄与葡萄酒的理化指标之间的联系。

评阅要点：给出分析酿酒葡萄与葡萄酒的成分之间关系的原理、模型和方法，得到葡萄酒的理化指标是否与葡萄的理化指标相关的结论，相关时给出具体的依赖关系。求解时最好先对葡萄的理化指标（包括芳香物质）进行分类和筛选，然后进行评价。注：仅把葡萄的全部理化指标进行简单回归不够完整。

说明：先应该对理化指标（包括芳香物质）进行分类整理和筛选，按照一级求和或者平均；最好给出通过葡萄理化指标来确定酒的理化指标的模型（表达式）；葡萄和酒中共有成分应该有较强的正相关关系；与葡萄酒的每个成分相关性较大的葡萄成分数量最好不超过**10**个。

先进行筛选，再进行回归

$$y_1 = 0.441x_4 - 0.085x_{26} - 0.085x_{29} + 0.209x_{12} + 0.461x_6 + 0.201x_{28} + 0.122x_{15} \\ + 0.057x_7 + 0.306x_{10} - 0.199x_{13} - 0.104x_{25}$$

$$y_2 = 0.347x_9 + 0.360x_1 + 0.396x_{18} + 0.360x_{10} - 0.250x_{16} + 0.138x_{22} - 0.081x_{25} \\ + 0.059x_{24}$$

$$y_3 = 0.386x_{13} + 0.215x_4 + 0.358x_1 - 0.183x_{26} + 0.789x_{10} - 0.189x_2 + 0.111x_{22} \\ + 0.335x_9 - 0.156x_{25} - 0.195x_{15} - 0.262x_{11} + 0.072x_{12}$$

$$y_4 = 0.671x_{11} + 0.160x_9 - 0.155x_{25} + 0.136x_{28} + 0.194x_{13} + 0.144x_4$$

$$y_5 = 0.951x_{13} + 0.334x_1 - 0.479x_{11}$$

$$y_6 = -0.303x_2 + 0.413x_1 + 0.375x_9 + 0.249x_{13} - 0.316x_{25} + 0.817x_{10} + 0.077x_{18} \\ + 0.086x_{28} - 0.055x_{29}$$

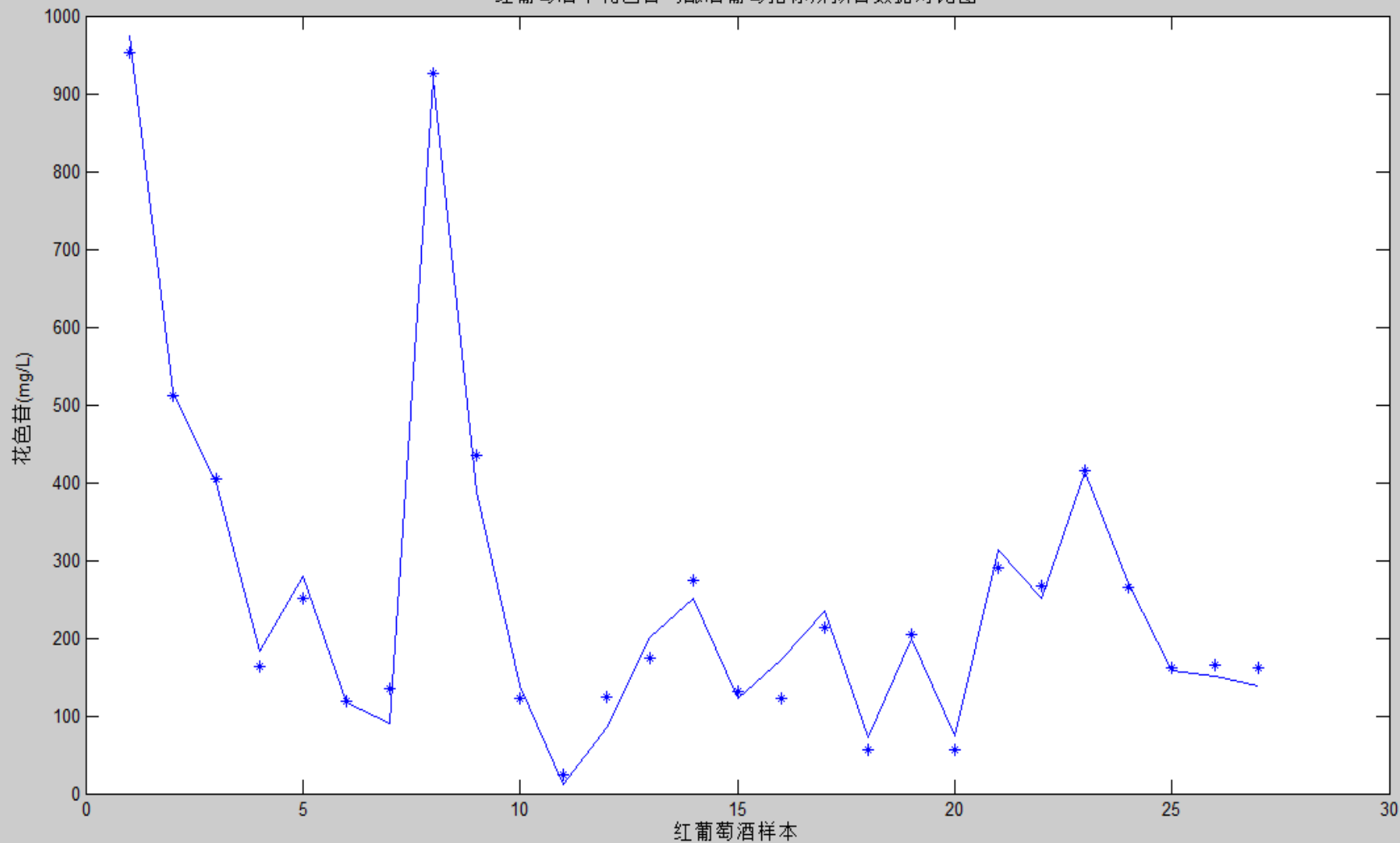
$$y_7 = -0.280x_4 + 1.100x_{29} - 0.221x_{15} - 0.638x_{10} - 0.666x_{30} + 0.274x_{24} - 0.216x_{28} \\ + 0.206x_{13} + 0.415x_{25} + 0.102x_2$$

$$y_8 = -0.917x_{29} - 0.527x_4 - 0.177x_{21} + 0.226x_{14} - 0.123x_9$$

$$y_9 = 0.540x_{17} - 0.383x_6 + 0.324x_5 - 0.137x_3$$

结果的对比与验证（花色苷）

红葡萄酒中花色苷与酿酒葡萄指标所拟合数据对比图



红葡萄酒与红葡萄理化指标回归分析结果

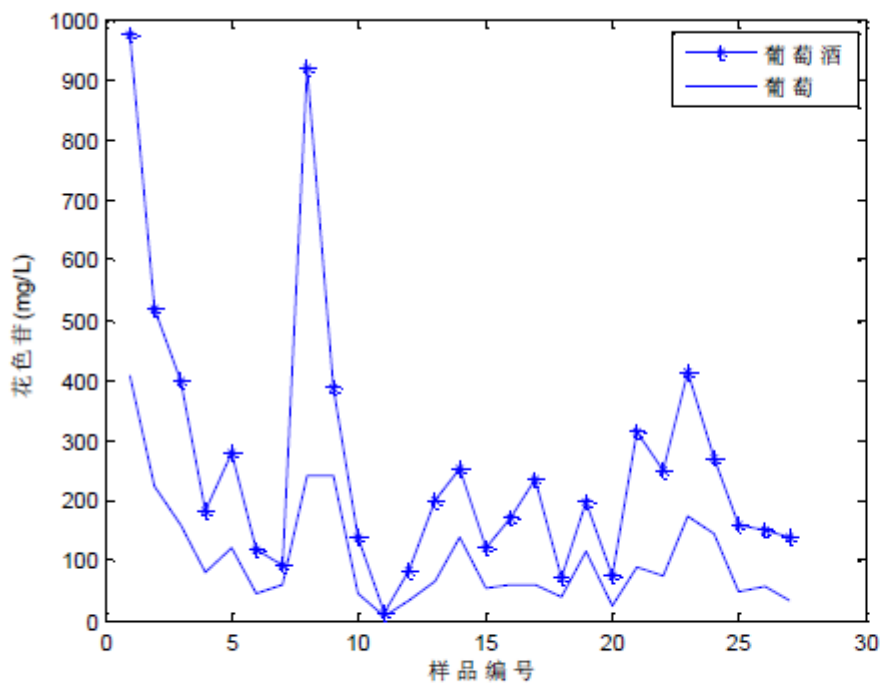
指标	相关自变量	回归方程	R 平方
花色(mg/L)	x4 x6 x9	$y=0.71*x4+0.11*x6+0.20*x9$	0.8883
单宁(mmol/L)	x4 x11 x12 x15	$y=0.173*x4+0.517*x11+0.09*x12+0.3*x15$	0.7886
总酚(mmol/L)	x4 x10 x11 x13	$y=0.33*x4+0.14*x10+0.29*x11+0.25*x13$	0.8276
酒总黄酮(mmol/L)	x4 x11 x13 x26	$y=0.205*x4+0.513*x11+0.276*x13-0.05*x26$	0.8012
白藜芦醇(mg/L)	x1 x11 x13	$y=0.334*x1-0.48*x11+0.953*x13$	0.43
DPPH 半抑制体积	x1 x2 x4 x10 x13	$y=0.285*x1-0.3*x2+0.229*x3+0.49*x10+0.4*x13$	0.8588
L*(D65)	x4 x10 x15 x29 x30	$y=-0.4*x4-0.28*x10-0.2*x15+0.68*x29-0.37*x30$	0.8965
a*(D65)	x6 x21 x29	$y=-0.43*x6-0.27*x21-0.47*x29$	0.6143
b*(D65)	x5 x17 x22	$y=0.44*x5+0.74*x17-0.29*x22$	0.5122

一些求解方法

- 相关性分析，典型相关性分析；
 - 多元回归分析，偏最小二乘回归分析；
 - 多因素优势的灰色关联度分析；
 - 主成分分析。
-
- 根据附录三中得到葡萄的芳香类理化指标，根据其化学性质，将其划分为酯类，芳香烃类，醛类，萜类。以各类芳香物质的相对分子质量作为权重，进行加权平均。

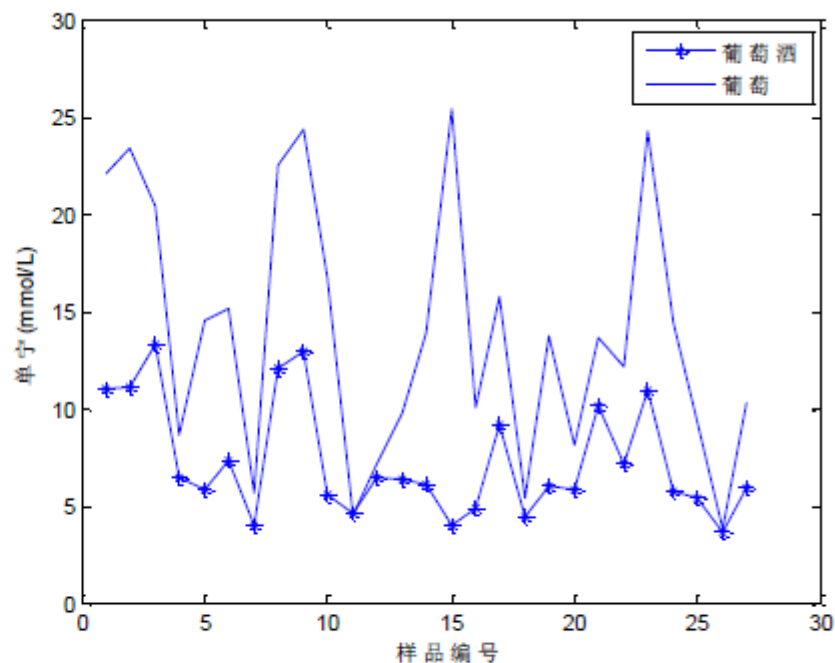
红葡萄酒与葡萄酒中一些指标的直观比较

花色苷



花色苷相关性曲线

单宁



单宁相关性曲线

问题4

问题：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？

评阅要点：建立模型分析酿酒葡萄和葡萄酒的理化指标与葡萄酒质量之间的关系，在模型的基础上给出具体结论，并对结论给出详细的分析说明。注：评价葡萄酒质量时不一定需要包含所有的理化指标，但根据经验知道花色苷、总酚和单宁是红葡萄酒的重要指标。

说明：在第二问基础上给出葡萄酒质量与葡萄酒理化指标及葡萄的理化指标关系的模型；在模型的基础上给出结论及其说明。红葡萄酒的理化指标与酒的质量关系比较密切红葡萄的理化指标与酒的质量的关系稍差一些；白葡萄酒的理化指标、白葡萄的理化指标与其对应酒的质量的关系与红酒相比更差一些。

方法：回归模型，神经网络，主成分分析，判别分析

葡萄酒的理化指标与葡萄酒质量的回归模型

- 葡萄酒的全部11个指标的回归模型

$$Y=0.029 \times \text{花色苷}+1.711 \times \text{单宁}-0.910 \times \text{总酚}+1.995 \times \text{酒总黄酮}+0.833 \times \text{白藜芦醇}-6.839 \times \text{DPPH半抑制体积}+0.456 \times L^*(D65)-1.639 \times a^*(D65)-0.474 \times b^*(D65)+0.174 \times H(D65)+2.172 \times C(D65)$$

- 葡萄酒的7个指标的回归模型

$$Y=0.018 \times \text{花色苷}+4.16 \times \text{单宁}+1.474 \times \text{总酚}+1.571 \times \text{酒总黄酮}+2.126 \times \text{白藜芦醇}-68.402 \times \text{DPPH半抑制体积}+0.667 \times L^*(D65)$$

- 芳香成份测试难度和成本高，故不选芳香成份

A004: 桂林理工大学

- 分析葡萄酒的理化指标与葡萄酒质量的相关性，通过比较选出与葡萄酒质量指标相关性程度大的葡萄酒成分，结合第三问中建立的酿酒葡萄的理化指标和葡萄酒理化指标之间的关系，计算相关性系数，用回归分析法建立酿酒葡萄的理化指标与葡萄酒质量之间的关系。
- 附表一中列出了十位品酒员对葡萄酒外观、香气和口感分析三者的数据，分别对四项指标求27（28）种红（白）葡萄酒样品权重平均值作为葡萄酒质量的评价指标。

主要步骤

对相关系数、回归方程、回归方程效果的检验进行了简单介绍；分别建立酿酒葡萄的理化指标与葡萄酒质量所涉及的外观、香气、口感、整体四个方面的回归关系，最后加权得到总的回归方程

外观与红葡萄酒理化指标的线性回归方程为

$$f_1(x_i) = 0.001x_1 + 0.237x_2 + 0.031x_3 + 6.442$$

其中 x_1 、 x_2 、 x_3 为花色苷、白藜芦醇、和L*(D65)的含量

香气与红葡萄酒理化指标的线性回归方程为

$$f_2(x_i) = 0.045x_1 + 0.043x_2 + 0.279x_3 + 4.798$$

其中 x_1 、 x_2 、 x_3 分别代表单宁、总黄酮、白藜芦醇的含量、

口感与红葡萄酒理化指标的线性回归方程为

$$f_3(x_i) = 0.246x_1 - 0.545x_2 + 0.253x_3 + 0.748x_4 + 0.298x_5 + 10.633$$

x_1 、 x_2 、 x_3 、 x_4 、 x_5 为单宁、总酚、总黄酮、白藜芦醇、H平均的含量、

平衡及整体评价与红葡萄酒理化指标的回归方程

$$f_4(x_i) = 0.078x_1 - 0.164x_2 + 0.21x_3 + 7.616$$

其中 x_1 、 x_2 、 x_3 为总黄酮、H平均、C平均的含量

综合评价的回归方程

$$F = 0.15f_1(x_i) + 0.3f_2(x_i) + 0.44f_3(x_i) + 0.11f_4(x_i)$$

优点：分四个方面建立回归方程是将复杂问题分解为几个简单的情况；
最后给出的结果简单、明确、实用；

不足：没有给出效果的实际检验；

建议：建立回归方程时留几组数据，最后检验质量评价的效果。

红葡萄酒的评分方式确定

评分 _ρ	因子及多元回归方程 _ρ
外观 F1 _ρ	<p>H1: 红葡萄酒花色苷 H2: 红葡萄酒色泽L*(D65) ρ</p> <p>H3: 红葡萄酒色泽a*(D65) H4: 红葡萄酒色泽b*(D65) ρ</p> <p>H5: 红葡萄酒色泽H(D65) H6: 红葡萄酒色泽C(D65)_ρ</p> <p>$F1=13.058-0.0023H1-0.0555H2-0.438H3-0.397H4-0.739H5+0.596H6$</p>
香味 F2 _ρ	<p>H7: 红葡萄黄酮醇 H8: 红葡萄可溶性固形物 H9: 红葡萄酒总酚_{~~~~~}</p> <p>H10: 红葡萄酒白藜芦醇_{~~~~~} H11: 红葡萄酒芳香物_ρ</p> <p>$F2=20.5717-0.0068H7+0.0180H8+0.7108H9+0.1059H10-0.0524H11$_ρ</p>
口感 F3 _ρ	<p>H7: 红葡萄黄酮醇 H8: 红葡萄固酸比_{~~~~~} H10: 红葡萄酒白藜芦醇_{~~~~~} ρ</p> <p>H12: 红葡萄酒单宁 H13: 红葡萄还原糖_ρ</p> <p>$F3=23.0102+0.0204H7+0.1447H8+0.5625H10+0.0174H12+0.0025H13$_ρ</p>
平衡/ 整体 F4 _ρ	<p>H9: 红葡萄酒总酚_{~~~~~} H10: 红葡萄酒白藜芦醇_{~~~~~} H11: 红葡萄酒芳香物</p> <p>H12: 红葡萄酒单宁 H13: 红葡萄还原糖 H14: 红葡萄总酸_ρ</p> <p>$F4=8.94-0.082H9+0.095H10-0.0083H11+0.16H12+0.0006H13-0.02H14$_ρ</p>

多元线性回归结果的验证

- 多元线性回归方程具有统计学意义，但是对与实际问题的拟合情况如何有待检验。
- 利用多元线性回归方程，代入相应指标进行验证，可以获得未进入回归分析的3组理化指标的各项分数

葡萄样品	外观		香味		口感		平衡/整体评价	
	原始分数	验证分数	原始分数	验证分数	原始分数	验证分数	原始分数	验证分数
红葡萄 25	10.4	10.8	21.7	21.3	28.8	27.2	8.3	8
红葡萄 26	11.4	11.2	23.5	23.2	30	39.4	8.9	8.6
红葡萄 27	9.9	10.1	21.6	21.3	32.5	33.1	9	9.3
白葡萄 25	10.60	10.20	24.20	24.6	35.10	34	9.60	9.4
白葡萄 26	11.10	10.80	22.00	22.3	32.00	33.2	9.20	9.2
白葡萄 27	10.90	9.50	23.80	22.6	33.00	33.5	9.30	9.4
白葡萄 28	11.00	10.00	24.10	24.5	35.00	35.4	9.50	9.3

- 由表中数据看出，两种分数间相差都不超过1，可认为回归模型符合实际，能用理化指标评价酒的质量

RBF 径向基函数神经网络

- RBF 径向基函数神经网络是一种三层前向网络，分别为输入层、隐藏层和输出层。在此，输出层为在某一项质量指标的得分，输入层为对应该项质量指标的几项理化指标。
- 将数据导入**SPSS**，分析求解后与人工评分平均值作误差分析。其中，选取“训练—检验比”约为**7:3**，得到用神经网络模型评价葡萄酒各方面指标的结果（红葡萄酒为例）

外观分析		香气分析		口感分析		平衡/整体		总和	
评分	模型	评分	模型	评分	模型	评分	模型	评分	模型
11.5	12.29	23.0	23.22	33.4	33.75	9.2	8.92	77.1	78.18
11.9	11.84	24.0	22.95	32.3	33.81	9.0	8.88	77.2	77.48
11.4	11.63	23.5	23.80	30.0	30.00	8.9	8.94	73.8	74.37
12.1	11.32	23.7	20.69	33.1	32.03	9.1	8.68	78.0	72.72
10.4	10.62	21.7	22.44	28.8	29.54	8.3	8.55	69.2	71.15
11.4	11.55	23.5	22.57	30.0	30.41	8.9	8.90	73.8	73.43
9.9	10.28	21.6	22.85	32.5	32.37	9.0	8.77	73.0	74.27

无用或者不合理的假设

- 葡萄酒质量完好，未变质；
- 葡萄的产地对其质量无影响；
- 评酒员的感官评价不受外界等因素的影响；
- 酿酒葡萄转化为葡萄酒的过程中基本无损耗；
- 葡萄酒的酿造工艺相同；
- 测量仪器没有误差；
- 不考虑葡萄中各个元素间的相互影响；
- 不考虑芳香物质的挥发带来的影响。

一些不满意的方面

- 大部分都没有验证评分的数据是否服从正态分布，直接进行检验（**t**检验、**F**检验）。也有些用多因素方差分析；只给出了一些统计量、或者软件计算后的表格、给出了结论。但没有说明从具体问题和数据方面说明有显著差异和无显著差异的实际含义，也没有说明哪一组结果更可信一些的实际含义。
- 没有给出分类合理性的直观解释，阐述评分排名和理化指标聚类如何结合，且未说明最后分类结果的合理性，未说明哪一类是优质酒、哪些是一般酒，哪些是差一些的酒。
- 对葡萄酒和酿酒葡萄理化指标关系的验证方法，从实际应用的角度说明筛选出主要成分的合理性，说明所建立的关系的好坏，主成分分析筛选出的主成分和那些指标对应，从实际情况说明这些指标的选取是否合理。
- 给出具体的评价模型，或者说明不能评价的原因。如果结论是理化指标能确定酒的质量。就需要给出通过理化指标的值确定酒的质量的模型（方法、公式），并对结果的合理性进行阐述。最好要对模型进行验证，用一些样本的数值确定参数，利用剩下的一些样本验证结果。

有的论文得到的结论实际意义不清楚

对于第三问，葡萄的理化指标和葡萄酒的理化指标是多对多的关系，首先对葡萄酒的理化指标做主成分分析，得到相互独立的主成分，将其转化为一组多对一的关系；然后建立多元线性回归模型，求解回归模型的系数矩阵；最后用逐步回归法分析回归显著性并剔除不显著参量。最终得到回归系数矩阵，体现了它们之间的关系，例如红葡萄酒理化指标第二主成分与红葡萄的理化指标各主成分的关系式为：

$$Y'_2 = 0.13Y_1 - 0.07Y_2 + 0.06Y_3 + 0.06Y_4 - 0.17Y_5 - 0.13Y_6$$

- 主要问题：没有将自己的分析结果和应用问题建立起联系，实际意义不清楚。

几个建议

- 方法或模型的使用：要理解方法或模型的原理和适用范围，验证前提条件是否满足，对结果进行说明；
- 软件的使用：尽量使用，但要给出较为仔细的描述，包括数据的准备，输出结果的说明；
- 解决一个问题是用模型的多少：尽量选恰当的模型或方法，不宜太多，几个模型时要进行比较，分析结果是否一致，或者哪一个更好一些；
- 结果的验证：除过数学方法的验证，还需要和实际的结果进行比较。
- 问题的分析和建模研究一定要和具体的实际情况联系

- “葡萄酒的评价”问题涉及的统计只是多，数据量大，与近几年的题目的类型有较大的差异
- 学习了相关的统计方法，给出了一些好的结果
- 希望我们的同学通过这个问题的解答过程更对统计方法的应用有更好的体会，在以后应用这些方法时更加关注这些方法的实际含义和结果的解释
- 希望我们的教师在统计课程的教学之中通过一些应用范例更多的介绍统计方法的应用过程，不仅使学生学会统计方法的使用过程，而且理解和关注解决问题时这些统计量的实际背景，使学生能更好的利用统计方法解决实际问题

谢谢大家！