

葡萄酒感官评价的差异性及可信度方法研究^①

曹志强¹, 刘娅楠¹, 杨婷¹, 马蕾¹, 齐亚娥²

(1. 河西学院数学与统计学院, 甘肃 张掖 734000; 2. 河西学院化学化工学院, 甘肃 张掖 734000)

摘要: 运用配对 T 检验和灰色关联性分析两种方法分析了评酒员间评价结果的差异性, 同时根据 Kendall 协调系数 W - 检验和标准差分析了评价结果的可信度. 结果表明: 评酒员的评价结果存在显著性差异, 得到了第二组评酒员评价结果更可信的评价结果.

关键词: 配对 T 检验; 灰色关联度分析; Kendall 协调系数; 标准差; 可信度

中图分类号: TS262 **文献标识码:** A

0 引言

葡萄酒的质量是其外观、香气、口味、典型性的综合表现, 主要依靠品酒员的感官进行评价^[1,2]. 由于葡萄酒质量与其成分密切相关且关系复杂, 以及受人为主观因素的影响, 不同评酒员对同一酒样的评价有差异. 因此, 如何度量评酒员感官评价的差异性以及可信度就成为一个非常重要的问题.

2006年, 李华等采用置信区间检验方法对该问题进行了研究, 该方法极大地降低了同一酒样的变异系数, 使不同品酒员对同一酒样的评价趋于一致, 在降低品酒员间差异性的同时不能反映品酒员间的真实差异^[3]. 2012年, 刘洋洋和薛凌云则采用 Wilcoxon 符号秩和检验和 Grubbs 准则对评酒员感官评价的差异性以及可信度问题做了研究^[4].

为更准确的分析不同评酒员对酒样的评价结果之间的差异, 基于 2012 年全国大学生建模竞赛 A 题数据样本^[5], 我们运用配对 T 检验和灰色关联性两种方法对评酒员的评价结果是否有显著性差异的问题进行了分析, 同时分别根据 Kendall 协调系数 W - 检验和标准差进行了可信度分析, 得出了比较可靠的结论.

1 感官评价的差异性分析

根据数据样本中两组评酒员的评价结果, 为

分析结果是否有显著性差异, 先对评酒员的数据进行预处理, 将葡萄酒品尝评分表中葡萄酒的评价结果按照葡萄酒样品按标号从小到大进行排序, 求出十个评酒员对葡萄酒分类指标打分的均值, 通过拟合平均数的图像来比较两组评酒员评价结果差异的显著性, 如图 1 所示.

从图 1 可以看出两组之间评价结果存在差异, 但观察得到的结果往往误差较大且可靠性不高, 下面通过统计方法来分析评价结果的差异性.

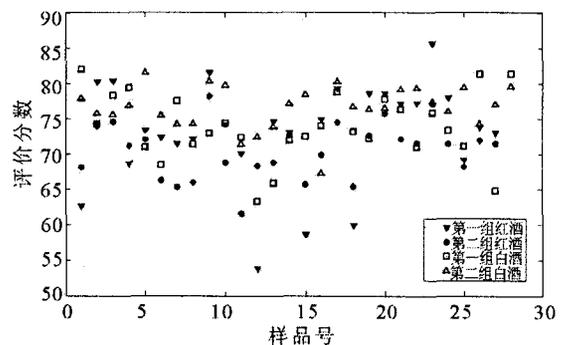


图 1 评价分数均值图

1.1 配对 T 检验

将红葡萄酒和白葡萄酒分开考虑, 即分别比较两组评酒员对红葡萄酒和白葡萄酒的评分有无显著性差异. 由于考虑的是两组评酒员分别对同种酒样品评价的结果, 数据是成对出现且相互影响的, 可认为总体近似服从正态分布, 其方差一

① 收稿日期: 2013-02-25

基金项目: 河西学院大学生科研活动项目(131).

作者简介: 曹志强(1990-), 男, 甘肃天水人, 河西学院学生,

通讯作者: 齐亚娥(1981-), 女, 甘肃陇西人, 河西学院讲师, 硕士.

致. 因此选用配对样本 T - 检验来对二者是否存在显著性进行检验^[6,7]. 根据配对样本 T 检验得出 t 统计量和 p 值, 如 $p < \alpha$, 则两组评分结果有显著性差异, 如 $p > \alpha$, 则两组评分结果无显著性差异.

用 SPSS 命令 Analyze→Compare mean →Paired - Samples T Test, 把需要检验的选项选入对话框, 可以得到检验结果, 如表 1 所示.

表 1 是相关性检验结果, 对两个样本进行相关性检验是对进行配对样本 T - 检验的前提, 即两个样本的数据之间的相关关系.

表 1 成对样本相关系数

| | | N | 相关系数 | Sig. |
|-----|-----------|-----|------|------|
| 对 1 | dt1 & dt2 | 27 | .700 | .000 |
| 对 2 | dt3 & dt4 | 28 | .207 | .291 |

表 2 成对样本检验

| 成对样本 | 成对差分 | | | | | t | df | Sig. (双侧) | |
|------|-----------|--------|--------|----------|---------|--------|--------|--------------|------|
| | 均值 | 标准差 | 均值标准误差 | 95% 置信区间 | | | | | |
| | | | | 下限 | 上限 | | | | |
| 对 1 | dt1 - dt2 | 2.5407 | 5.3718 | 1.0338 | .41569 | 4.6657 | 2.458 | 26 | .021 |
| 对 2 | dt3 - dt4 | -2.271 | 5.5038 | 1.0401 | -4.4056 | -.1372 | -2.184 | 27 | .038 |

表 2 是对两独立样本进行 T 检验的结果, 包括两配对样本差分的均值、标准差、均值的标准误差、95% 的置信区间, 还有 T 检验的 t 值、自由度 (df) 和显著性概率 ($Sig.$). “dt1 - dt2” 指第一组评酒员对红葡萄酒打分总和减去第二组评酒员对红葡萄酒打分总和, 即“成对差分”, 表中统计量即是对应于这种差值.

结果表明, 各种统计量的数值都比较大, 这说明这两组评价结果有差异, 再根据显著性概率 $p = 0.021 < 0.05$, 进一步证明在 95% 的置信水平上两组评酒员对红葡萄酒的评价结果存在显著性差异.

到在 95% 的置信水平上两组评酒员对白葡萄酒的评价结果差异也显著.

1.2 灰色关联度分析

基于灰色关联分析的基本思想^[8], 根据两组评酒员的打分情况, 对数据进行处理, 得到每位评酒员对每个酒样的评价总分, 记第一组评酒员对 28 个白葡萄酒、27 个红葡萄酒样品的评价总分为 Y , 第二组评酒员对 28 个白葡萄酒、27 个红葡萄酒样品的评价总分为 X_i , 计算出第一组参照序列 Y 与第二组比较数列 X_i 的关联度 v_i , 如表 3 和表 4 所示.

对白葡萄酒两组评价结果采用相同分析, 可得

表 3 白葡萄酒一、二两组评价结果的关联度

| 序号 | 关联度 | 序号 | 关联度 | 序号 | 关联度 | 序号 | 关联度 |
|----|--------|----|--------|----|--------|----|--------|
| 1 | 0.6526 | 8 | 0.6265 | 15 | 0.5951 | 22 | 0.5507 |
| 2 | 0.6407 | 9 | 0.6161 | 16 | 0.6565 | 23 | 0.5668 |
| 3 | 0.6863 | 10 | 0.5682 | 17 | 0.5951 | 24 | 0.5969 |
| 4 | 0.5069 | 11 | 0.7347 | 18 | 0.6027 | 25 | 0.5775 |
| 5 | 0.6797 | 12 | 0.6144 | 19 | 0.6776 | 26 | 0.6472 |
| 6 | 0.5592 | 13 | 0.5837 | 20 | 0.6870 | 27 | 0.6062 |
| 7 | 0.6265 | 14 | 0.7189 | 21 | 0.7304 | 28 | 0.5436 |

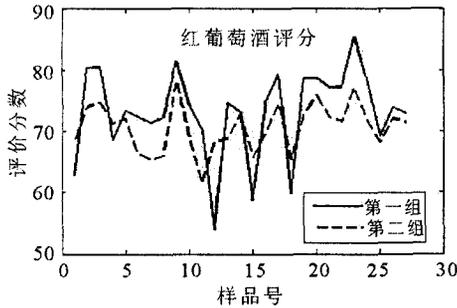
表 4 红葡萄酒一、二两组评价结果的关联度

| 序号 | 关联度 | 序号 | 关联度 | 序号 | 关联度 | 序号 | 关联度 |
|----|--------|----|--------|----|--------|----|--------|
| 1 | 0.6257 | 8 | 0.5932 | 15 | 0.6028 | 22 | 0.6592 |
| 2 | 0.6127 | 9 | 0.5966 | 16 | 0.6747 | 23 | 0.7257 |
| 3 | 0.6213 | 10 | 0.5572 | 17 | 0.6162 | 24 | 0.5241 |
| 4 | 0.5651 | 11 | 0.6125 | 18 | 0.6843 | 25 | 0.6094 |

| | | | | | | | |
|---|--------|----|--------|----|--------|----|--------|
| 5 | 0.6633 | 12 | 0.6694 | 19 | 0.7433 | 26 | 0.6191 |
| 6 | 0.5592 | 13 | 0.5893 | 20 | 0.6081 | 27 | 0.5139 |
| 7 | 0.5128 | 14 | 0.6037 | 21 | 0.5754 | | |

通过对第一组参照序列 Y 与第二组比较数列 X_i 的关联度 v_i 的分析,其关联度在 0.5 到 0.75 之间,因此可以认为两组评酒员的评价结果有显著性差异。

根据配对 T 检验和灰色关联度分析结果,两组评酒员的评价结果有显著性差异。



2 感官评价的可信度分析

以样本为 x 轴,运用 Matlab 绘制两组评酒员分别对红葡萄酒和白葡萄酒评价分数的均值图,根据图像的波动性可直观判定评价结果的可信度,如图 2 所示。

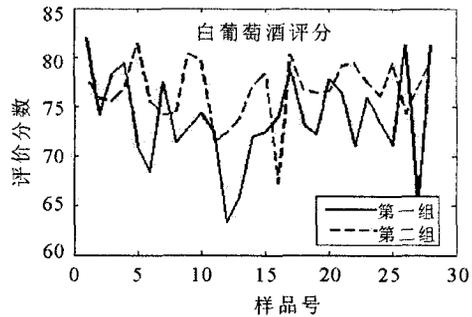


图2 分数均值图

从图 2 可以看出第一组评酒员评价波动较之于第二组评酒员评价更大,相对来说稳定性不高。因此,第二组评酒员的评价结果更可信。

为使结果更可靠,下面根据 Kendall 协调系数 W - 检验和标准差进行可信度分析。

根据两组评酒员(各 10 人)对 27(28) 种酒样品的评价分数,首先检验两组评酒员的评价结果是否具有-致性,然后将数据转化为序列数据,采用 Kendall 协调系数 W - 检验^[6,9] 的分析方法。

2.1 Kendall 协调系数 W 检验

原假设 H_0 : 两组评酒员对红(白)葡萄酒样品的评分没有-致性;

备择假设 H_1 : 两组评酒员对红(白)葡萄酒样品的评分存在-致性。

运行 SPSS 命令: Analyze → Nonparametric Test → K Related Samples, 把需要检验的选项选入对话框,就可以得到检验结果,如表 5 所示。

表 5 Test Statistics

| N | 10(dt1) | 10(dt2) | 10(dt3) | 10(dt4) |
|------------------|---------|---------|---------|---------|
| Kendall's $W(a)$ | 0.076 | 0.021 | 0.029 | 0.02 |
| Chi - Square | 19.772 | 5.549 | 7.809 | 5.377 |
| df | 26 | 26 | 27 | 27 |
| Asymp. Sig. | 0.802 | 1 | 1 | 1 |

从表 5 中可以看出,第一组评酒员对红葡萄

酒的 Kendall 调系数 $W = 0.076, \chi^2 = 19.772, Asymp. Sig = 0.802 > 0.05$, 由于卡方值 19.772 小于界值 $\chi^2_{(0.05,26)} = 38.885$, 故在 $\alpha = 0.05$ 水准上接受 H_0 , 即可认为第一组评酒员对红葡萄酒的评分有显著差异,也可认为第一组评酒员对红葡萄酒的评分不一致。

同理,用同样的方法可得到两组评酒员对红(白)葡萄酒的评分亦不一致的结论。

χ^2 值反映了实际频数与理论频数的吻合程度,若检验假设 H_0 成立,则实际频数与理论频数的差值会小, χ^2 值也会小;反之,若检验假设不成立,则实际频数与理论频数的差值会大, χ^2 值也会大。由上分析可知检验假设 H_0 成立,且由表 5 知第二组评酒员对红(白)葡萄酒评分的 χ^2 值均小于第一组,故实际频数与理论频数的差值小,即第二组评酒员的评价结果更可信。

2.2 标准差分析

比较两组的可信度,选择用标准差 S 的方法。将两组评酒员对红葡萄和白葡萄的打分分别一一比较。设第一组评酒员对红葡萄打分为 S_1 , 第一组评酒员对白葡萄打分为 S_2 , 第二组评酒员对红葡萄打分为 S_3 , 第二组评酒员对白葡萄打分为 S_4 。

根据标准差计算公式:

$$S = \sqrt{\frac{(w_1 - \bar{w})^2 + (w_2 - \bar{w})^2 + \dots + (w_n - \bar{w})^2}{n}}$$

其中 \bar{w} 为平均值. 可得 $S_1 = 7.343$, $S_2 = 5.110$, $S_3 = 3.978$, $S_4 = 3.173$.

由于 $S_3 < S_1$, $S_4 < S_2$, 第二组红葡萄及白葡萄的标准差值都比第一组的小, 所以第二组评酒员的评价结果更可信.

综合上述两种方法, 结论是一致的, 即第二组评酒员评价结果的可信度更高.

3 结 论

本文基于2012年全国大学生建模竞赛A题数据样本, 运用配对T检验和灰色关联度分析两种方法对评酒员的评价结果是否有显著性差异的问题进行了分析, 结果一致表明: 评酒员的评价结果有显著性差异; 同时根据Kendall协调系数W检验和标准差对两组评酒员的评价结果进行了可信度分析, 得到第二组评酒员评价结果更可信的结论. 两个问题都通过两种方法得到了一致的结论, 对于葡萄酒的感官评价的可信度评判具有重要的

意义.

参考文献:

- [1] 靡川清, 郭安鹤, 王华. 感官分析及仪器分析在葡萄酒香气研究中的应用[J]. 食品科学, 2012, 23: 1-7.
- [2] 李记明, 宋长兵, 贺普超. 葡萄与葡萄芳香物质研究进展[M]. 西北农业大学学报, 1998, 26(5): 105-109.
- [3] 李华, 刘曙东, 王华, 等. 葡萄酒感官评价结果的统计分析方法研究[J]. 中国食品学报, 2006, 6(2): 126-131.
- [4] 刘洋洋, 薛凌云. 葡萄酒评价结果差异性及可信度分析方法研究[J]. 学术论坛, 2012, 21: 180-182, 2012.
- [5] 2012年全国大学生数学建模竞赛A题数据[EB/OL]. http://www.mcm.edu.cn/html_cn/block/61dfec317d7a5b3b218efed81c8af3.html, 2012.9.8.
- [6] 李卫东. 应用多元统计分析[M]. 北京: 北京大学出版社, 2008.
- [7] 袁志发, 宋世德. 多元统计分析(第二版)[M]. 北京: 科学出版社, 2009.
- [8] 苏玉华, 韦师. 葡萄酒感官评价的差异性及可信度研究[J]. 学苑论坛, 2013, 1: 254-255.
- [9] 程琮, 刘一志, 王如德. Kendall协调系数W检验及其SPSS实现[J]. 泰山医学院学报, 2010, 31(7): 487-490.

The Differences and Reliability Analysis Methods of Grape Wine Sensory Evaluation

CAO Zhi-qiang¹, LIU Ya-nan¹, YANG Ting¹, MA Lei¹, QI Ya-e²

(1. School of Mathematics and Statistics, Hexi University, Zhangye 734000, China; 2. College of Chemistry and Chemical Engineering, Hexi University, Zhangye 734000, China)

Abstract: In this paper, the difference of the sensory evaluation between difference tasters was analyzed using the paired T-test and grey relational grade analysis, and the reliability of the evaluation results was analyzed using Kendall coordination coefficient W-test and standard deviation. The test results showed that there was significant different between the two evaluation groups, and the second group evaluation is more reliable.

Key words: paired T-test; grey relational grade analysis; Kendall coordination coefficient; standard deviation; reliability