

Research Article

Analysis on Reliability of Wine Tasters' Evaluation Results Based on the Analysis of Variance

¹Wang Yufei, ¹Qian Qiuye, ²Weixia Luan and ¹Wang Guizhou

¹International Business School,

²School of Translation Studies, Jinan University, 519070, Guangdong, China

Abstract: Based on the data related to the evaluation score of wine taster provided in 2012 CUMCM, this study firstly adopts confidence interval method to eliminate the effect of wine tasters' personal differences. Then, by using analysis of variance, we make a test of significance on evaluation results of wine tasters from Group A and B at the significance level of 0.05. Results show that there is no significant difference in the sensory evaluation results of wine tasters from the two groups. By comparing the variance of comprehensive scores given by wine tasters from the two groups, we confirm the evaluation results of wine tasters from which group are more reliable. Results of the model shows that variances of evaluation results given by wine tasters from Group B are all smaller than that of Group A, which prove that evaluation result of wine tasters from Group B is more reliable.

Keywords: Analysis of variance, test of significance, wine evaluation

INTRODUCTION

During the evolutions process of wine, we usually invite some qualified wine tasters to evaluate the quality of wine. After a thorough taste of a kind of wine, every wine taster will give scores on its classification indicators; sum these scores to figure out the aggregate score in order to evaluate the quality of wine. There exists a direct relationship between the quality of wine grape and the quality of the wine made out of the grape. The physical and chemical indicators for wine and wine grape evaluation can reflect the quality of wine and wine grape at some extent. Most traditional evaluation methods are based on the evaluating scores given by senior wine tasters, thus their judgments are often inevitably influenced by wine tasters' own habit, preference, experience, emotion.

In the sensory evaluation of wine, because of sommelier evaluation scale, assessment and evaluation of the direction of the position and other aspects of differences, resulting in different wine sommelier for the same kind of evaluation vary widely and thus cannot truly reflect differences between wine samples differences. Therefore, the sensory evaluation results of the statistical analysis. Must sommelier correspond processing the raw data to reflect the differences between samples.

Song Yuyang (Year) evaluated several samples by the multi-factor comprehensive evaluation method of fuzzy mathematics and the results showed that the organoleptic taste and comprehensive evaluation has

the same trend and comprehensive can better reflect the taste of the identification of the degree of dispersion.

Li *et al.* (2006) uses the processed relevant data method and the comparative analysis lead to the results: standardization not only not eliminate the heterogeneity between sommelier, but increase the differences between the sommelier, while the confidence interval method to adjust the raw data can be effectively reduce the differences between the sommelier, truly reflect the objective differences between wine samples.

Based on the data and assumption in Problem A of 2012 China Undergraduate Mathematical Contest in Modeling (2012 CUMCM), the data provide us with two groups' evaluation scores on specific factors of wine samples. This study makes a study on how to make an analysis on reliability of wine tasters' sensory evaluation by using analysis of variance in mathematical analysis. As a result, we will figure out which Group's evaluation and judgment on wine samples are preciser and better.

ANALYSIS AND DATA PROCESSING

Analysis: By adopting the data in Problem A of 2012 CUMCM to analyze whether there is significant difference between the evaluation results of wine tasters from two different groups, Group A and B, this study further make a judgment on evaluation results of wine tasters from which group is more reliable.

At first, analyze and process the data to get the comprehensive score of each kind of sample wine given by the wine tasters from the two groups. The data we

get from 2012 CUMCM presented the sensory evaluation scores for 27 kinds of red wine samples and 28 kinds of white wine samples given by the wine tasters from Group A and B (each group with 10 members); the evaluation indicators are mainly divided into four aspects including appearance, aroma, flavor and whole, respectively accounting for 15, 30, 44 and 11%, respectively of the aggregate score. Besides, we make detailed subdivision under each aspect. Since the aggregate score for every kind of wine sample is 100, we can simply sum scores on different evaluation indicators given by one wine taster to figure out the aggregate score of this kind of wine sample in one wine taster's view. Later, sum up aggregate score of this kind of wine sample given by 10 wine tasters from one group and take an average, we can get the final comprehensive score of this wine sample by one wine taster group.

Secondly, during the sensory evaluation process of wine, difference lying in the personal preferences of different wine tasters may result in great difference among the evaluation results given by different wine tasters for the same kind of wine. Thus, this method cannot truly reflect the differences among different kinds of wine. Therefore, before judging whether there is significant difference lying in the evaluation results of wine tasters from Group A and B, we should initially process the comprehensive scores to reduce or even

eliminate the personal difference among wine tasters. Then, we can analyze whether there is significant difference or not. There are many ways to process the comprehensive scores, for example, standardization method and minimization and maximization method. Based on the conclusion in the reference (Li *et al.*, 2006) this study applies confidence interval to process the comprehensive scores.

After that, determine the method to judge whether there is significant difference lying in the evaluation results of wine tasters from Group A and B. Many methods can be used to analyze significance in statistics, such as t-test and chi-square test. But all these methods are based on the premise that the distribution of sample data follows some specific distribution. Considering the distribution of data in our paper, this method can't be adopted. Therefore, this study applies analysis of variance judge whether there is significant difference lying in the evaluation results (Ristic and Bindon, 2010).

Finally, confirm evaluation results of wine tasters in which group are more reliable. Referring to the data in this study, every wine taster here is qualified, which means there should not be too big difference among wine tasters. Therefore, we can determine evaluation results of wine tasters in which group is more reliable by comparing the variance of comprehensive scores given by wine tasters in Group A and B.

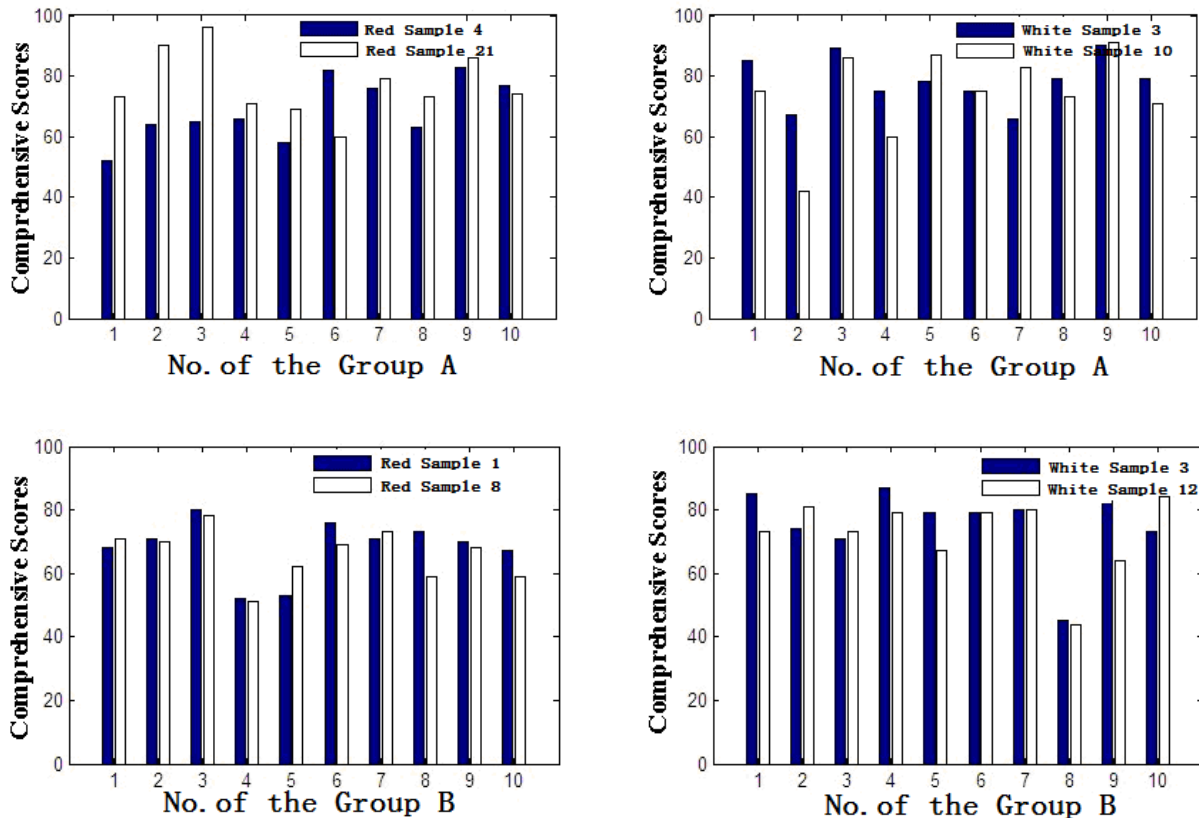


Fig. 1: Differences among the comprehensive scores given by wine tasters

Difference among comprehensive scores of wine given by wine tasters: Based on the data, we can figure out the comprehensive scores of every kind of wine sample given by the wine tasters from Group A and B, which is worked out by summing scores on all sensory indicators together. Figure 1 shows the difference among the comprehensive scores of two random selected kinds of wine samples given by wine tasters in Group A and B.

As Fig. 1 shows, great differences lie between the comprehensive scores of the same kind of wine sample given by the wine tasters from Group A and B. Through analysis, we can know the reason that these differences resulting from are mainly divided into two aspects as follows.

Different wine tasters have different personal preferences and evaluation standards. During the process of sensory evaluation, each wine taster will evaluate the wine according to their own evaluation standards, which leads to the great differences among their evaluation scores, ranging from 41 to 90. Meanwhile, to evaluate the same kind of wine sample, differences among evaluation scores of different wine tasters also occur, due to their own preference for wine.

Differences also occur among different wine samples of the same kind of wine. This kind of difference really exists and it is determined by the nature of the wine sample. This nature is also part of the data we should get in the process of evaluation.

Due to the differences among wine tasters, we should process the comprehensive scores we have got before analyzing the significant difference of the evaluation results to eliminate the influence of the differences among wine tasters.

MODELING, SOLUTION AND ANALYSIS OF RESULTS

Comprehensive scores given by wine tasters from two groups for wine samples:

- **Original comprehensive scores of wine samples ZHAO and DAN (2003):** Sensory evaluation standards of wine tasters are mainly divided into the following four aspects: appearance, aroma, flavor and whole. The total score is 100 and every aspects account for certain marks. Sum the scores on every aspects of sensory evaluation given by wine tasters together, we can get the total score, which can be regarded as comprehensive scores for wine. Therefore, the comprehensive scores of red wine samples given by wine taster can calculated as follows:

$$SR_{ij} = \sum_{k=1}^{10} s_{jk}$$

where,

SR_{ij} = The comprehensive scores for the i^{th} red wine sample given by the j^{th} wine taster from Group A.

s_{jk} = Scores on ten sensory indicators for the wine given by the j^{th} wine taster.

This method is also available for calculating the comprehensive scores of white wine samples.

Comprehensive scores figured out through the method above take all aspects of sensory quality into consideration. There are scientific and reasonable, which can be taken as comprehensive scores of wine. By using Excel, we can work out the comprehensive scores of some red wine samples given by wine tasters from Group A as Table 1.

Seen from Table 1, we can know that there are great differences among the evaluation score for the same wine sample given by different wine tasters, with a score range of 41-90. This is the same as our analysis in the data preprocessing. Thus, we should adopt a corresponding method to reduce or even eliminate the difference in scores resulting from the personal differences among wine tasters.

- **Comprehensive scores of wine samples processed by the confidence interval method:** By analyzing the difference among evaluation scores given by different wine tasters in the data preprocessing, we can conclude that the personal preferences and evaluation standards of different wine tasters can result in great differences among original comprehensive scores, part of which results from differences of wine tasters themselves. In order to reduce or even eliminate these differences, we use confidence interval method to process the original comprehensive scores. Specific steps are listed as follows:

Step 1: Figure out average value of original comprehensive scores, standard deviation and confidence interval for different wine samples given by wine taster in each group.

Step 2: If the original comprehensive scores of wine samples given by wine tasters are in the range of confidence interval, we can directly use the original comprehensive scores without process; If not, we should add its standard deviation when the original comprehensive score is larger than the average value and we should minus its standard deviation when the original comprehensive score is smaller than the average value.

By using Excel, parts of comprehensive scores of part of the red wine samples are presented Table 2. These scores are given by wine tasters from Group A, processing by using confidence interval method.

Table 1: Comprehensive scores of some red wine samples given by wine tasters from Group A

Wine No.	Wine taster 1	Wine taster 2	Wine taster 3	Wine taster 4	Wine taster 5	Wine taster 6	Wine taster 7	Wine taster 8	Wine taster 9	Wine taster 10
1	51	66	49	54	77	61	72	61	74	62
2	71	81	86	74	91	80	83	79	85	73
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
25	60	78	81	62	70	67	64	62	81	67
26	73	80	71	61	78	71	72	76	79	77
27	70	77	63	64	80	76	73	67	85	75

Table 2: Part of the comprehensive scores for red wine samples given by wine tasters from Group A, after processing with the confidence interval method

WineNo.	Wine taste 1	Wine taster 2	Wine taster 3	Wine taster4	Wine taster 5	Wine taster 6	Wine taster 7	Wine taster 8	Wine taster 9	Wine taster 10
1	68	71	80	52	53	76	71	73	70	67
2	75	76	76	71	68	74	83	73	73	71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
25	68	68	84	62	60	66	69	73	66	66
26	68	67	83	64	73	74	77	78	63	73
27	71	64	72	71	69	71	82	73	73	69

Table 3: Part of the final average comprehensive scores of red and white wine

Wine samples No.	Red wine final average comprehensive scores		White wine final average comprehensive scores	
	Group A	Group B	Group A	Group B
1	62.76	67.98	79.89	77.60
2	81.40	73.58	74.85	75.80
⋮	⋮	⋮	⋮	⋮
26	74.50	72.22	79.97	75.82
27	73.36	71.20	62.80	76.20
28	-	-	82.05	79.30

Seen from Table 2, after processing with the confidence interval method, differences of scores given by wine tasters are greatly reduced, due to the reduction or elimination of the effect of wine tasters' personal difference on evaluation. Hence, it is more scientific to analyze the scores after processing.

- **Average comprehensive scores processed by the confidence interval method:** Take the average value of the comprehensive scores processed by the confidence interval method as final comprehensive scores wine samples given by one wine taster. Part of the final average comprehensive scores of red and white wine given by wine tasters from Group A and B are listed in Table 3. Average comprehensive scores after processing successfully eliminate the effect of wine tasters' personal differences on evaluation. Based on this, we make a test of significance.

Concluded from Table 3, we find that certain difference exists among scores of the same wine sample given by wine taster from two different groups. But, we cannot judge whether there is significant difference in the evaluation results given by wine tasters only by observing data.

Test of significance in the evaluation results of wine tasters from Group A and B: By using Excel, we figure out the final average comprehensive scores given

by two groups. Set significance level to be 0.05 and make an analysis of variance to get the statistical result as we can see from Table 4, the significance probability of red grape and white grape are, respectively 0.0802 and 0.0525, which are greater than 0.05. Thus, we can conclude that there is no significant difference in the results of sensory evaluation given by wine tasters from the two groups. It also means that the overall evaluation results of wine tasters from the two groups are consistent, which conforms to the assumption that all wine tasters are qualified and able to give objective evaluation about the sensory quality of wine.

Comparison of credibility's of wine samples' evaluation result given by wine tasters from two different groups YUAN (1990): For the comparison of credibility's, since all wine tasters are qualified, we can simply consider that the group whose wine tasters have the smaller variance of comprehensive scores processed by confidence interval is better. That is to say that the evaluation results of wine tasters within a group is more consistent, which also means the evaluation results are more reliable. Thus, we can judge the evaluation results given by wine tasters from which group is more reliable by comparing the variance of evaluation results given by these two groups.

Based on the data we get after processing by confidence interval, we can figure out the variance of final comprehensive scores for red and white wine given by wine tasters from Group A and B, which is shown in Table 5.

Table 4: Analysis result of the variance of final average comprehensive scores given by wine tasters from Group A and B

Red grape						
	SS	df	MS	F	p-value	F crit
Among groups	115.6637	1	115.6637	3.183624	0.080214	4.026631
Within group	1889.203	52	36.33082			
White grape						
Among groups	77.74473	1	77.74473	3.930545	0.052514	4.019541
Within group	1068.100	54	19.77963			

Table 5: Variance of final comprehensive scores for wine given by wine tasters from Group A and B

Wine samples No.	Final comprehensive scores variance of red wine		Final comprehensive scores variance of white wine	
	Group A	Group B	Group A	Group B
1	9.14	8.58	9.11	4.83
2	5.98	3.82	13.45	6.65
3	6.42	5.26	18.13	11.32
4	9.86	6.10	6.34	5.50
5	7.47	3.51	10.67	4.86
6	7.33	4.36	12.10	4.52
7	9.66	7.51	5.94	6.16
8	6.29	7.66	12.85	5.29
9	5.45	4.81	9.14	9.78
10	5.23	5.71	13.84	7.96
11	7.98	5.85	12.63	8.89
12	8.47	4.75	10.21	11.23
13	6.36	3.71	12.40	6.49
14	5.69	4.57	10.14	3.78
15	8.78	6.10	10.88	6.97
16	4.04	3.91	12.66	8.60
17	8.90	2.87	11.39	5.88
18	6.52	6.73	11.87	5.22
19	6.53	7.05	6.46	4.84
20	4.84	5.93	7.61	6.71
21	10.22	5.65	12.47	7.61
22	6.75	4.67	11.17	6.95
23	5.41	4.72	6.27	3.23
24	8.21	3.11	10.00	5.89
25	7.63	6.27	5.52	9.79
26	5.31	6.12	8.10	9.62
27	6.69	4.30	11.40	5.66
28	-	-	8.51	4.78
Sum	191.16	143.61	282.74	184.24

Seen from Table 5, we can conclude that the variance of final comprehensive scores for wine given by wine tasters from Group B is smaller than that of Group A. Generally, the evaluation results of wine tasters from Group B are more similar to each other's. Then, we can believe that final comprehensive scores for wine given by wine tasters from Group B are more reliable than that of Group A. At the same time, by further comparison, we find that for every kind of wine sample, variances of evaluation results given by wine tasters from Group B are all smaller than that of Group A, which further prove that evaluation result of wine tasters from Group B is more consistent and more reliable.

CONCLUSION

By adopting various methods such as confidence interval method and analysis of variance, this study makes a mathematical analysis on evaluation results of wine tasters from Group A and B, avoiding the subjective effect wine tasters' personal preference on

the evaluation results to some extent. Then we come to the conclusion that evaluation results of wine tasters from Group B are more consistent and more reliable. This study not only simplifies the problem but also successfully solve it. Our model is visual, which can be further promoted and improved.

Preprocess the data by statistics, analysis, and classification and so on, this study applies data processing methods such as confidence interval method to get more reliable results with certain reference value. In the meantime, further analyze and improve the model by increasing the number of wine tasters in one group and so on. Then we can get the results which are more scientific.

REFERENCES

- Li, H., S. Liu, H. Wang and Y. Zhang, 2006. Studies on the statistical analyses methods for sensory evaluation results of wine. *J. Chinese Inst. Food Sci.*, 6(2): 126-131.

- Ristic, R. and K. Bindon, 2010. Flavonoids and C13-norisoprenoids in *Vitis vinifera* L. cv. Shiraz: Relationships between grape and wine composition, wine colour and wine sensory properties. *Aust. J. Grape Wine Res.*, 16: 369-388.
- Song, Y. *et al.*, 2002. Application of the fuzzy synthetic judgement in the taste of wine. *Sino-Overseas Grapevine Wine*, 2: 35-34.
- Yuan, Y., 1990. *Probability Theory and Mathematical Statistics*. China Renmin University Press, Beijing, pp: 167-184.
- Zhao, J. and Q. Dan, 2003. *Mathematical Modeling and Mathematical Experiment*. Higher Education Press, Beijing, pp: 260-275.