Establishment of the Wine Quality Evaluation Model

Shaodong Chen[*] and Song Liang

School of Mathematics and Physics, Nanyang Institute of Technology Henan Nanyang 473500, China

Abstract: In accordance with data provided from Mathematics Modeling Competition for National University students in 2012, it can be found whether grapes are high-quality or not as well as the direct relationship between the grape and wine quality by means of SPSS and other statistical softwares, through multivariate statistical analysis and data analysis, and by verifying the credibility what comments made on wine quality by members of liquor, additionally, quality classification on grapes is made through factor analysis. The physical and chemical indicators detected from wine and grapes reflect their quality to a certain extent. Through analysis of relationship between physical and chemical indicators concerning grapes and wine, a mathematical model used for evaluating wine quality with the help of physical and chemical indicators concerning grapes and wine is established, providing references for accurately evaluating wine quality.

Keywords: 1, T –tes 2, canonical correlation analysis 4, factor analysis 3, regression analysis.

1. INTRODUCTION

Wine quality evaluation can be determined by retaining a batch of qualified members of liquor to comment. Analyzing the relationship of physicochemical indexes between wine grapes and wine, the mathematical model can be established. Which of these physicochemical indexes have an influence on the wine quality is sought, in order to utilize these physicochemical indexes to evaluate the wine quality.

2. MATERIALS AND METHODS

2.1. Reliability Comparison of Members of Liquor Evaluation Results Between Two Groups

On the assumption that qualification level of members of liquor in the same group is equivalent, the arithmetic mean can be worked out using each group's wine data in the first accessory, which is supplied by the National University Colleges Mathematics Modeling Competition in 2012 [5]. There is a significant test of the members of liquor evaluation results between two groups. We can see from P-P diagram that data points are basically distributed on or near the diagonal, which is a normal distribution. Utilizing Levene's method to work out the test results of homogeneity of variance about the overall arithmetic mean between two groups, we can see that F=3.61 and P=0.55 (Less than 0.1). Therefore, under the significance level α of=0.1, the population variance of arithmetic mean is unequal. So, we choose the t test results on the condition of heteroscedasticity. Sig (bilateral) is 0.12 (more than 0.1), which we cannot deny the null hypothesis that the population mean is equal. What is more, the mean difference is 2.5407. 90% of the confidence

intervals that are (-0.1654, 5.2368), including 0, which indicates that there are no significant differences in the population mean. On the condition of heteroscedasticity, the difference of marking standard of the second group's members of liquor is small and stable. Therefore, the evaluation results of the second group's members of liquor are more reliable [1].

2.2. Decision on the Classification of Grapes

We should carry out the factor analysis based on the physicochemical index of grapes, by determining the common factor, utilizing factor for getting the score function, and working out to integrate the score of the grape samples. We should analyze the obvious change trend of the score and put forward several classification schemes, referencing the quality classification of grapes. And then, we should utilize the scores of members of liquor coming from different levels, by determining the most premium classification scheme to estimate that different grape levels have different score ranges of wine using the parameter intervals. The steps taken are as follows:

Having the standardization processing on the data of the second accessory, the formula is:

$$x_{ij} = \frac{x_{ij} \quad \overline{x_j}}{s_j} \tag{1}$$

(x_{ij} indicates that the data is located in the j column of the i sample. $\overline{x_j}$ indicates that it is the data mean of the j column. s_j indicates that it is the data standardization difference of the j column). We thus get the standardized data.

It should be noted that the data in the second accessory will be operated at the first grade indexes. We should make the average as the index data in many measurements. If the data of the grapes' pericarp color is a plusminus number, in case that pericarp colors of L*, a*, b* will be positive and

*Address correspondence to this author at the School of Mathematics and Physics, Nanyang Institute of Technology, Nanyang, 473500, China; Tel: +8613137792688; E-mail: chenshaodong2006@163.com

negative offset, we should work out the arithmetic square root of the sum of squares of these three color indexes.

There are twenty-eight physicochemical indexes of grapes in the second accessory, which is supplied by the National University Colleges Mathematics Modeling Competition in 2012. After factor analysis, these indices can be integrated into eight common factors. According to the variance contribution and accumulative contribution rate, we can make the common factors by combining those whose characteristic roots of the first eight are more than one. The interpretation of the total variance is about up to 85%.

The comprehensive score of each grape sample is taken on the basis of eight common factor data and factor score function [1]. The expression of the factor score function is:

$$F_1 = a_{11}F_{11} + a_{12}F_{12} + ... + a_{18}F_{18}$$

$$a_{1k} = \frac{\lambda_{1k}}{\lambda_{11} + \lambda_{12} + ... + \lambda_{18}} \text{ (k=1,2,....,8)}$$

$$\lambda_{11} = 7.142$$

$$\lambda_{12} = 5.071 \quad \lambda_{13} = 3.318 \quad \lambda_{14} = 2.950$$

$$\lambda_{15} = 2.063 \quad \lambda_{16} = 1.578 \quad \lambda_{17} = 1.341 \quad \lambda_{18} = 1.118$$

The scores of $F_{11}, F_{12} \cdots\cdots F_{18}$ of physicochemical index factors have been taken in the process of factor analysis such as, F_1, which is the comprehensive score of each sample [3].

The classification schemes of grapes can be determined based on wine quality evaluation and grape samples' comprehensive scores. The grapes will initially be divided into three levels, *i.e.* Level 1, Level 2 and Level 3, comprehensively referencing the classification standards of grape quality. There are several specific schemes. There are two main kinds as are shown in the following Table **1**:

Table 1. Classification scheme.

The first kind	Level1	3, 17, 2, 12, 23, 6, 9, 5、8, 1, 21
	Level2	22, 24, 19, 14, 18, 20, 15, 13, 7, 26, 27, 16, 4, 11
	Level3	25, 10
The second kind	Level1	3, 17, 2, 12, 23, 6, 9, 5, 8, 1, 21, 22
	Level2	24, 19, 14, 18, 20, 15, 13, 7, 26, 27, 16, 4, 11
	Level3	25, 10

The scheme is examined by utilizing evaluation scores that are provided by the masters of liquor commenting on the wine quality. It can be identified that the first kind is the premium scheme. The classification range can be set by using the samples of different levels to estimate the parameter intervals in general. The formula for estimating the parameter intervals is:

$$\bar{\mu} \pm t_{\frac{\alpha}{2}}(n-1) \times \frac{S}{\sqrt{n}} \tag{2}$$

Among them, $\bar{\mu}$ is the sample quality evaluation average of grapes with different levels. n is the sample number of different levels. Taking α as 0.05, S is the sample standard deviation [4].

The score ranges of quality evaluation are as follows after grapes of different levels are brewed into wine.

The first level: (68.8951, 74.9449);

The second level: (67.8352, 71.8581);

The third level: (65.8046, 71.1954).

3. CANONICAL CORRELATION ANALYSIS ON ALL THE PHYSICOCHEMICAL INDEXES BETWEEN GRAPES AND WINE

Taking eight common factors combining all the physicochemical indexes of grapes and seven physicochemical indexes of wine as two related variables, we can carry out the canonical correlation analysis. According to the SPSS canonical correlation analysis measures and steps, the results of the main canonical correlation analysis show that there are seven canonical correlation coefficients in the two variables. The canonical correlation coefficients will be operated by chi-square test. Assuming that the corresponding canonical correlation coefficients are 0, the significant probabilities (Sig.) of the first canonical correlation coefficient and the second canonical correlation coefficient are 0 and 0.004 respectively. On the condition that α =0.05, the assumption that canonical correlation coefficient is 0, is untenable. It also shows that the relevance of the first two pairs of canonical variables is obvious. And the relevance of the latter five pairs of canonical variables is not obvious. Therefore, the study on the relevance of the two related variables can be converted into the study on the relationship between the first pair of canonical related variables and the relationship between the second pair of canonical related variables.

The coefficient of the two pairs of canonical variables, can be used to conduct a survey on the correlation between the physicochemical indexes of grapes and physicochemical indexes of wine. The formula of the first canonical variables derived from common factors of grapes is:

$$U_{11} = -0.418F_{11} + 0.261F_{12} + 0.027F_{13} + 0.750F_{14} - 0.229F_{15}$$
$$+ 0.359F_{16} - 0.112F_{17} + 0.002F_{18}$$

$$V_{11} = -1.144Y_{11} + 1.043Y_{12} - 0.376Y_{13} - 0.650Y_{14} + 0.163Y_{15}$$
$$+ 0.465Y_{16} - 0.086Y_{17}$$

The coefficient of the physicochemical indexes of grapes is larger for the first pair of canonical variables, which indicates that they have better interpretation degree.

The second pair of canonical variables, is as follows;

$$U_{12} = 0.861F_{11} + 0.126F_{12} + 0.231F_{13} + 0.321F_{14} - 0.157F_{15}$$
$$+ 0.164F_{16} + 0.134F_{17} + 0.125F_{18}$$

$$V_{12} = -0.212Y_{11} + 1.315Y_{12} - 0.423Y_{13} + 0.227Y_{14} + 0.214Y_{15}$$

$$- 0.072Y_{16} - 0.028Y_{17}$$

Among them, F_{1i} is the common factor obtained by factor analysis on the physicochemical properties of grapes. $i \in (1,8)$ Y_{1i} is the standardized indicator obtained from the physicochemical properties of wine.

Thus, most coefficients are more uniform in the physicochemical indexes of grapes, except that F_{15} is negative, Y_{12} is even more bigger and positive in the physicochemical properties of wine, and some others are more uniform. It can be interpreted that the bigger the test indexes are, the better the physicochemical properties of wine [2].

3. RESULTS AND DISCUSSION

The models are established for evaluating wine quality by using the physicochemical properties of grapes and wine.

According to the chemical property of aromatic substances, they can be divided into six main categories: alcohols, esters, acids, carbonyl compounds, phenols, terpenes and other categories. Firstly, we should make up the missing data in the list of aromatic substances. All the data for each sample should be summed, which will be set as the value of each category and be used as standardized parameters.

From calculation, it has been found that the observed value of F statistics of regression models is 1.704 and Sig is 0.165. On this condition, the significance level is 0.1 and on the basis of above analyses, we conclude that there is a linear relationship among Y_1 (expert grading), X_{11} (esters), X_{12} (alcohols), X_{13} (carbonyl compounds), X_{14} (acids), X_{15} (phenols and terpenes), X_{16} (other categories), X_{17} (U11), X_{18} (U12), X_{19} (V11), and X_{110} (V12).

According to the Model1, the multivariate linear regression equation is

$$Y_1 = 0.392X_{11} - 0.294X_{12} - 0.134X_{13} - 0.092X_{14}$$

$$- 0.148X_{15} + 0.103X_{16} - 0.139X_{17} + 0.635X_{18}$$

$$+ 0.155X_{19} - 0.100X_{110}$$

The model's multiple correlation coefficient is 0.718 and its coefficient of determination is 0.516, which is even more higher. It indicates a significant relationship among wine quality, grapes and the regression model of the physicochemical properties of wine.
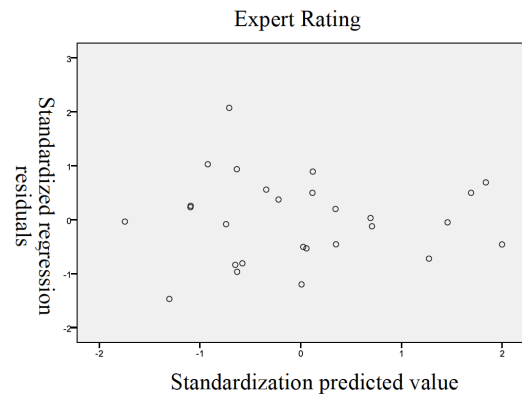

Fig. (1). Scatterplot.

It shows from Fig. (**1**) that no matter how the standard predictive value changes, the fluctuation range of standard residuals basically retains stable, which indicates that the residual variance is homogeneous and the regression equation of both is in line with normality hypothesis. Therefore it is meaningful to examine the regression model.

Firstly, this paper establishes the authority that masters of liquor in the second group have evaluated the wine. Secondly, the canonical correlation relationship between grapes and wine has been confirmed by the canonical correlation analyses. Finally, physicochemical indexes of grapes and wine were utilized to evaluate the wine quality.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Y. Li, J. M. Li, and Z. J. Jiang, "Application of Statistical Analysis in the Evaluation of Grape Wine Quality", *Liquor-Making Science & Technology*, vol. 4, pp. 79-82, 2009.
[2] X. Xiao, L. Li and B. S. Xu, "Quality Evaluation of Grape Wine Based on Discrete Hopfield NeuralNetwork", *Journal of Shanghai University of Engineering Science,* vol. 27, no. 3, pp. 266-269, 2013.
[3] P. Cortez, A. Cerdeira, and F. Almeida, "Modeling wine preferences by data mining from physicochemical properties", *Decision Support Systems*, vol. 47, no. 4, pp. 547-553, 2009.
[4] W. H. Yu, SPSS and Statistical Analysis, Beijing: Publishing House of Electronics Industry, 2007.
[5] The National University Colleges Mathematics Modeling. http://www.mcm.edu.cn/problem/2012/cumcm2012problems. ra-rompetition, DB/O