ISSN : 0974 - 7435 Volume 10 Issue 23

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(23), 2014 [14052-14056]

Study on the wine quality evaluation model

Chen Shaodong*[1], Song Liang[2]
[1,2]School of Mathematics and Physics, Nanyang Institute of Technology Henan
Nanyang, (CHINA)
E-mail: Chenshaodong2006@163.com, Songliang0903@163.com

ABSTRACT

According to the data supplied by the National University Colleges Mathematics Modeling Competition in 2012, we can find that whether the grapes are good or not, and the direct relationship between the grapes and the wine quality by the help of the SPSS and other statistical softwares, by comprehensively utilizing the method of multivariate statistical analysis and then carrying out data analysis, by distinguishing the credibility that members of liquor comment on the wine quality, and by utilizing the factor analysis to make quality classification on the grapes. The physical and chemical indicators detected by wine and grapes reflect the quality of wine and grapes to a certain extent. Analyzing the relationship among the physical and chemical indicators of the grapes and wine, we establish a mathematical model evaluating the wine quality by utilizing the physical and chemical indicators of the grapes and wine, which provides some references to accurately evaluate the wine quality in practical problems.

KEYWORDS

T –tes; Factor analysis; Canonical correlation analysis; Regression analysis.

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 these physicochemical indexes have an influence on the wine quality, in order to utilizing these physicochemical indexes to evaluate the wine quality.

MATERIALS AND METHODS

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 supplied by the National University Colleges Mathematics Modeling Competition in 2012.

(http://www.mcm.edu.cn/problem/2012/cumcm2012problems.rar)There is a significant test on 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 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 of α =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 are (-0.1654, 5.2368), including 0, which indicates that there are no significant differences on 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.

Decision on the classification of grapes

We should carry out the factor analysis on the physicochemical index of grapes, determine the common factor, utilize factor getting the score function, work out the integrate 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, determine the most premium classification scheme and 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 standardization processing formula is：

$$x_{ij}^* = \frac{x_{ij} - \overline{x_j}}{s_j}$$

(x_{ij} indicates that the date 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 can 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 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 supplied by the National University Colleges Mathematics Modeling Competition in 2012.

(http://www.mcm.edu.cn/problem/2012/cumcm2012problems.rar)After factor analysis, they can be integrated into eight common factors. According to the variance contribution and accumulative contribution rate, we can make the factors, whose characteristic roots of the first eight are more than one, as the common factors. The interpretation of the total variance is about up to 85％.

We can get the comprehensive scores of every grape sample on the basis of eight common factor data and factor score function (Li Yun, Li Jiming, and Jiang Zhongjun,2009). The expression of the factor score function is:

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

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

$\lambda_{11} = 7.142$,

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

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

The scores of $F_{11}, F_{12} \cdots\cdots F_{18}$ of physicochemical index factors have been given out in the process of factor analysis. We can get F_1, which is the comprehensive score of every sample.

The classification schemes of grapes can be determined by the sort of wine quality evaluation and the sort of grape samples' comprehensive scores. The grapes will initially divide into three levels, which are 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 follows:

TABLE 2 : 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

We can examine the scheme by utilizing evaluation scores that are given by the masters of liquor commenting on the wine quality. We can get that the first kind is the premium scheme. The classification range can be gotten by using the samples of different levels to estimated the parameter intervals in general. The formula of estimating the parameter intervals is:

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

Among of 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.

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）

Canonical correlation analysis on all the physicochemical indexes between grapes and wine

Taking eight common factors of 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 of that α =0.05, the assumption of 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 transferred 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.

Since the coefficient of the two pairs of canonical variables, we can have a survey on the correlation between the physicochemical indexes of grapes and physicochemical indexes of wine. The formula of the first canonical variables 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 in the first pair of canonical variables, which indicates that they have better interpretation degree.

As to the second pair of canonical variables,

$$U_{12} = 0.861F_{11} + 0.126F_{12} + 0.231F_{13} + 0.321F_{14} - 0.157F_{15}$$

$$+ 0.164F_{16} + 0.134F_{17} + 0.125F_{18}$$

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

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

Among of 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. $i \in (1,7)$

We can see that most of coefficients are more uniform in the physicochemical indexes of grapes, except F_{15} is negative. Y_{12} is more bigger and positive in the physicochemical properties of wine, others are also more uniform. It can be interpreted that the bigger the test indexes are, the better the physicochemical properties of wine (Xiao Xiang, Li Lu and Xu Bosheng,2013).

RESULTS AND DISCUSSION

The models establishment of evaluating wine quality by using the physicochemical properties of grapes and wine

According to the chemical property of aromatic substances, it 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 of each sample should be summed, which can be considered as the value of each category and be standardized.

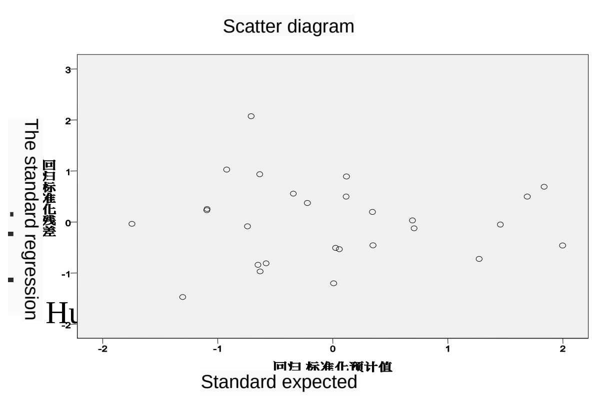
From calculation, we can see that the observed value of F statistics of regression models is 1.704 and Sig is 0.165. On the condition of that the significance level is 0.1 and on the basis of above analysises, we can conclude that there is linear relationship among Y_1(expert grading), X_{11}(esters), 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) (Cortez P, Cerdeira A,Almeida F,et al.2009).

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 more higher. It indicates that there is significance among wine quality, grapes and the regression model of the physicochemical properties of wine.



In accordance with the residual plot, we can know no matter how the standard predictive value of Y_1 changes, the fluctuation range of standard residuals retains basically stable, which indicates that the residual variance is homogeneous and the regression equation of both is in line with normality hypothesis. It is meaningful to examine the regression model.

After all, firstly, this paper establishes the authority that masters of liquor in the second group have evaluated on the wine. Secondly, the canonical correlation relationship between grapes and wine has been confirmed by the canonical correlation analysises. At last, utilizing the physicochemical indexes of grapes and wine evaluates the wine quality.

REFERENCES

[1] Li Yun, Li Jiming, Jiang Zhongjun; Application of Statistical Analysis in the Evaluation of Grape Wine Quality. J. Liquor-Making Science & Technology, **(4)**, 79-82 **(2009)**.
[2] Xiao Xiang, Li Lu, Xu Bosheng; Quality evaluation of grape wine based on discrete hopfield neural network, J.Journal of Shanghai University of Engineering Science, **27(3)**, 266-269 **(2013)**.
[3] P.Cortez, A.Cerdeira, F.Almeida et al.; Modeling wine preferences by data mining from physicochemical properties, J.Decision Support Systems, **47(4)**, 547-553 **(2009)**.
[4] Yu Wenhua; SPSS and statistical analysis, M.Publishing House of Electronics Industry, Beijing, **(2007)**.
[5] The National University Colleges Mathematics Modeling Competition, DB/OL, http://www.mcm.edu.cn/problem/2012/cumcm2012problems.rar