

Evaluation and Analysis Model of Wine Quality Based on Mathematical Model

Yunhui Zeng¹, Yingxia Liu¹, Lubin Wu¹, Hanjiang Dong¹, Yuanbiao Zhang¹, Hongfei Guo¹, Zisheng Guo¹, Shuyang Wang¹, Yao Lan¹

¹College of Intelligent Science and Engineering, Jinan University, Zhuhai, China

Correspondence: Zeng Yunhui, College of Intelligent Science and Engineering, Jinan University, No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China.

Received: September 17, 2018 Accepted: October 24, 2018 Online Published: November 1, 2018

doi:10.11114/set.v6i1.3626

URL: <https://doi.org/10.11114/set.v6i1.3626>

Abstract

This paper takes wine quality evaluation as the research object, establishes the analysis and evaluation model of wine quality, and explores the influence of physical with chemical indicators of wine grapes and wine on the wine quality. Firstly, the Mann-Whitney U test is used to analyze the wine evaluation results of the two wine tasters, and it is found that the significant difference between the two is small. Then this paper uses the Cronbach Alpha coefficient method to analyze the credibility of the two groups of data. It is found that the credibility of the first group of wine scores is significantly greater than that of the second group and the white wine scores are more reliable than the red wine. Therefore, the first set of data and white wine can be applied for follow-up studies. Next, the principal component analysis is used to extract the main indicators and calculate the factor coefficients as the Ward method in cluster analysis is used to classify the wine into four grades according to the quality score of the wine. Then, based on the extracted principal components that is physical with chemical indicators, this paper does the multiple linear regression analysis of wine quality, and takes the influence of aromatic substances on the aroma of wine in physical with chemical indicators as an example. Regression analysis shows that there is a positive correlation linear relationship between the scores of the aroma of wine and C_2H_6O , $C_6H_{12}O_2$, C_3H_8O , $C_{11}H_{24}$, $C_7H_{12}O_2$, $C_5H_{10}O_2$ and $C_{10}H_{16}$. It can be judged that the aromatic substances in the wine such as C_2H_6O have a regular influence on the odor of the wine, and it is inferred that other physical and chemical properties have a similar regular relationship with the wine quality. This provides an effective reference for the analysis and evaluation of wine quality by using physical with chemical indicators such as aromatic substances in wine in the future.

Keywords: wine quality, Mann-Whitney U test, Cronbach Alpha coefficient, Ward cluster method, multiple linear regression analysis

1. Introduction

In recent years, wine has been favored by more and more consumers with its unique taste and high nutritional value and in the meantime the evaluation of wine quality has also increasingly become the focus of attention. The evaluation of wine quality is usually to hire some professional wine assessors, who will review and score, and finally determine the overall quality of the wine. However, determining the quality of wine in this way is susceptible to a number of subjective factors, and the evaluation of the quality of the same batch of wine is often divergent. In fact, there are two main directions for the indicator analysis of wine quality evaluation. One is the statistical analysis idea, which is to establish the quantitative index of wine evaluation based on experimental data and analyze the credibility of the index. For example, obtaining the wine quality evaluation model by establishing BP neural network trained by the error reverse propagation algorithm (Wang & Guan, 2016). And establishing the quality discriminant function by summarizing the well-known wine quality evaluation methods, and obtaining the criteria for wine classification (Yakuba, 2016). The other is the physical and chemical analysis idea, which is based on theoretical research, analyzing the influence of the chemical composition of the wine and its physical environment on its evaluation, and then verifies the guess according to the experimental data. For example, using solid-phase Microextraction-gas chromatography-mass spectrometry establishes an identification model to study 74 volatile compounds under the physical conditions of short-term high temperature and continuous vibration, and conducting experiments on 23 volatile compounds with the highest odor activity values is used to obtain data and verify the accuracy of the model (Zhao,

Wang, & Li, 2018). In addition, directly explore the influence of headspace volume, ascorbic acid and sulfur dioxide on the oxidation state of Riesling wines and sensory characteristics (Morozova, Schmidt, & Schwack, 2015).

Under the statistical analysis, the outstanding advantage of BP neural network is that it has strong nonlinear mapping ability and flexible network structure (Wang et al., 2016). However, this method also has the following defects: Firstly, it's speed of learning is slow. Even if it is a simple question, it usually takes hundreds or even thousands of times to learn. Secondly, easy to fall into local minimums. Thirdly, there is no corresponding theoretical guidance on the choice of the number of network layers and the number of neurons. Fourthly, Capabilities of network promotion is limited. The method of establishing the quality discriminant function fully summarizes the well-known evaluation methods, and thus has certain universality, but the evaluation method of wine will seem to be insufficiently innovative. The solid-phase Microextraction-gas chromatography-mass spectrometry has the advantages of academic strength but it requires knowledge of extraction and chromatography with the theoretical analysis, so it does not have the characteristics of simplicity. And the theoretical analysis of the influence of headspace volume, ascorbic acid and sulfur dioxide on the oxidation state of wine and sensory characteristics has the same advantages and disadvantages, that is, it needs to have a full understanding of the chemical properties of ascorbic acid and sulfur dioxide.

In summary, according to the data and questions of topic A of the National College Mathematical Modeling Contest (2012), this paper takes wine quality evaluation as the research object, adopts statistical ideas, uses correlation analysis, establishes wine quality analysis and evaluation model, explores the influence of physical with chemical indicators of wine grapes and wine on wine quality, hoping to obtain the relationship between physical with chemical indicators and wine quality, providing reference for improving the quality of wine.

2. Method

2.1 Evaluation Results of the Two Wine Drinkers

As described in the introduction, after the data is collected, to determine whether the data satisfies the general law, it can be tested whether the sample data is from the same normal distribution population, that is, it is tested for significance. The following is an example of two groups' data of wine tasters in a sample of white wine, using a normal distribution based on the Lilliefors test for significant analysis (Lilliefors, 1967).

2.1.1 Normal Distribution Test Establishes Lilliefors Method

The Kolmogorov-Smirnov test is usually used to test the normal distribution of samples. This test is used to test a continuous distribution function $F_0(x): H_0: F(x) = F_0(x)$ which is fully known. ($F_0(x)$ is a completely known continuous distribution function fully known). When H_0 is established and n is large, the difference between $F_n(x)$ and $F(x)$ should not be too large. Therefore, the statistic B is used as the test statistic, and the limit distributions of the exact distributions A and C are derived. The specific method of determination and the determination of the rejection region are given:

Step1: The sample observations X_1, X_2, \dots, X_n are arranged in $X(1) \leq X(2) \leq \dots \leq X(n)$ in not descending order;

Step2: Calculate the value of d_i, D_n :

$$d_i = \max \left\{ \left| F_0(x_i) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F_0(x_i) \right| \right\}, i = 1, 2, \dots, n \quad (1)$$

$$D_n = \max \{ d_1, d_2, \dots, d_n \} \quad (2)$$

Step3: For a α given level of significance, when $n \leq 100$, the critical value $D_{n, \alpha}$ is checked in the critical value table $D_{n, \alpha}$ of the Kolmogorov-Smirnov test;

Step4: If $D_{n, \alpha} \leq D_n$, reject the null hypothesis, that is, the sample is not taken from the population with the distribution $F_0(x)$; Otherwise accept H_0 , that is, the sample is taken from the distribution of the total $F_0(x)$.

The Kolmogorov-Smirnov test compares the sample data with the expected theoretical distribution and derives the test results based on whether the two are consistent. This method is a nonparametric test that only applies to standard normal distribution problems and does not apply to non-standard normal distributions. At the same time, the Smirnov test assumes that the parameters of the population are the mean μ , and the variance σ^2 is known, but it is difficult to know these parameters in practice, Therefore, based on the Kolmogorov-Smirnov test, the modified Lilliefors test not only applies to all normal distribution test problems, but also replaces the overall parameters according to the sample mean and variance. It has stronger universality. The basic operations are as follows:

For the unknown parameters μ, σ^2 , if replaced by their unbiased estimator:

$$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})^2 \quad (3)$$

the hypothesis to be tested is actually:

$$H_0 : F(x) = F_0(x; \bar{X}, \hat{\sigma}^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \int_{-\infty}^x \exp\left\{-\frac{(t-\bar{X})^2}{2\hat{\sigma}^2}\right\} dt \tag{4}$$

For a given level α of significance, the test rule is:

If $D_{n,\alpha} \leq \hat{D}_n$, then reject the null hypothesis H_0 , otherwise accept H_0 . To find $\hat{D}_{n,\alpha}$, we need to know the distribution function of \hat{D}_n . The calculation method of \hat{D}_n value is the same as the calculation method of D_n value.

2.1.2 Solution Result of Lilliefors Test

The results obtained by using the Lilliefors test of the tasting score data of the first group of white wines are shown in Table 1 below. Due to space limitations, only the top five data are shown here.

Table 1. Lilliefors test of the first set of white wine tasting score data

| Parameter | | Clarity | tone | Aroma purity | Aroma concentration | Aroma quality |
|---------------------------------------|----------|---------|-------|--------------|---------------------|---------------|
| Average | | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| standard deviation | | 3.60 | 6.40 | 3.10 | 5.60 | 10.20 |
| Most extreme difference | absolute | 0.52 | 0.84 | 0.74 | 1.51 | 1.48 |
| | positive | 0.38 | 0.48 | 0.25 | 0.41 | 0.25 |
| | negative | 0.28 | 0.48 | 0.25 | 0.20 | 0.25 |
| Test statistics | | -0.38 | -0.38 | -0.32 | -0.25 | -0.41 |
| Progressive significance (two-tailed) | | 0.38 | 0.38 | 0.48 | 0.25 | 0.41 |

It can be seen from the above significant results that most of the significant indexes are less than 0.05, thus it can be judged that the normal distribution of the data of the group is not obvious, so the significant difference analysis cannot be performed by the T test. Therefore, another method that does not require data to follow a normal distribution will be tested next.

2.2 Significance Test Based on Mann-Whitney U

Considering that the Lilliefors test requires the sample data to satisfy the normal distribution, but the collected data does not meet this condition, therefore, the sample will be analyzed by Mann-Whitney U test.

The Mann-Whitney U test is also a nonparametric test method (Nachar, 2008). Unlike the Lilliefors method, it does not require the sample data to follow a normal distribution and its validity is also comparable to the T test of the normal distribution. The main steps of the Mann-Whitney U test are as follows:

Before conducting the test, first make the following assumptions for the two groups of white wine sample data sets:

- (1) H_0 : There is no significant difference in the evaluation results in the two sample (sample 1 and 2) data sets;
- (2) H_1 : There is a significant difference in the evaluation results in the two sample (sample 1 and 2) data sets.

Next, taking sample 1 as an example, the score data of the two wine tasters are separately mixed and the data is sorted in ascending order according to the size of the data. Then the rank is arranged, and the minimum rank is recorded as 1, in the meantime the second rank is recorded as 2. If the evaluation results are the same size, the average of the data bit sequence is taken. And the sum of the levels of the two samples is obtained, which are respectively recorded as T_A and T_B . Due to space limitations, the table below only shows the clarity score and rating of the whiteness sample 1.

Table 2. The clarity score and grade of the white wine sample 1 (out of 5 points)

| First set ratings | grade | Second set ratings | grade |
|-------------------|-------|--------------------|-------|
| 3 | 5 | 2 | 1 |
| 3 | 5 | 3 | 5 |
| 3 | 5 | 3 | 5 |
| 3 | 5 | 3 | 5 |
| 4 | 13.5 | 4 | 13.5 |
| 4 | 13.5 | 4 | 13.5 |
| 4 | 13.5 | 4 | 13.5 |
| 4 | 13.5 | 4 | 13.5 |
| 5 | 19.5 | 4 | 13.5 |
| 5 | 19.5 | 4 | 13.5 |

Calculated, $T_A = 113$, $T_B = 178$.

The test statistic Z was obtained from the above calculation, and the relevant data of the various indexes of the white wine sample 1 are shown in Table 3 below.

Table 3. Liquor sample 1 test statistic

| | Mann-Whitney U | Wilcoxon W | | Significant (two-tailed) |
|---------------------|----------------|------------|-------|--------------------------|
| Clarity | 42.00 | 97.00 | -0.70 | 0.51 |
| tone | 27.00 | 82.00 | -1.90 | 0.05 |
| Aroma purity | 32.00 | 87.00 | -1.70 | 0.09 |
| Aroma concentration | 28.00 | 83.00 | -1.80 | 0.07 |
| Aroma quality | 35.00 | 90.00 | -1.30 | 0.19 |
| Pureness of taste | 47.50 | 102.50 | -0.20 | 0.84 |
| Taste concentration | 30.50 | 85.50 | -1.60 | 0.11 |
| Persistence | 47.50 | 102.50 | -0.20 | 0.84 |
| Taste quality | 44.00 | 99.00 | -0.50 | 0.62 |
| Evaluation | 39.00 | 94.00 | -0.90 | 0.37 |
| Total scores | 30.50 | 85.50 | -1.50 | 0.14 |

Here, $\alpha=0.05$, that is, the upper quantile, $Z_{\alpha/2}=1.96$. As shown in Table 3, the indicators of white wine sample 1 are all less than 1.96, so there is no significant difference.

Then, the 605 scores of the red and white wines including the total score were subjected to the Mann-Whitney U test according to the above steps. It was found that only 90 of the 605 indicators had significant differences, accounting for only 14.9% of all items; There were no significant differences among the remaining 515 items, accounting for 85.1% of all scores. Therefore, this paper can initially draw conclusions that there is no significant difference in the evaluation results of the two wine drinkers, indicating that the data of the two groups are similar, that is, the distribution of the data is in normal.

2.3 Compare the Credibility of the Evaluation Results of the Two Wine Tasters

It is known from the above that the distribution of the evaluation results data of the two wine drinkers is normal, and then the credibility of the two groups should be determined. The Cronbach's Alpha analysis method will be used to analyze the credibility of the two sets of evaluation results. A group of appraisers with more objective and fair scores is used to determine the data set used for later analysis.

Cronbach's Alpha analysis was first proposed by Lee Cronbach in 1951 and was one of the most commonly used reliability analysis methods in the field of social science research such as psychological testing reliability (Peterson, 1994). This paper will also use Cronbach's Alpha method to analyze the credibility of the judges' scores. The main formula is:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_X^2} \right) \quad (5)$$

Where K is the number of items in a certain scale, $\sigma_{Y_i}^2$ is the variance of the current observed sample, and σ_X^2 is the variance of the total sample.

Since the influence of each scoring result on the credibility is needed while doing the credibility analysis based on Cronbach's Alpha method. To eliminate the influence of the dimension, the data is first standardized, and then the coefficient is solved. The values of the first and second groups of red wines are 0.852 and 0.750 and the values of the first and second groups of white wines are 0.8810 and 0.838. Based on this, two conclusions can be drawn: First, the Cronbach's Alpha values corresponding to the first group of judges are greater than 0.8 and larger than the second group, indicating that the internal consistency is better than the second and in the meantime the credibility is higher than the second. Second, the white wine's scoring data is more reliable than the red wine, so the white wine quality of the first group of wine judges will be analyzed below, and wine is taken as an example.

2.4 Use the Characteristic of Wine and Wine Grapes to Classify Wine

According to the data used in this paper, the chemical composition and related properties of wine grapes and wine constitute physical with chemical indicators, which may affect the quality of wine. Based on this, the wine can be graded using relevant data of physical with chemical indicators. Since physical with chemical indicators include anthocyanins, tannins, total phenols, total flavonoids, resveratrol, DPPH semi-inhibitory volume and color, etc., the factors that play a major role can be extracted and analyzed. Next, the principal component analysis (Jolliffe, 2002) and cluster analysis (Wilks, 2011) are used to analyze the two sets of scoring data.

2.4.1 Principal Component Analysis of Physical With Chemical Indicators of Wine Grapes and Wine

Principal component analysis, which is also known as principal component analysis, aims to use the idea of dimensionality reduction to transform multiple indicators into a few principal components. And the each of principal component can reflect most of the information of the original variables and is not duplicated. This method combines complex factors into several principal components while introducing multivariable, which simplifies the problem and makes the results more scientific and effective. From the above, the physical with chemical indicators of wine grapes and wine are too cumbersome, so it is necessary to extract the main influence factor, that is, the main component. The main steps are as follows:

First, the data is analyzed by the common factor analysis of variance, limited to the length, only part of the indicator data is intercepted.

Table 4. Common factor variance statistics

| Factor (g/L) | Start | extract |
|---------------|-------|---------|
| amino acid | 1.000 | 0.809 |
| VC content | 1.000 | 0.503 |
| Anthocyanin | 1.000 | 0.898 |
| protein | 1.000 | 0.809 |
| tartaric acid | 1.000 | 0.845 |
| Malic acid | 1.000 | 0.896 |
| Citric acid | 1.000 | 0.815 |

It can be seen from the above table that the extracted variables have strong commonality, which indicates that most of the information in the variables can be extracted by the factors, indicating that the results of the factor analysis are valid.

Then this paper selects 31 primary indicators for principal component analysis and normalize the data, calculating the contribution rate of all parties and the contribution rate of cumulative variance, the eigenvalues of the first eight factors are all greater than 1, and the cumulative contribution rate is above 80%. Therefore, the first eight factors can be extracted as the main factor.

Then, the matrix normalized by the original variables is multiplied by the eigenvector matrix to obtain the factor matrix. The linear relationship between the principal component and the physical with chemical indicators of the sample grape is as follows. Due to space limitations, only part of the data is intercepted.

Table 5. Linear relationship between principal components and physical with chemical indicators of wine grapes

| Red grapes | Z1 | Z2 | Z3 | Z4 | Z5 | Z6 | Z7 | Z8 |
|----------------|-------|-------|-------|-------|-------|-------|-------|------|
| Grape sample 1 | -1.04 | -1.27 | -0.98 | -0.83 | 1.07 | -1.00 | 0.07 | 0.75 |
| Grape sample 2 | -1.23 | -1.47 | 1.15 | 1.24 | 0.02 | -0.38 | -0.05 | 1.15 |
| Grape sample 3 | 0.63 | 0.68 | 0.57 | 2.98 | -1.28 | 2.57 | 1.33 | 0.82 |

2.4.2 Cluster Analysis of Factor Coefficients

For the classification of factor coefficients, European countries have strict hierarchical management of wine quality, and their classification levels are also different. For example, France divides wine into three levels (four levels before 2011), while Spain divides wine into five levels. This paper may be used to group data into four categories.

First determine the clustering method. Common clustering methods include k-means clustering algorithm, system clustering algorithm, SOM clustering algorithm and FCM clustering algorithm.

The K-means clustering method randomly selects K objects as the initial cluster center, then calculates the distance between each object and each seed cluster center, and assigning each object to the cluster center closest to it. Once all objects have been assigned, the cluster center of each cluster is recalculated based on the existing objects in the cluster until a termination condition is met. This method is fast and efficient, especially when dealing with large amounts of data, and the accuracy is high, but it needs to specify the category of the cluster manually.

System clustering is the most widely used clustering method at home and abroad. This method first considers the clustered samples or variables as a group, then determines the similar statistics between the classes and the classes, and chooses the closest combine two or more classes into one new class, next calculate the similarity statistics between the new class and other categories, finally select the closest two groups to merge into a new class until all the samples or variables are combined into one class. The method is that the system automatically lists the categories according to the distance between the data, and a tree diagram is obtained by the system clustering method. As for the meaning of the categories, it needs to be determined according to the tree diagram and experience and has certain subjective.

The SOM clustering algorithm is a neural network-based algorithm. The algorithm mainly finds the output unit with the minimum distance from the input node during the network learning process, and performs network update until the classification is completed, But the algorithm takes longer to update the iteration process. The FCM clustering algorithm applies the knowledge of the field of fuzzy set theory to overcome the shortcomings of the traditional algorithm. The membership degree determines that each data point belongs to a certain cluster, like the SOM clustering algorithm, the FCM clustering algorithm needs to perform iterative operations in the implementation, which will affect the application efficiency in specific use.

Since the system analysis method can automatically list the four categories needed in this paper, the generated tree diagram is conducive to intuitive conclusion, and the time is short and easy to implement. This paper chooses the system clustering method to get the wine classification. In the system clustering, the Ward clustering method adopts the variance analysis method. According to the same sample, the squared sum of the deviations should be small, and the squared sum of the deviations between the classes should be larger, compared with the shortest distance method, the longest distance method, the class average method, the center of gravity method and the intermediate distance method in system clustering, the error is smaller and the classification effect is better. Therefore, this paper chooses Ward clustering.

Then the Ward method is used for cluster analysis. The clustering results are shown in Figure 1:

It is known from Table 6 that in the cluster analysis of red wine, the average value of the first category is 74 points and above, the second type is 72-74, the third type is 70-72, and the fourth type is 70 and below; In the cluster analysis of white wine, the average value of the first category is 77 and above, the second category is 75-77, the third category is 73-75, and the fourth category is 73 and below. The classification criteria for red and white wines are thus determined. In addition, as can be seen from the above classification results, the average score value of white wine is almost always higher than the average value of red wine, so it is known that the quality of white wine is slightly higher than that of red wine.

2.5 Analysis of the Impact of Physical With Chemical Indicators On Wine Quality (Taking White Wine as an Example)

According to the classification of the above wines, the wines will have different mean scores under the physical with chemical indicators, and the law of quality classification will be initially revealed. Next, the influence of physical with chemical indicators on the quality classification should be analyzed. For the analysis of the relationship between physical with chemical indicators of wine grapes with wine and wine quality, this paper uses a typical correlation analysis method (Thompson, 2005). Take white grapes as an example, First, the composition of all white grapes X_1, X_2, \dots, X_{27} are expressed in the form of vectors, and correlation analysis is performed with a major component Y of white wine to obtain multiple sets of related data Z_1, Z_2, \dots, Z_8 . Then, the correlation matrix between the physical with chemical indicators of wine grapes and wine can be obtained by multiple linear regression analysis, in the meantime the regression coefficients corresponding to various components are shown in the following table. Due to space limitations, this article will only analyze white wine, and the analysis of white wine is also the same.

Table 7. Regression analysis coefficient table of physical with chemical indicators of white grape

| | Z1 | Z2 | Z3 | Z4 | Z5 | Z6 | Z7 | Z8 |
|------------------|-------|-------|-----|-----|----|------|------|-----|
| Anthocyanin | 21.6 | -22.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tannin | 19.8 | -20.3 | 0 | 0.3 | 0 | 0 | 0.2 | 0.2 |
| Total phenol | 21.9 | -27.4 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total flavonoids | 15.5 | -16.1 | 0 | 0.3 | 0 | 0 | 0 | 0.4 |
| Resveratrol | 0 | 0 | 0 | 0.4 | 0 | 0 | 0 | 0 |
| DPPH | 18.9 | -19.4 | 0 | 0.3 | 0 | 0 | 0 | 0.3 |
| L*(D65) | -17.5 | 18.5 | 0.3 | 0 | 0 | -0.4 | 0 | 0 |
| a*(D65) | 0 | 0 | 0 | 0 | 0 | 0 | -0.5 | 0 |
| b*(D65) | 0 | 0 | 0 | 0 | 0 | 0.5 | 0 | 0 |

The regression relationship of each component can be obtained from Table 7 as shown in the following formula:

$$\left\{ \begin{array}{l}
 \text{anthocyanin} = 21.603Z_1 - 22.178Z_2 \\
 \text{Tannin} = 19.766Z_1 - 20.273Z_2 + 0.178Z_7 + 0.189Z_8 \\
 \text{Total phenol} = 21.85Z_1 - 27.377Z_2 \\
 \text{Total flavonoids of total liquor} = 15.452Z_1 - 16.08Z_2 + 0.265Z_4 + 0.388Z_8 \\
 \text{Resveratrol} = 0.4Z_4 \\
 \text{DPPH} = 18.85Z_1 - 19.4Z_2 + 0.305Z_4 + 0.283Z_8 \\
 \text{L}^*(\text{D65}) = -17.5Z_1 + 18.46Z_2 + 0.287Z_3 - 0.37Z_6 \\
 \text{a}^*(\text{D65}) = -0.46Z_7 \\
 \text{b}^*(\text{D65}) = 0.502Z_6
 \end{array} \right. \quad (6)$$

According to the above results, there is a strong linear relationship between the physical with chemical indicators of wine and the physical with chemical indicators of wine grapes. If it can be proved that there is a certain relationship between the physical with chemical indicators of wine and the quality of wine, then the effect of the specific values of physical with chemical indicators of grapes and wine on the quality of wine can be obtained. Because the physical with chemical indicators of wine are numerous, each physical with chemical indicator affects the quality of the wine. Therefore, the aromatic substances will be selected for analysis to explore the effect on the aroma of wine quality, that is, using the method of multiple linear regression, the first 20 samples of white wine were selected, and the aromatic substances were analyzed. The results obtained are shown in the following table.

Table 8. Multiple linear regression results

| substance | B | Standard error | Beta coefficient |
|----------------|--------|----------------|------------------|
| C_2H_6O | 0.071 | 0.000 | 0.358 |
| $C_5H_{10}O_2$ | -5.995 | 0.000 | -0.985 |
| $C_6H_{12}O_2$ | -0.690 | 0.000 | -0.118 |
| $C_6H_{12}O_2$ | -0.336 | 0.000 | -0.087 |
| C_3H_8O | 5.808 | 0.000 | 0.764 |
| $C_{11}H_{24}$ | 2.444 | 0.000 | 0.505 |
| $C_7H_{14}O_2$ | 0.374 | 0.000 | 2.726 |
| $C_{10}H_{16}$ | 1.360 | 0.000 | 0.561 |

The Beta coefficient corresponds to the coefficient before each aromatic substance, indicating that the aroma score in wine quality has a positive correlation with C_2H_6O , $C_6H_{12}O_2$, C_3H_8O , $C_{11}H_{24}$, $C_7H_{14}O_2$, $C_5H_{10}O_2$ and $C_{10}H_{16}$. The standard error in the table is 0, indicating that the model is well established. In addition, when the red wine was selected to perform the same operation on its aromatic substance, the obtained result was consistent with the red wine, which proves the robustness of the model. Due to space limitations, this article will not describe its process in detail.

3. Conclusions

This paper takes wine quality evaluation as the research object, establishes the analysis and evaluation model of wine quality, and explores the influence of physical with chemical indicators of wine grapes and wine on the wine quality. Firstly, the Mann-Whitney U test is used to show that the data of the two groups are similar, that is, the distribution of the data conforms to the normal law. Then using the Cronbach Alpha coefficient method, the credibility of the first group of wine scores is significantly greater than that of the second group and the white wine scores are more reliable than the red wine, so that the first set of data and white wine can be applied for follow-up studies. Next, the main indicators are extracted by principal component analysis, and the Ward method in cluster analysis is used to classify the wine into four grades according to the quality score of the wine. Then, based on the extracted principal components which is the same as physical with chemical indicators, this paper does the multiple linear regression analysis of wine quality and finds that there is a positive correlation linear relationship between the scores of the aroma of wine quality and C_2H_6O , $C_6H_{12}O_2$, C_3H_8O , $C_{11}H_{24}$, $C_7H_{14}O_2$, $C_5H_{10}O_2$ and $C_{10}H_{16}$. It can be judged that the aromatic substances in the wine such as C_2H_6O have a regular influence on the odor of the wine, and it is inferred that other physical and chemical properties have a similar regular relationship with the wine quality. This provides an effective reference for the analysis and evaluation of wine quality by using physical with chemical indicators such as aromatic substances in wine in the future.

The statistical ideas embodied in the process of establishing wine analysis and evaluation models have the following disadvantages: Firstly, the experimental data is over reliant, and the conclusions obtained when the experimental data is unreliable are not strictly scientific. Secondly, the lack of strict theoretical derivation rationality makes this model have certain contingency. Furthermore, for statistical raw data, the analytical methods used are also important, and the flaws in the methods can also lead to irrational conclusions.

But the advantages of this model are also obvious: First, due to the development of modern science and technology, the

precision of the measuring instrument strongly guarantees the accuracy of the physical with chemical indicators of the measured wine and wine grapes. Secondly, the use of physical with chemical indicators to analyze the quality of wine can not only avoid the subjective mistakes of the tasters, but also save time in testing and reduce labor waste. Furthermore, using physical with chemical indicators to evaluate the quality of wine can effectively avoid artificial operability, making the results more convincing.

Therefore, this paper uses statistical ideas to ensure a certain reliability and can easily obtain a model that is more accurate for wine analysis and evaluation. It confirms that the physical with chemical indicators of grapes and wine can reflect the quality of wine to a certain extent, providing an effective reference for the improvement of wine quality evaluation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (51475095), the Fundamental Research Funds for the Central Universities (21618412), Key Project of Guangdong Natural Science Foundation (2016A030311041), 2015 Guangdong Special Support Scheme (2014TQ01X706), High-level Talent Scheme of Guangdong Education Department (2014-2016), the Guangdong Natural Science Foundation (2017A030313401).

References

- China Society for Industrial and Applied Mathematics. (2012). Higher Education Club Cup National Contest on Mathematical Modeling for College Students. Retrieved September 7, 2009, from <http://www.mcm.edu.cn>
- Jolliffe, I. T. (2002). *Principal component analysis*. <http://dx.doi.org/10.1007/b98835>
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318), 399-402. <https://doi.org/10.2307/2283970>
- Morozova, K., Schmidt, O., & Schwack, W. (2015). Effect of headspace volume, ascorbic acid and sulphur dioxide on oxidative status and sensory profile of Riesling wine. *European Food Research and Technology*, 240(1), 205-221. <https://doi.org/10.1007/s00217-014-2321-x>
- Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13-20. <http://dx.doi.org/10.20982/tqmp.04.1.p013>
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of consumer research*, 21(2), 381-391. <http://dx.doi.org/10.1086/209405>
- Thompson, B. (2005). Canonical correlation analysis. *Encyclopedia of statistics in behavioral science*. In B. Everitt & D.C. Howell (Eds.), 1, 192-196.
- Wang, X. J., & Guan, Z. L. (2016). Evaluation model of grape wine quality based on BP neural network. *Proceedings of the 2016 INTERNATIONAL CONFERENCE ON LOGISTICS, INFORMATICS AND SERVICE SCIENCES (LISS' 2016)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA.
- Wilks, D. S. (2011). Cluster analysis. *International geophysics*, 100, 603-616. <https://doi.org/10.1016/B978-0-12-385022-5.00015-4>
- Yakuba, Y. F., Temerdashev, Z. A., & Khalaf'yan, A. A. (2016). Application of Ranging Analysis to the Quality Assessment of Wines on a Nominal Scale. *Journal of analytical chemistry*, 71(2), 205-214. <https://doi.org/10.1134/S1061934816020155>
- Zhao, P., Wang, H., & Li, H. (2018). Characterization of the effect of short-term high temperature and vibration on wine by quantitative descriptive analysis and solid phase microextraction-gas chromatography-mass spectrometry. *Acta Alimentaria*, 47(2), 236-244. <https://doi.org/10.1556/066.2017.0007>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.