

Research Article

Relation between Grape Wine Quality and Related Physicochemical Indexes

¹Zi-Yue Chen, ²Yuan-Biao Zhang and ¹Qiu-Ye Qian

¹International Business School, Jinan University, Zhuhai 519070, China

²Mathematical Modeling Innovative Practice Base, Jinan University, Zhuhai, 519070, China Packaging Engineering Institute, Jinan University, Zhuhai 519070, China and Key Laboratory of Product Packaging and Logistics of Guangdong Higher Education Institutes, Jinan University, Zhuhai, 519070, China

Abstract: The aim of this study is to evaluate grape wine quality more objectively by reducing the error of traditional grape-wine-quality evaluation. On combining grape wine quality and physicochemical index of grapevine, we provided a grape-wine-quality evaluation model by grapevine's physicochemical index in this study. Firstly, evaluations of the tasters are analyzed, for eliminating the disturbance caused by their individual difference. Then, relationship between grape wine and grapevines are analyzed. Inherent mechanism which affects the grape wine quality was figured out based on description of grape wine quality by physicochemical index of grapevine. Finally, we evaluated the grape wine quality by physicochemical index of grapevine. Additionally, rationality of the model is verified by statistical test while the accuracy of the results is verified by comparison with the evaluating results made by tasters.

Keywords: Grape wine quality evaluation, multiple linear stepwise regression, principal component analysis, significant test

INTRODUCTION

At present, classification of grape wine differs from country to country while distinguish of grape wine quality are similar which mainly depends on sensory quality (Wen-Jing, 2007). Main components in aroma of grape wines are summarized in a research on aromatic substance of grape wine since aroma is a pivotal index of grape wine quality evaluation (Yu *et al.*, 2005). The 12 main aromatic sources in grape wine, their sensory characteristics and their influence on grape wine quality are described in a research (Ji-Ming, 2005). In fact, evaluation of grape wine quality includes the appraisal, taste and so on besides aroma. By building regression equation between grape wine quality and four factors including aging time, alcohol content and residual sugar, relationship between grape wine sensory quality and each factor has been figured out in a research (Li *et al.*, 2005).

Sensory evaluation by tasters is commonly used on evaluating the grape wine sensory quality. During the evaluation, tasters grade several indexes of the grape wines after tasting them. Based on the summation of the indexes, the quality of grape wine is finally evaluated. However, the side effect of the evaluation method is the evaluation error due to individual difference of the

tasters, which reduce the objectivity of the results. To reduce evaluation scale of tasters, confidence interval method is better than standardization method, leading the difference of grape wine quality more objectively (Li *et al.*, 2006).

Considering the direct relationship between grapevine and grape wine quality, physicochemical indexes of grapevines reflect the quality of grape wines to some extent. As a result, the relationship between grape wine quality and physicochemical indexes of grapevines are researched in this study. Based on the data on <http://www.mcm.edu.cn/>, a model which estimates grape wine quality by physicochemical indexes of grapevines was established.

SELECTION OF THE TASTERS BASED ON SIGNIFICANT DIFFERENCE TESTING MODEL

Two groups of tasters were chosen to evaluating the 27 samples of red wine. The evaluation dimension of the evaluation includes appearance analysis, aromatic analysis, taste analysis and overall assessment. The appearance analysis contains clarity and hue while aromatic analysis and taste analysis contains purity, concentration and quality.

Table 1: Standardization result of red wine sample 1 by tasters of group 2

		Taster 1	Taster 2	Taster 3	Taster 4	Taster 5	Taster 6	Taster 7	Taster 8	Taster 9	Taster 10
Appearance	Clarity	-0.354	0.196	-0.267	0.000	-2.736	0.477	-1.622	0.000	-2.098	0.000
Analysis	Hue	1.800	0.598	0.964	0.730	0.151	0.707	1.069	0.603	1.500	0.492
	Purity	1.718	-1.386	-0.251	-2.065	-0.866	-0.251	-1.369	0.870	-0.371	-1.494
Aromatic Analysis	Concentration	0.175	0.120	-0.073	-0.461	-0.964	0.863	-0.029	-1.136	0.577	-0.240
	Quality	-0.137	-0.945	0.725	-2.005	-0.725	0.347	-0.957	0.224	-2.138	-1.520
	Purity	-1.671	0.064	0.990	-2.475	-0.870	1.768	-0.058	-0.649	-0.073	0.354
	Concentration	-1.735	1.414	0.755	-2.082	-0.724	1.151	1.237	0.226	0.417	1.832
Taste	Persistence	-0.755	1.497	-0.707	-0.884	-0.598	1.218	1.206	-0.653	1.778	0.894
Analysis	Quality	-1.293	0.535	1.020	-2.000	-1.975	-0.981	-0.649	0.277	-1.055	-1.541
Overall assessment		-0.251	0.058	-1.500	-1.651	-1.956	0.392	-0.578	0.283	0.592	0.196

Table 2: F-test results of the red wine sample 1 by two group of tasters

		S_1^2	S_2^2	F	F0.025	F0.975	Whether significant difference exist
Appearance analysis	Clarity	0.663	0.163	4.061	4.026	0.248	Y
	Hue	0.846	0.117	7.204	4.026	0.248	Y
Aromatic analysis	Purity	0.699	0.177	3.950	4.026	0.248	N
	Concentration	0.951	0.578	1.644	4.026	0.248	N
Taste analysis	Quality	0.977	0.163	6.000	4.026	0.248	Y
	Purity	0.629	0.052	12.009	4.026	0.248	Y
	Concentration	1.128	1.598	0.706	4.026	0.248	N
	Persistence	0.637	0.757	0.842	4.026	0.248	N
	Quality	0.367	0.193	1.899	4.026	0.248	N
Overall assessment		0.255	0.036	7.078	4.026	0.248	Y

It is variance analysis that could deal with the problem that whether there are significant differences between the two groups of tasters. However, further calculation is needed for judging whether the variance of grades are difference between the two groups of tasters in one index of a sample, since the presumption of testing significant difference of mean value is that the variance of each sample equals.

Based on the analysis, firstly, the same indexes of each taster in different samples were standardized, avoiding the influence of individual difference. Then each standardized index of the two groups was tested by *F-test* and *t-test* with the salience value of 0.05, for judging whether the significant difference of each index graded by two groups of tasters exists. Finally, the grades of tasters with higher reliability were selected as the evaluation standard of red wine according to the rules that the group with smaller variation is better.

Standardization of the data: Firstly, standardize the data with standard deviation, as is in Formula (1):

$$A_{ijkn} = \frac{P_{ijkn} - \bar{P}_{ijn}}{\sigma_{ijn}} \quad (1)$$

According to the formula

P_{ijkn} = The original mark of the Index n by Taster No. j of Group i in Sample k .

A_{ijkn} = The standardized result of P_{ijkn} .

\bar{P}_{ijn} = The mean value of the k groups of original mark of the Index n by Taster No. j of Group i .

σ_{ijn} = The standard deviation of the Index n on all the samples of the original marks graded by Taster No. j of Group i .

The standardization result of red wine graded by tasters of Group 2 is figured out as the following Table 1.

F-test: significant difference test on standard deviation of grades: Firstly, build the hypothesis H_0 : $\sigma_{2kn} = \sigma_{1kn}$, which means that there is little significant difference between the grades of an index in the same sample marked by two group tasters.

The value of $\frac{S_{1kn}^2}{S_{2kn}^2}$ cannot neither be too large nor too small if H_0 is established. Hence, statistic F is selected as Formula (2):

$$F = \frac{S_{1kn}^2}{S_{2kn}^2} \quad (2)$$

F presents the F distribution with the degrees of the freedom valuing 9. As $\alpha=0.05$, it could be figured out that $F_{0.975} = 0.2484$ and $F_{0.025} = 4.0260$.

After calculating each F Statistics of each index graded by different tasters, we judged that whether the value is between $F_{0.975}$ and $F_{0.025}$. If the value is not between the intervals, the significant difference exists.

Data in Table 2 shows the results of F test of the two-group tasters based on Red Wine Sample 1.

For 27 red wine samples, 270 times of F test was calculated since there are 10 indexes in each sample, which contains 58 significant difference analyses.

t Test: testing whether the mean value of indexes with no significant difference equals: Choose the t Statistic as Formula (3):

Table 3: Results of the significant difference test

F-Statistics	t-statistics	Red wine
No significant difference	/	211
Significant difference	/	58
No significant difference	No significant difference	64
Significant difference	Significant difference	206

Table 4: Reliability comparison

Index Number in F Test	Red wine
Variance of group 1 is smaller	57
Variance of group 2 is smaller	1

$$t = \frac{\bar{A}_1 - \bar{A}_2}{\sqrt{(S_1^2 - S_2^2)/N}} \quad (3)$$

In Formula (3), S_1^2 and S_2^2 are the sample variances while N is the sample size of each group. In this study, $N = 10$ while \bar{A}_1 and \bar{A}_2 represents the mean value of the same indexes graded by 2 groups of tasters.

t test was used in the indexes which have the same variance according to F test. The results are as Table 3.

We compared variance of the indexes with the same mean value and established that the evaluation index with smaller variance is better as is in Table 4.

Tasters in Group 1 enjoy higher reliability of evaluation results in red wine according to the comparison result of variance.

PHYSICOCHEMICAL INDEX EXTRACTION OF RED WINE BASED ON PRINCIPAL COMPONENT ANALYSIS

Due to the large size of physicochemical index and the uncertainty relation between every two indexes, physicochemical indexes were classified and processed

before merging indexes with strong relationship based on the characteristic of the indexes.

Firstly, we made correlation analysis on standardized indexes for judging that whether the multi-collinearity exists among the indexes. Then, based on principal component analysis, we merged the remaining indexes with strong relationship, aiming to simplifying calculation and eliminating the multi-collinearity among indexes.

Correlation coefficient matrix: We made the correlation analysis of standardized physicochemical indexes with SPSS 17.0. The results are shown in Table 5.

Based on the correlation coefficient matrix of grapevines' physicochemical indexes, results could be figured out that reducing sugar, total sugar and soluble solids present to be remarkably positive correlation, which means there is huge overlap of information among them.

However, the test of coefficient matrix may face difficulty when the multi-collinear is too strong in multiple-linear-regression model. It may cause the situation that F test is passed while the t test of the coefficient matrix cannot pass, which may further leading to that the meaning of estimated coefficient contradict common sense.

As a result, principal component of n grapevine physicochemical index is extracted by principal component analysis, aiming at getting the independent principal component, which are F_1, F_2, \dots, F_m , based on simplifying the calculation of physicochemical index. The principal component analysis used here could not only simplify the regression equation but eliminate the influence caused by correlation among varieties.

Table 5: Correlation matrix of the physicochemical indexes

	Aspartic acid	Threonine	Serine	...	Juice yield	Lightness L^*	Red-green value a^*	Yellow-blue value b^*
Aspartic acid	1.00							
Threonine	0.31	1.00						
Serine	0.39	0.25	1.00					
...				...				
Juice yield	0.17	0.10	-0.12	...	1.00			
Lightness L^*	-0.20	-0.29	-0.19	...	-0.44	1.00		
Red-green value a^*	0.11	-0.17	0.09	...	-0.32	0.33	1.00	
Yellow-blue value b^*	0.24	0.06	0.19	...	-0.22	-0.05	0.87	1.00

Table 6: Variance contribution and cumulative of physicochemical indexes

Component	Initial feature value			Extraction		
	Sum	Variance %	Accumulation %	Sum	Variance %	Accumulation %
1	9.01	17.02	17.02	9.02	17.02	17.02
2	7.38	13.93	30.94	7.38	13.93	30.94
3	6.09	11.49	42.44	6.09	11.49	42.44
...
12	1.23	2.32	86.88	1.23	2.32	86.88
13	1.06	1.99	88.87	1.06	1.99	88.87
...
52	-6.75E-16	-1.27E-15	100.00			
53	-6.90E-16	-1.30E-15	100.00			

Table 7: Coefficient matrix of principal component

	Principal component									
	1	2	3	4	...	10	11	12	13	
	0.373	0.450	0.334	0.067	...	0.142	-0.254	-0.174	0.141	
	0.485	0.153	0.398	0.432	...	0.049	0.002	0.128	0.053	
Aspartic acid Threonine Serine	-0.090	0.638	0.487	-0.060	...	0.175	-0.044	-0.046	-0.037	
...	
Juice yield	0.479	-0.110	-0.139	-0.350	...	0.025	-0.055	0.100	-0.211	
Lightness L*	-0.475	0.086	-0.408	0.286	...	-0.211	0.042	-0.001	0.018	
Red-green value a*	-0.266	0.058	0.245	0.426	...	0.067	0.204	0.127	-0.028	
Yellow-blue value b*	-0.099	0.059	0.507	0.371	...	0.256	0.233	0.078	0.068	

Principal component analysis: Due to the assumption of multiple linear regressions that there is no accurate linear relationship between each variety, principal component analysis on grapevine physicochemical index has been made for better regression effects, which mean the analysis could reduce multiple linear regressions among the indexes.

A large number of indexes could be merged to several comprehensive indexes by principal component analysis without losing information. According to the theory of principal component analysis, enough information is reflected, if cumulative of the first *R* principal component reaches 85%. Hence, after principal component analysis on grapevine physicochemical index, analysis of cumulative was made for simplifying the main indexes of grapevine quality.

Extraction of principal component: Variance contribution and cumulative were figured out by SPSS 17.0 as is shown in Table 6. As the cumulative of the first 13 principal components is 88.87%, we combined feature value of the components to new comprehensive indexes which are independent to each other, leading to a rounded reflection on grapevine quality.

Coefficient matrix of principal component: Coefficient matrix of principal component was figured out by SPSS 17.0 as is shown in Table 7.

The data in Table 7 represent the load that principal components have on variables. Based on the data, expression of each principal component was figured out as Formula (4):

$$F_i = \sum_{j=1}^n c_{ij} \cdot x_j \tag{4}$$

In the formula,

c_{ij} = The coefficient of the physicochemical index *j* in principal component *i*.

x_j = Physicochemical index after standardization.

GRAPE WINE QUALITY EVALUATION MODEL BASED ON MULTIPLE LINEAR STEPWISE REGRESSIONS

Large number of information could be extracted by stepwise regression since influence of each factor is

considered after the comprehensive influence by other factors. Therefore, stepwise regression could be used for describing the factors which influence grape wine quality since the influences between factors are controlled. There are two main advantages of the stepwise regression. One is to extract the factors which affect grape wine quality among quantities of factors. The other is to express the saliency of each factor which is easy for comparison and selection.

Establishment of multiple stepwise regression equation:

Multiple stepwise regressions were mainly used on selecting indexes in this study. Since many factors of grapevine could lead to a characteristic of grape wine, factors with remarkable influence should be extracted. Firstly, influences on grape wine quality by all independent variable, principal component, were considered. Then, principal components were introduced to the stepwise regression equation based on the saliency. Principal Components with large saliency enjoys the priority of introducing to the equation while components with small saliency might never been introduced to the equation. Additionally, introduced components may lose its significance when a new component is introduced to the equation, which would be eliminated from the multiple stepwise regression equation.

Firstly, grape wine quality was chosen as dependent variable and physicochemical indexes are chosen as the independent variable of the regression before *F* value set. Since the evaluation reliability of the taster in Group 2 is higher, evaluation of Group 2 has been chosen as the dependent variable. Meanwhile, independent variables were represented by 13 principal components by principal component analysis.

Before the stepwise regression, we tested whether each variable is in the interval of *F* test for ensuring that the regression equation contains principal component with great influence only. In this period, we established that the significance level $\alpha=0.05$. When a variable is introduced, critical value of *F* test is *F1*, while critical value of *F* test is *F2* when a variable is eliminated. Additionally, $F1 > F2$, is established as a standard when a principal component is introduced or eliminated.

Then, stepwise regression was made for a simple linear regression model as the basic model according to

Table 8: Partial-regression-coefficient significant test of independent variable

	Non-standardized regressive coefficient		Standard coefficient	t	Sig.
	B	S.E.			
Constant	0.005	0.112	-0.514	0.040	0.968
X4	-0.072	0.024		-3.000	0.006
Constant	0.005	0.102	-0.515	0.044	0.965
X4	-0.072	0.022	0.384	-3.288	0.003
X2	0.034	0.014		2.457	0.022
Constant	0.005	0.094	-0.515	0.048	0.026
X4	-0.072	0.020	0.384	-3.580	0.002
X2	0.034	0.013	0.335	2.674	0.014
X1	0.024	0.010		2.333	0.029

Table 9: Classification of the grapevines

Classification	1 st rate grapevine	2 nd rate grapevine	3 rd rate grapevine	4 th rate grapevine
Numeration of grapevine variety	23	5	24	27
	9	19	7	25
	3	1	14	6
	21	17	26	12
	22	13	8	11
	2	15	10	18
	20	16	4	

the fitting effect. Principal Component 2 to Principal 13 was added to the basic model according to the test for eliminating variable with unapparent estimating parameters.

According to the stepwise regression, Principal Component 4, Principal Component 2 and Principal Component 1 are in the regression equation. Since the three principal components contains 53 physicochemical indexes such as aspartic acid, threonine and glutamate, the regression on grape wine quality based on the 53 physicochemical indexes of grapevine as is in Formula (5):

$$Q = C_1 \sum_{j=1}^{54} c_{4j} x_j + C_2 \sum_{j=1}^{54} c_{2j} x_j + C_3 \sum_{j=1}^{54} c_{1j} x_j \quad (5)$$

In the formula, $C_1 = -0.515$, $C_2 = 0.384$, $C_3 = 0.335$. The value of c_{4j} is $\alpha_1 = (0.067, 0.432, -0.060, \dots, 0.371)$. The value of c_{2j} is $\alpha_2 = (0.450, 0.153, 0.638, \dots, 0.059)$. The value of c_{1j} is $\alpha_3 = (0.373, 0.485, -0.090, \dots, -0.099)$.

Rationality test of the model: Analysis on the statistical results of the stepwise regression model was figured out.

- Variance Analysis: $F = 8.469$, $Sig. = 0.001$, $p < 0.05$ means that the multiple regression equation is of great statistical significance.
- Comprehensive Analysis of the Model: R (Correlation Coefficient) = 0.724. R^2 (Coefficient of Determination) = 0.525. R^2 (Adjusted R Square) = 0.463. (Std. Error of the Estimate) = 0.489.
- Multi-collinearity Analysis: The tolerance of the three variable is 1.000 while $VIF = 1.000 < 2$, which means the weak Multi-collinearity.

- Autocorrelation Test: The autocorrelation is weak since the value of DW is 0.02
- Significance Test of Partial Regression Coefficient: The related parameters namely constant term of the multiple regression equation, Partial regression coefficient of the variable (B) and sampling error are figured out as is shown on Table 8.

According to the regression coefficient, the grape wine quality is mainly affected by three principal components, which are all made up by 53 physicochemical indexes of grapevine. Therefore, it is the 53 physicochemical indexes of grapevine that influence the grape wine quality.

The *Sig.* value of both constant and independent variable is far smaller than 0.05 and the *p* value of the model is 0.001 based on the variance analysis, which means the model is apparent due to the variables with statistical significance. Thus, the established multivariate linear regression equation is the optimality equation for the problem.

Accuracy of the model: Due to the unique application of the grapevine and the direct relationship between grape wine quality and the physicochemical indexes of grapevine, we classified the grapevine based on the quality of grape wine.

Firstly, physicochemical indexes of each cultivar of grapevine were substituted to the regression equation before the scheduling of the 27 cultivar of red grapevine. Based on the ranking result, we defined that the first 7 cultivar of grapevine are First Rate Grapevine, the next 7 are Second Rate Grapevine and the rest were deduced by analogy as is shown on Table 9.

Then, the grape wine quality were evaluated based on the regression equation, then 5 best cultivar of grapevines could be figured out, whose numeration are

23, 9, 21 and 22. Due to the direct effect that grapevine has on grape wine quality, we scheduled the grape wine by grape wine quality based on evaluation of the tasters. The numerations of the top 5 grapevine are 23, 9, 22, 3 and 19. By comparing the two ranking results, a conclusion could be made that the results calculated by the two methods is of high similarity, which verify the rationality and accuracy of the multiple linear regression.

CONCLUSION

Grape wine quality and physicochemical indexes of grapevine are connected in this study, leading to a model of evaluating grape wine quality by physicochemical indexes.

Firstly, *F* test and *t* test were used in the analysis whether the significance difference of the two groups of tasters exists. *t* test was made after the *F* test passed, avoiding the assumption for variance analysis. Then, regression was used for describing the relationship between grape wine quality and the physicochemical indexes since the chemical reaction during the brewing time is too comprehensive to describe with mechanism analysis, making the abstract problem concrete. As the multiple linear regression is greatly influenced by Multi-collinearity, principal component analysis was used in this study to reduce the side effect. Finally, the

rationality of the model was verified by statistical test while the accuracy of the model was verified by the comparison between data and the result calculated by the model.

Heteroscedasticity cannot be avoided despite of the standardization in this study. Therefore, the model would be of more accuracy if the influence of heteroscedasticity is considered.

REFERENCES

- Ji-Ming, L., 2005. Oak aroma of grape wine. *Sino-Overseas Grapevine Wine*, 8: 47-48.
- Li, H., L. Shu-Dong, W. Hua and Z. Yu-Lin, 2006. Studies on the statistical analyses methods for sensory evaluation results of wine. *J. Chinese Inst. Food Sci. Technol.*, 6(2): 126-131.
- Li, H., Y. Yong-Feng, G. Ming-Hao and L. Shu-Wen, 2005. Effects of different factors on tasting results of dry red wine. *J. Biomath.*, 20: 223-228.
- Wen-Jing, W., 2007. Development of study on sensory evaluation in wine. *Liquor Mak.*, 13(4): 57-59.
- Yu, J., L. Jing-Ming, W. Ji-Hong and G. Yi-Qiang, 2005. Research progress of aromatic substance in grape wine. *Sino-Overseas Grapevine Wine*, 3: 48-51.