Advance Journal of Food Science and Technology 9(4): 296-301, 2015
DOI: 10.19026/ajfst.9.2012
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: March 3, 2015 Accepted: March 25, 2015 Published: August 15, 2015

Research Article

The Evaluation Model of Grape Wine Quality Based on Multivariate Statistical Methods

[1]Lihui Zhou and [2]Ruiqi Song
[1]College of Sciences,
[2]Hebei United University, Tangshan 063009, China

Abstract: The purpose of this study is to establish contacts of the physiochemical indexes between the Wine grapes and the Wines. Because of a wide range of physiochemical indexes, in order to more clearly reflect the contact between Wine grapes and Wines, firstly, the principal component analysis is used to select principal components and the correlation matrix is established based on the corresponding variables of principal components. And then, by stepwise regression method, the function of the relationship of physiochemical indexes between the Wines and Wine grapes is fitted, through which shows a strong correlation between the physiochemical indexes of Wines and Wine grapes.

Keywords: Principal component analysis, stepwise regression analysis, wine evaluation

INTRODUCTION

During determining the quality of wine, a number of qualified wine-tasting are usually employed to tasting. Each wine-tasting gives a score to classification index of wine after the tasting of the wines, then summed to obtain the total score to determine the quality of the wine. There is a direct relationship between the quality of Wine grapes and the quality of Wines, so the physiochemical indexes of Wine and Wine grapes will reflect the quality of the Wines and Wine grapes to some extent (Li *et al*., 2011). The examples give the composition data of some Wines and Wine grapes in a given year. This study will attempt to establish a mathematical model to analyze the connection with the physiochemical indexes of Wines and Wine grapes (Gao, 2004).

From this study, we want to know which are the important physiochemical indicators having a significant impact on Wines and Wine grapes. And we also want to know the models of the important physiochemical indicators and Wines and Wine grapes. There are too many physicochemical indicators to establish the model of Wine grapes and wines. Therefore, in order to obtain a clearer regression equation on the physicochemical indicators between Wine grapes and wines, firstly the principal component analysis is used to obtain the main component, further regression equation is based on the corresponding physiochemical indexes by stepwise regression analysis and then the connection of the physicochemical indicators between Wine grapes and Wines is obtained.

MATERIALS AND METHODS

Model assumes: It is assumed that the data used in this study are real and effective and have a systematic analysis of the value.

In this study, obviously erroneous data are manually modified and data are accurate and objective, that is not considered view error.

The sample data can be approximated as from a normal or near-normal distribution.

Symbol description: The original variables of Wines are:

X_1: Anthocyanins, X_2: Tannin, X_3: Total phenols, X_4: Wine total falconoid, X_5: Resveratrol, X_6: DPPH inhibition half volume, X_9: Color and b * (D65), X_{10}: Color H (D65), X_{11}: Color C (D65), X_{13}: Grape total flavonoids, X_{16}: Total sugar, X_{17}: Sugar, X_{18}: Soluble solids, X_{21}: Solid acid ratio, X_{22}: Dry matter content, X_{25}: Stems ratio, X_{29}: Skin color a * (+red; -green), X_{31}: Skin color H, X_{32}: Skin color C (Chernev, 1997)

The original variables of Wine grapes are:

Y_1: Anthocyanins, Y_2: Tannin, Y_3: Total phenolic, Y_4: Wine total flavonoids, Y_5: DPPH inhibition half volume, Y_6: Color (HD65), Y_7: Color (CD65)

Model establish:

The basic principles of the principal component analysis: Assume that there are n samples, each sample

Corresponding Author: Lihui Zhou, College of Sciences, Hebei United University, Tangshan 063009, China

has a total of *p* variables, an n*p matrix of order data is constituted:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

When p is large, it is problematic for expedition in p-dimensional space. To overcome this difficulty, the dimension is needed to reduce, which uses relatively few comprehensive index instead of the original variables more indicators, but these less comprehensive index can reflect as much as possible of the original indicators that are more variable reflects the information, while between them is independent of each other (Fang and Pan, 1982).

Definition: Remember $x_1, x_2, \cdots, x_p$ original variable index, $z_1, z_2, \cdots, z_m$ $(m \le p)$ for the new variable index:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ \cdots\cdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{cases}$$

The determining principle of coefficient of l_{ij} is: $l_{i1}^2 + \cdots + l_{ip}^2 = 1$:

- $Cov(z_i, z_j) = 0$ $(i = j; i, j = 1,2,\cdots,m)$
- z_1 has the greatest variance in all linear combinations of $x_1, x_2, \cdots, x_p$. z_2 is not related to z_1, and has the second largest variance in all linear combinations of $x_1, x_2, \cdots, x_p$. z_m is not related to $z_1, z_2, \cdots, z_{m-1}$ and has the M^{th} largest variance in all linear combinations of $x_1, x_2, \cdots, x_p$ (Yu, 1993)

From the above analysis, the essence of principal component analysis is to determine the load l_{ij} and proved mathematically, they are the eigenvectors of the m larger eigenvalues to the correlation matrix.

The calculation step of principal component analysis:

- **The normalization processing of raw data standardization:** Because of different dimension of various indicators, it is first necessary to normalize the data. Standardized formula is as follows:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{s_i}$$

- Establish the correlation coefficient matrix variable:

$$r_{ij} = \frac{\sum (x_{ik}^* - \bar{x}_i^*)(x_{jk}^* - \bar{x}_j^*)}{\sqrt{\sum (x_{ik}^* - \bar{x}_i^*)^2} \sqrt{\sum (x_{jk}^* - \bar{x}_j^*)^2}}$$

$$R = (r_{ij})_{p \times p}$$

- Seeking the eigenvalues and their corresponding eigenvectors of R: (He, 2004):

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$$

$$a_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{pmatrix}, a_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{pmatrix}, \cdots\cdots, a_p = \begin{pmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{pmatrix}$$

- Write the principal components:

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{pi}X_p,\ i = 1,2,\cdots,p$$

- Calculate the contribution rate and the cumulative contribution rate of the principal components:

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}, \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Generally the cumulative contribution rate is required to be above 80%.

RESULTS AND DISCUSSION

First, there is need to extract the principal component of the physiochemical indexes data of Wines (including red wine and white wine) and the Wine Grapes (including red grapes and white grapes). The contribution rate and the cumulative contribution rate are calculated and scatter plot of the contribution rate are below (Itamar *et al*., 1994).

Table 1: The contribution rate of each main component

Principal component	Contribution rate
f_1	45.7917
f_2	20.7300
f_3	14.9704
f_4	7.5678
f_5	5.6449
f_6	2.7243
f_7	1.2205
f_8	0.5880
f_9	0.3181
f_{10}	0.2456
f_{11}	0.1917
f_{12}	0.0070

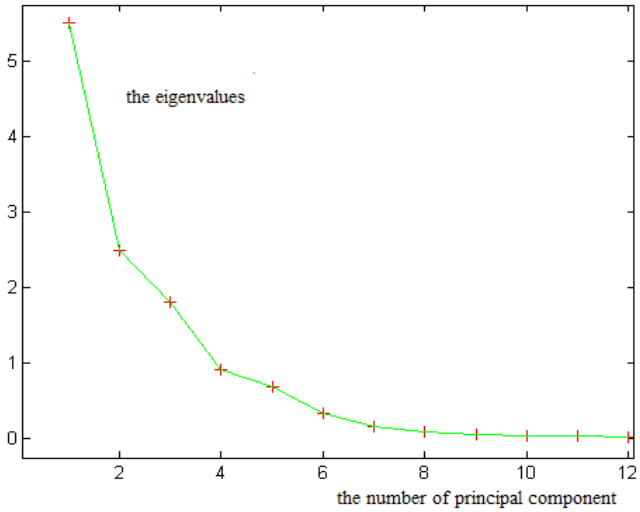

Fig. 1: The distribution of eigenvalues of red wine

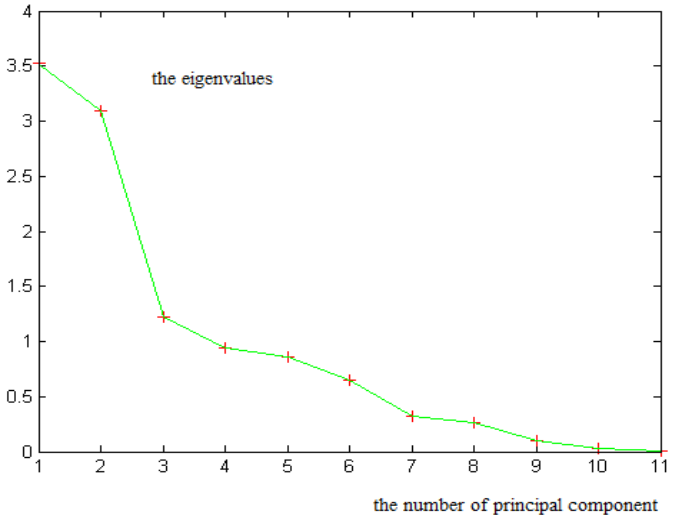

Fig. 2: The distribution of eigenvalues of white wine

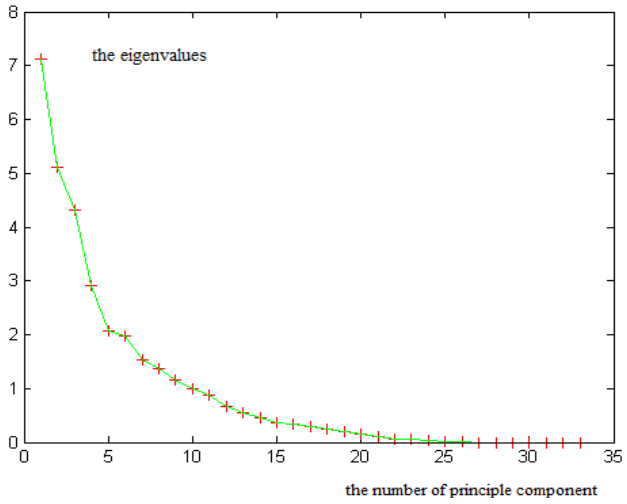

Fig. 3: The distribution of eigenvalues of red wine grapes

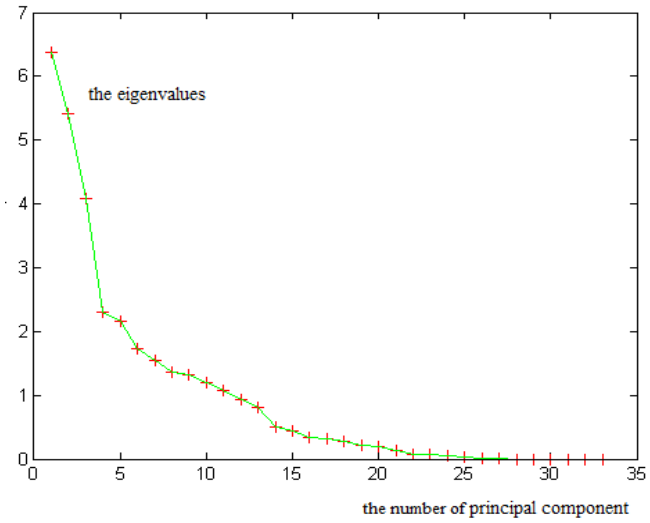

Fig. 4: The distribution of eigenvalues of white wine grapes

Due to the large amount of data, take the red wine for example. The contribution rate of each main component of red wine is in Table 1.

From Fig. 1 and Table 1, the operating results above show that: the contribution rate of first principal component is 45.7917% and the contribution rate of the second main components is 20.7300% and the contribution rate of the third main components is 14.9704%, so the cumulative contribution rate of the first three principal components is 81.4921%. Because the cumulative contribution rate is more than 80% (Zhou *et al*., 2010), the first three new factors are chosen.

From Fig. 2 about white wine, the cumulative contribution rate of the first three principal components is more than 80%, so the first three new factors are chosen.

From Fig. 3 about red wine grapes, the first nine new factors are chosen, because the eigenvalues of the first new nine factors are more than 1 and the others are less than 1.

From Fig. 4 about white wine grapes, the first twelve new factors are chosen for the same reason as Fig. 3.

For the red wine, from the three new factors, the main representatives of the variables extracted are: X_1 (anthocyanin), X_2 (tannin), X_3 (total phenols), X_4 (wine total flavonoids), X_6 (DPPH inhibition half volume), X_{10} (color H (D65)), X_{11} (color C (D65))) (Luo *et al*., 2000).

The main representatives of the variables extracted of the other three goals are in follow.

The main representative of variable about white wine: X_2 (tannin), X_3 (total phenols), X_5 (resveratrol), X_6 (DPPH inhibition half volume), X_9 (color, b* (D65)), X_{11} (color C (D65)).

The main representative of variable about white grapes: X_{11} (total phenols), X_{13} (grape total flavonoids), X_{18} (soluble solids), X_{21} (solid acid ratio), X_{22} (dry matter content), X_{29} (skin color a * (+red; -green)), X_{31} (skin color H).

The main representative of variable about red grapes: X_2 (tannin), X_4 (anthocyanin), X_5 (resveratrol), X_{16} (total sugar), X_{17} (sugar), X_{25} (stems ratio), X_{29} (skin color a * (+red; -green)), X_{32} (skin color C).

To analyze the relationship between red wine and red grapes among the main variables, Pearson correlation coefficient is calculated. Thus the correlation matrix shows: the anthocyanins and tannins of Wine Grape are significantly positively correlated to the anthocyanins and tannins in the Wines.

The results obtained through regression analysis are:

$$Y_1 = 0.9762X_4 - 0.2114X_{12} + 0.1436X_{14} + 0.1217X_{16} + 0.1158X_{17} + 0.0779X_{25} + 0.2974X_{29} - 0.4773X_{32} - 0.0583$$

R-square = 0.8512

$$Y_2 = 0.1917X_4 + 0.5243X_{12} + 0.1852X_{14} + 0.2529X_{16} + 0.2037X_{25} - 1.5924$$

R-square = 0.6679

$$Y_3 = 0.2576X_4 + 0.6315X_{12} + 0.2609X_{14} - 0.0292X_{17} + 0.1921$$

R-square = 0.8055

$$Y_4 = 0.2295X_4 + 0.7427X_{12} + 0.7427_{14} + 0.7257X_{32} + 0.4517$$

R-square = 0.7798

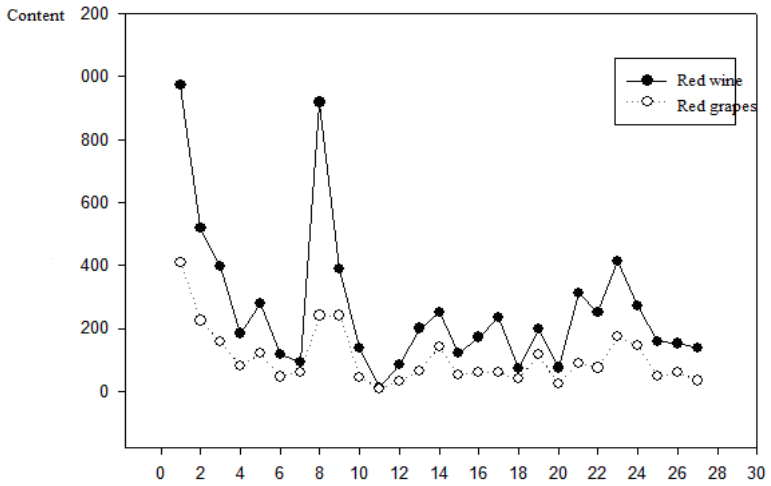

Fig. 5: The anthocyanin content of red grapes and red wine

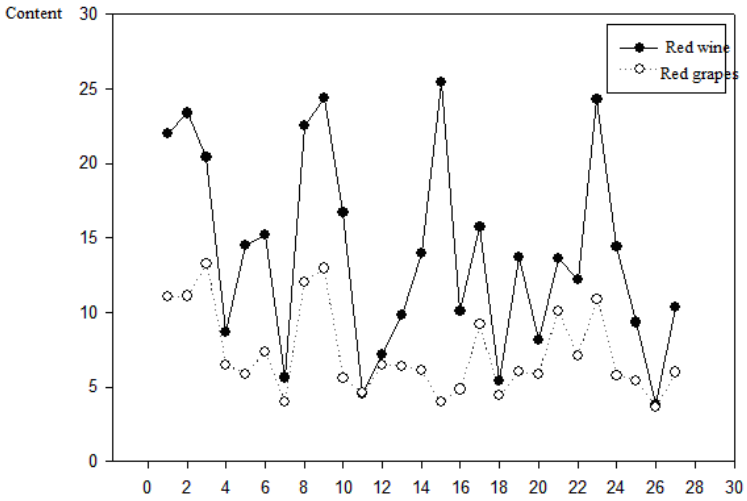

Fig. 6: The tannin content of red grapes and red wine

$$Y_5 = 0.7470X_{12} + 0.3582X_{14} - 0.8356X_{32} - 1.1652$$

R-square = 0.7646

$$Y_6 = 0.2732X_4 - 0.5152X_{12} + -0.5228X_{17} + -0.6604X_{29} - 6.9734$$

R-square = 0.3872

$$Y_7 = -0.7712X_4 - 0.7414X_{29} - 0.1224X_{32} - 5.3788$$

R-square = 0.5887

From the above regression equations established, the color (HD65) in physiochemical indexes of red wine has lesser extent related to the physiochemical indexes of wine grapes and only has 38.72% of the goodness of fit to anthocyanins, tannins, sugar and fruit color a *. Color (CD65) also shows the general goodness of fit and only has 58.87% of the goodness of fit to anthocyanins, skin color a * and skin color C. For the five other Physicochemical indexes of red wine, most showed strong correlation to the physiochemical indexes of the wine grape (Li *et al*., 2011).

Here are the scatter plot of anthocyanins and tannins for wine and wine grapes. From Fig. 5 and 6, the red wine is higher than red wine grapes between anthocyanin content and tannin content.

For anthocyanin content, from Fig. 5, anthocyanin content of red wine significantly higher than red grapes on the point 1 and 8 and the gap between red wine and red grapes is up to 600 or more. At other points, the gap between the anthocyanin content of red grapes basically fluctuates around 150 and the extent of fluctuations is not big.

For tannin content, from Fig. 6, the extent of gap fluctuations of red wine and red wine is large and the tannin content of red wine is significantly higher than red grape on eight points, namely point 1, 2, 3, 8, 9, 10, 14 and 22, respectively.

CONCLUSION

In this study, we want to establish contacts of the physiochemical indexes between the Wine grapes and the Wines. Because of a wide range of physiochemical indexes, in order to more clearly reflect the contact between Wine grapes and Wines, firstly, the principal component analysis is used to select principal components and the correlation matrix is established based on the corresponding variables of principal components. Take the red wine for example, the first three new factors are chosen, because the cumulative contribution rate of the first three principal components is 81.4921%, which is more than 80%. For the red wine, from the three new factors, the main representatives of the variables extracted are: anthocyanin, tannin, total phenols, wine total flavonoids, DPPH inhibition half volume, color H (D65) and color C (D65)). And the correlation matrix shows: the anthocyanins and tannins of Wine Grape are significantly positively correlated to the anthocyanins and tannins in the Wines.

And then, by stepwise regression method, the function of the relationship of physiochemical indexes between the Wines and Wine grapes is fitted, through which shows a strong correlation between the physiochemical indexes of Wines and Wine grapes.

ACKNOWLEDGMENT

The authors wish to thank the helpful comments and suggestions from my teachers and colleagues in Hebei United University.

REFERENCES

Chernev, A., 1997. The effect of common features on brand choice: Moderating role of attribute importance. J. Consum. Res., 23: 304-311.

Fang, K.T. and E.P. Pan, 1982. Cluster Analysis [M]. Geological Publishing House, Beijing, pp: 23-78.

Gao, X., 2004. Fuzzy Clustering Analysis and Application [M]. Xi'an University of Electronic Science and Technology Press, Xi'an, pp: 37-46.

He, X., 2004. Multivariate Statistical Analysis [M]. China Renmin University Press, Beijing, pp: 99.

Itamar, S., C. Ziv and O. Suzanne, 1994. Experimental evidence on the negative effect of product features and sales promotions on brand choice. Market. Sci., 13(1): 23-40.

Li, Y., M.D. Hou and H. Liu, 2011. Probability and Statistics [M]. Harbin Institute of Technology Press, Harbin, pp: 143-145.

Luo, J., T. Lu and Q. Wang, 2000. Comprehensive evaluation method of product design quality gray system [J]. Mech. Sci. Technol., 19(5): 747-749.

Yu, X., 1993. Multivariate Statistical Analysis [M]. China Statistics Press, Beijing, pp: 158-161.

Zhou, L., H. Wang and L. Du, 2010. A balanced relationship analysis between Chinese economic growth and the iron and steel production based on time series. Proceeding of the International Conference on E-Business and E- Government, pp: 3490-3493.