The Evaluation of Wine- summary

Leilei Zhao

School of Energy Power And Mechanical Engineering, North China Electric Power University, Hebei 071003, China;

2773754282@qq.com

Keywords: The Evaluation of Wine Summary；Student's t test; K-S test; factor analysis; Cluster analysis; partial least squares regression

Abstract. In order to obtain an evaluation system of wine, we use K-S test, t test, factor analysis and partial least squares regression analysis in the article, to grade the wine grapes. Then, we discuss the relationship among the wine grapes, the physical and chemical indicators and the qualities of the wine to finish the evaluation of the wine. First, we calculate the final scores given by the taster of each red and white wine sample. Then, we use K-S test to find that the two groups of data both obey the normal distribution. After using t-test based on paired data, we find that the two taster groups' scores are both under the significance level of 0.05, which means there exists significant differences. For the next step, we use SPSS to do reliability analysis. We get the reliability coefficient of the two group, 0.921 and 0.861 respectively, variance are 13.506 and 12.589. Finally, after our comprehensive consideration, we think that the first group's score for the red wine sample is more reliable, and the second group's score for the white wine sample is more reliable.

1. Introduction

1.1 Problem Background

The way of determining the quality of wine is generally by hiring a group of qualified member, to taste the wine and give out their conclusions. After the wine tasting, the taster scores on each indicator, then we get the total grade, so as to determine the quality of the wine. The quality of wine grape has a direct relationship between the quality of the wine. And the physical and chemical indicators of wine and wine grape will to some extent reflect the quality of the wine and grapes. So in this problem, we have to complete two tasks:

- Find out which group's scores of the wine is more reliable;
- Develop an evaluation system for the wine.

2. Symbols, Definitions and Assumptions

2.1 Symbols and definitions

Table 1 Symbols and definitions

NO.	Symbols	Definitions
1	$h1_i$	The first groups' score on the i sample of the red wine
2	$h2_i$	The second groups' score on the i sample of the red wine
3	$b1_i$	The first groups' score on the i sample of the white wine
4	$b2_i$	The second groups' score on the i sample of the red wine
5	D_i	The two groups' score difference between the i sample of the red wine
6	a_{ij}	The i indicator of the j red wine sample
7	x_i	The ith indicator of the wine grape

2.2 General assumptions

- The scores from the professionals are fair and reliable. We can directly judge the qualities of wines from the scores.

- The main factor and the corresponding principal components selected by factor analysis can represent all original composition.
- All the data we collect online is accurate and reliable.

3. Task 1: Find the More Reliable Professional Group

3.1 Preprocess the Data

(1) The preprocess of missing data

For the phenomenon of the data missing, we use average substitution method. Because of the small difference among the scores on the same sample, we think it feasible using the average to replace the missing data, which can be calculated as follows:

$$h_{im} = \frac{1}{9}\left[\sum_{k=1,k\neq m}^{10} h_{ik}\right] (m = 1,2,\cdots,10) \quad (1)$$

-Where h_m represents the missing data.a

3.2 Find the Probability Distribution of the Scores

Different professional has a total score for each wine samples. In order to eliminate individual factors, we take the strategy of "take out a maximum value, remove a minimum, then calculate the average" to deal with the given points from the four tables, the calculation can be listed as follows:

$$\bar{p}_{kj} = \frac{\sum_{i=2}^{9} p_{ij}}{8} (i = 2,3,\cdots 9 \quad j = 1,2,\cdots 27 \quad k = 1,2,3,4) \quad (2)$$

Table 2 One-sample test kolmogorov-smirnov

		White 1	White 2	White difference	Red 1	Red 2	Red difference
N		28	28	28	27	27	27
Normal Parameters	Means	74.9688	77.3036	-2.3348	72.9833	70.6435	2.3398
	Standard deviation	4.72638	3.40894	5.13938	7.44885	3.98445	5.43159
The most extreme difference	Absolute value	.092	.135	.068	.162	.116	.226
	positive	.066	.069	.068	.108	.090	.160
	negative	-.092	-.135	-.064	-.162	-.116	-.226
Kolmogorov-Smimov Z		.488	.716	.361	.841	.602	1.176
Asymptotic significance(Bilateral)		.971	.684	.999	.479	.862	.126

a. Normal distribution test

b. According to the data obtained calculate

From the p-p figure, we know that both red and white wine data obey the normal distribution.

3.3 Judge the Significant Difference

From the conclusion we drew above, we use t-test to judge whether there exists significant difference between the paired data.

First, we do hypothesis test for the red wine. We

let $D_1 = h1_1 - h2_1, D_2 = h1_2 - h2_2$, …, $D_{27} = h1_{27} - h2_{27}$,the $D_1, D_2, \ldots, D_{27}$ are independent of each other, and they all obey the normal distribution .that means $D_1, D_2, \ldots, D_{27}$ from $N(\mu_D, {\sigma_D}^2)$ one sample.

Based on this hypothesis test sample:

$$H_0: \mu_D = 0, \quad H_1: \mu_D \neq 0$$

We make $D_1, D_2, \ldots, D_{27}$ the sample mean and variance of observations $\bar{d}$, s^2,the we can get the refuse domain:

$$|t| = \left|\frac{\bar{d}}{s_D/\sqrt{27}}\right| \geq t_{\alpha/2}(26) \quad (3)$$

The significant level α is 0.05, From the table, we find that $|t| \geq 1.7056$ .Then we get $\bar{d}=2.3398$, s_D =5.43159，$|t| \quad \frac{2.3398}{5.4398/\sqrt{2}} \geq 1.7056$

The result falls in the refused domain, so we accept H_1 ,thinking that the data of red wine exists significant difference. In the same way, we do t-test to the data of white wine,α=0.05， refused domain is $|t|$ 1.7033.Then we get $\bar{d}$ =-2.3348，s_D =5.13938，$|t|$=2,4039>1.7033,so we think that the data of white wine exists significant difference as well.

4 Task2 : Reliability Analysis

After the judgement of the significant difference, we can come to the conclusion that the two groups' scores on red and white wine both exist significant difference. As a result, it becomes important to find out which group's score is more reliable.

We import the two groups' scores on red and white wine into SPSS, analyse their reliability coefficients respectively. The results are showed as follow.

Reliability Statistics

Cronbach's Alpha	Based on standardized items Cronbach's Alpha	Number of items
.921	.921	10

Fig.1 Reliability Analysis of the first group of red wine Reliability Statistics

Cronbach's Alpha	Based on standardized items Cronbach's Alpha	Number of items
.861	.867	10

Fig.2 Reliability analysis of the second group of red wine Summary Statistics

	The mean	Minimum	Maximum	range	Maximum/Minimu	variance	Number of items
Each of The mean	73.056	65.889	79.185	13.296	1.202	13.056	10

Fig.3 The first group of red grapes Stability Analysis Summary Statistics

	The mean	Minimum	Maximum	range	Maximum/Minimum	variance	Number of items
Each of The mean	70.515	64.259	76.704	12.444	1.194	12.596	10

Fig.4 Stability analysis of the second group of red grapes Reliability Statistics

Cronbach's Alpha	Based on standardized items Cronbach's Alpha	Number of items
.771	.792	10

Fig.5 Reliability Analysis of the first group of white wine Reliability Statistics

Cronbach's Alpha	Based on standardized items Cronbach's Alpha	Number of items
.910	.945	10

Fig.6 Reliability analysis of the second group of white wine

From the data showed in figure 4 and 5, the reliability coefficients of red wine of the first and the second group is 0.912 and 0.861. So we find that the first group's score on red wine is more reliable. Figure 6 and 7 show that the variances are 13.056 and 12.598 respectively. Although the variance of the 2th group is smaller, the difference between the two groups of data is small as well. So, after our comprehensive consideration, we come to the conclusion that the first group's score on red wine is more reliable. In the same way, we find that the second group's score on white wine is more reliable.

References

[1]. Argyrous, G. Statistics for Research: With a Guide to SPSS. London: SAGE. ISBN 1-4129-1948-7.