Study and Identification of Ancient Glass Composition Based on Regression Equation and Statistical Analysis

Pu Zhang *

School of Anhui University of Science and Technology, Huainan, China

* Corresponding Author Email: 1746361415@qq.com

Abstract. Glass is valuable material evidence of the early trade along the ancient Silk Road, but the ancient glass is easily weathered by the buried environment, which leads to the change of its composition proportion, thus affecting the correct judgment of its classification. The study of the composition analysis and identification of ancient glass products is of great help to the understanding of the social culture at that time and the trade civilization between China and foreign countries. This paper mainly studies the composition analysis and identification of ancient glass products. Firstly, the chi-square test is used to analyze the correlation, the regression analysis model is established to complete the significance test, and dynamic clustering is used to classify the glass. It is concluded that high-potassium glass and lead-barium glass can be divided into three categories. The rationality and sensitivity of the classification results are analyzed by a decision tree. Finally, based on the least square method and dynamic clustering, the results show that: The comprehensive model used in this paper can accurately analyze the composition of ancient glass.

Keywords: Chi-square test, regression analysis model, decision tree, dynamic clustering, correlation analysis

1. Introduction

The Silk Road was a channel for cultural exchanges between China and the West in ancient times, in which glass was valuable material evidence of early trade exchanges. However, the main chemical composition of ancient glass is also different, and ancient glass is easily weathered by the influence of the burial environment, which affects the correct judgment of its classification, so it is necessary to propose a method to analyze and identify the composition of glass products.

In this paper, the correlation analysis was carried out by the chi-square test, and the regression equation was established for the significance test. Finally, the statistical law of weathering chemical composition content on the surface of cultural relics samples was analyzed and predicted according to the established regression equation.

Then the dynamic cluster analysis was carried out to further obtain that high potassium glass and lead-barium glass can be divided into three categories. We use SPSS to analyze the rationality and sensitivity of the classification results.

Finally, a correlation analysis model is established by using Python, which transforms the correlation between the chemical components of different types of glass cultural relics into the correlation between the chemical components. Through the establishment of a comprehensive evaluation model, the composition of glass products was analyzed and identified.

2. Model establishment and solution

Through data processing [1], we establish a correlation analysis model to analyze the relationship between weathering and possible impact factors. Calculate according to the data statistics of the independent variable and the dependent variable, establish a regression analysis equation on this basis, that is, a regression analysis prediction model, then carry out correlation analysis, calculate the predicted value by using the regression prediction model, and comprehensively analyze the predicted value to determine the final predicted value.

2.1. Chi-square test

The problem requires analyzing the relationship between the surface weathering of these glass relics and their glass type, decoration and colour. First of all, we should understand that the indicators we need to analyze are all classified variables, so we use the chi-square test here, that is, the deviation between the actual observed value of the statistical sample and the theoretical inferred value. The deviation degree between the actual observed value and the theoretical inferred value determines the size of the chi-square value, if the chi-square value is larger, the deviation degree between the two is larger; The smaller the deviation between the two values; if the two values are exactly equal, the chi-square value is 0, indicating that the theoretical value is completely consistent. By default, a progressive significance result of less than 0.05 is considered to be a certain correlation between rows and columns.

The chi-square [2] test on the imported data using SPSS shows that the progressive significance result is 0.009. Since 0.009 < 0.05, we believe that the chi-square test result rejects the original hypothesis (original hypothesis: rows and columns are not correlated), that is to say, rows and columns are correlated. Through the symmetry measurement test, it can be seen that the contingency coefficient is 0.326, and it is considered that the surface weathering is generally related to the type.

2.2. Principal component analysis

Through the given data [3], we can see that the glass contains more chemical components, which will lead it more difficult to deal with. Through further observation, we can find that some chemical components are interrelated and correlated. Therefore, we consider explaining the internal structure of variables through a few principal components, that is, deriving a few principal components from the original variables. So that they retain as much information as possible about the original variables and are independent of each other.The results of principal component analysis[4] by SPSS are as follows,which is shown in Figure 1 and 2.

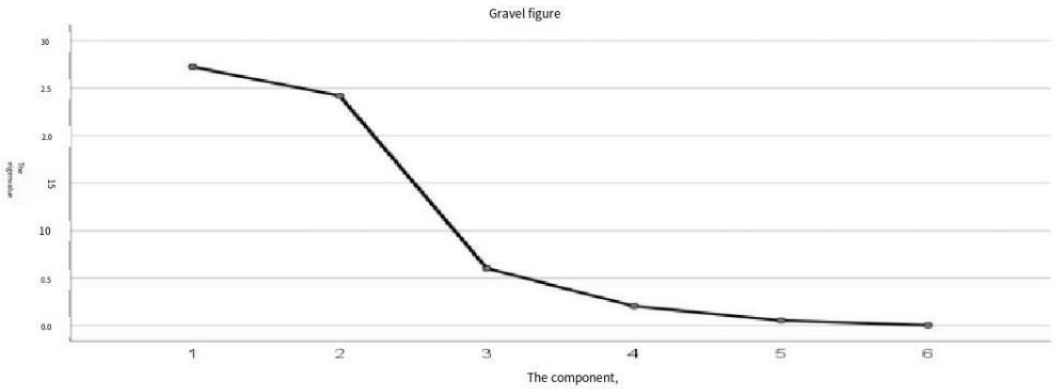

Figure 1. Gravel diagram of the indicators of the six items

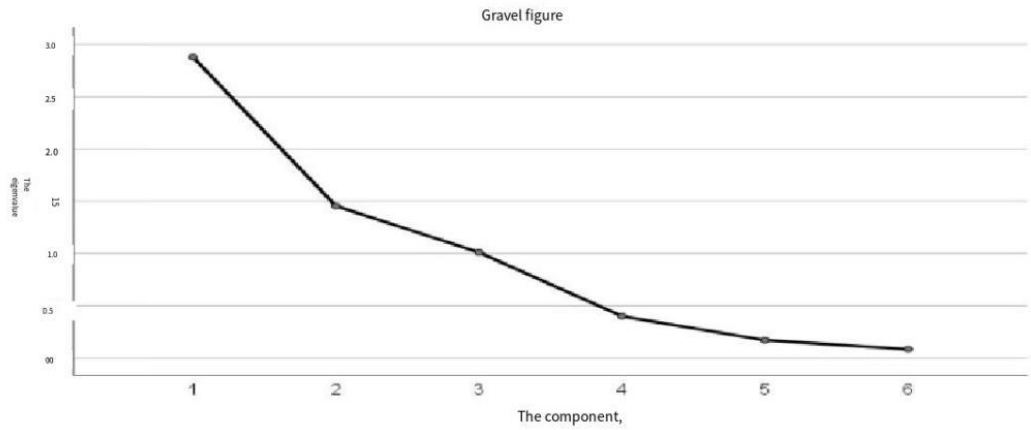

Figure 2. Gravel diagram of the indicators of the last six items

From the figure, we can see that there is an obvious inflexion point at component 2. To sum up, we regard components 1, components 2, 3, 4, 5 and 6 as two principal components respectively.

Similarly, components 7 and 8 are taken as one principal component, and components 9, 10, 11, 12, and 13 are taken as one principal component.

2.3. Establishment of regression analysis model

We respectively establish a linear regression model [5] of each index, which is recorded as:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

$$E\varepsilon = 0, D\varepsilon = \sigma^2 \tag{2}$$

The fixed location parameter is called the regression coefficient, and the independent variable X also becomes the regressor.

The parameter estimator can be written as:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_0 &= \hat{Y} - \hat{\beta}_1 \overline{X} \end{aligned} \tag{3}$$

We will get the results of the significance test of the regression equation, here we use the F test method to test when it is true H_0.

$$F = \frac{U}{Q_0 / (n-2)} \sim F(1, n-2) \tag{4}$$

Where (regression sum of squares) $U = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

2.4. Solution of Regression Analysis Model

The obtained regression equation can predict the composition before weathering, as shown in the figure 3:

Cultural relics number	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
8	35.26	0	0.58	0.21	0	4.45	1.51	4.78	18.6	23.55	5.75	0.1	0	0
23	59.28	2.66	0.11	0.54	0.74	6.1	0.08	0.53	15.99	10.96	0	0.53	0	0
25	56.54	3.66	0.31	0.98	0.61	3.06	0	0.59	24.4	8.23	0.1	0.85	0.12	0
26	65.81	0	0	1.71	0.9	3.11	4.59	0.4	16.43	3.42	1.58	0.31	0	0
28	56.01	2.44	0.21	0.57	1.06	4.28	0	0.65	22.21	10.08	0.1	0.5	0	0
29	58.52	1.89	1.06	0.72	1.02	3.97	0.02	0.46	20.2	10.06	0	0.76	0	0
42	51.54	4.66	0.29	0.87	0.61	3.06	0	0.65	25.4	9.23	0.1	0.85	0	0
43	30.62	0	0	1.61	0.6	1.65	0.2	0.79	42.25	15.45	2.54	0	0	0
44	37.36	0	0.71	0	0	5.45	1.51	4.78	9.3	23.55	5.75	0	0	0
49	60.12	0	0.23	0.89	0	2.72	0	3.01	17.24	10.34	1.46	0.31	0	3.66
50	62.24	0	0.26	0.86	0	2.59	0	3.11	17.12	10.33	1.46	0.28	0	0
51	75.51	0	0.15	0.64	1	2.35	0	0.47	16.16	3.55	0.13	0	0	0
53	61.28	2.66	0.11	0.84	0.74	5	0	0.53	15.99	10.96	0	0.23	0	0
54	54.91	3.62	0.16	0.58	0.58	3.44	0.17	0.56	21.05	6.68	0.42	0.23	0	0

Figure 3. Plot of prediction results

3. Classification model based on K-clustering and decision tree

The decision tree is a good mathematical model [6] to predict the classification of high-potassium glass and lead-barium glass, and dynamic cluster analysis is also necessary. The decision tree model is a tree structure. The classification problem, represents the process of classifying instances based on features. Dynamic cluster analysis provides a scientific and simple method for the classification of ancient glasses. We take the content of oxide in ancient glass as a variable [7], and take each ancient glass sample as an event for cluster analysis, which helps to eliminate subjective errors, and also can cluster the minor elements and major elements separately. Therefore, different series of ancient glass can also be compared in all aspects.

3.1. Decision Tree

The decision tree is a tree structure, which is suitable for the classification results of research objects. It starts from the root node, tests the data samples, and divides the data samples into different data sample subsets according to different results. Each data sample subset constitutes a sub-node. It is through the establishment of a model to study the attributes and characteristics of the object, to determine which classification group the object is most likely to fall into.

By establishing a decision tree model [8], the decision tree structure is shown in Figure 4.

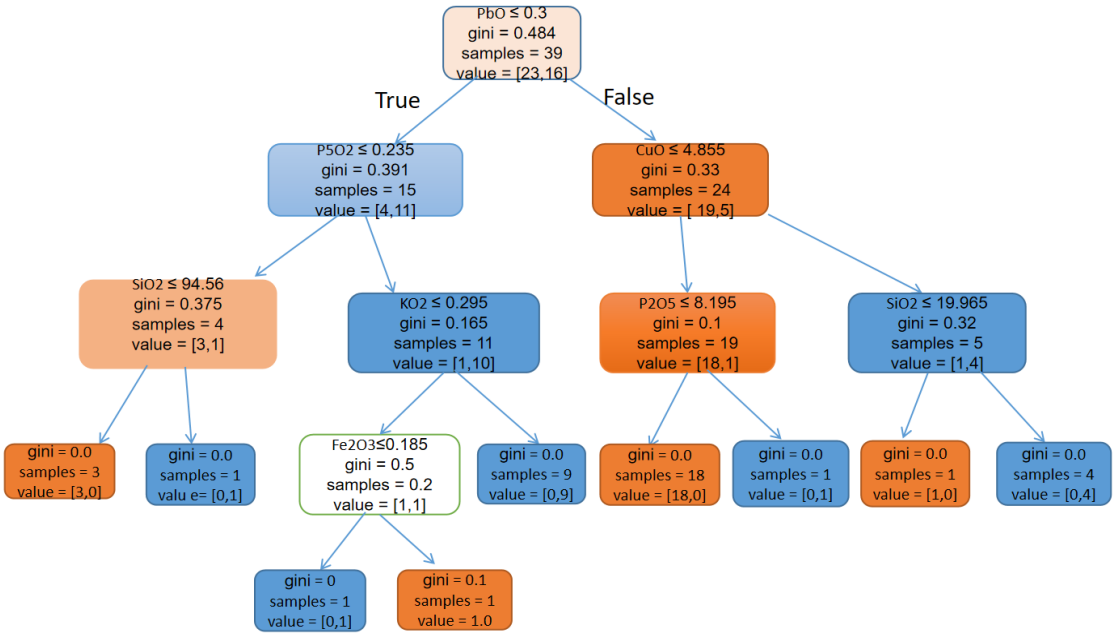

Figure 4. Decision Tree

In the decision tree structure shown in the figure above, the internal nodes give the specific segmentation of the branched features. The Gini is used to determine which feature to segment, and the sample type distribution is the number of samples belonging to each classification group in the node. The decision tree constructed above is mainly based on the chemical composition of the cultural relics.

Table 1. Results of the model evaluation

	Accuracy	Recall rate	Accuracy	F1
Training set	1	1	1	1
Test set	0.882	0.882	1	0.938

Result analysis: According to the results of the model evaluation table 1 , the accuracy of the training set is 1, and the accuracy of the test set is 0.882, both of which are better. The recall of the training set is 1, and the recall of the test set is 0. 882, both of which are better. The accuracy of the

training set is 1, the accuracy of the test set is 1, and the accuracy of both is very high. The F1 of the training set is 1, and the F1 of the test set is 1, so to sum up, the evaluation result of the model is good.

The basic steps of k-means clustering are Figure 5, The elbow method is used to select the value of clustering number K, which is shown in Figure 6.

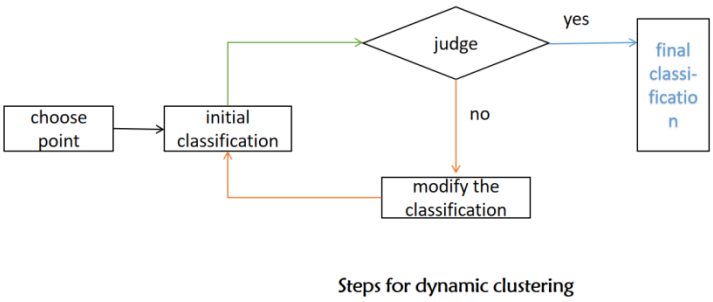

Figure 5. Flow chart of dynamic clustering

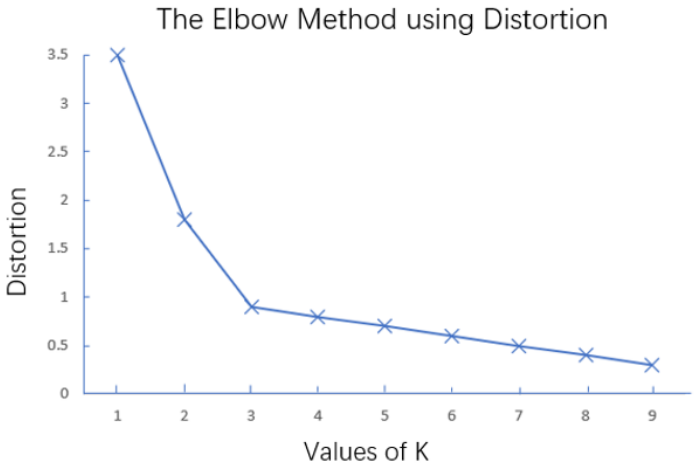

Figure 6. Selection of K value

As shown in the above figure, according to the elbow rule, we can divide the 56 numbered cultural relics into three categories, that is, K takes 3.

It is concluded that the cultural relics numbered 8, 11, 18, 21, 22, 23, 24, 25, 27, 30, 31, 32, 34, 36, 41, 43, 44, 45, 46, 47, 48, 50, 53 and 55 are grouped together. Objects numbered 2, 16, 29, 33, 35, 37, 38, 39, 40, 42, 49, 51, 52, 54, 56, 57, 58 are grouped together and objects numbered 1, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 19, 20, 26, 28 are grouped together. which is shown in Figure 7.

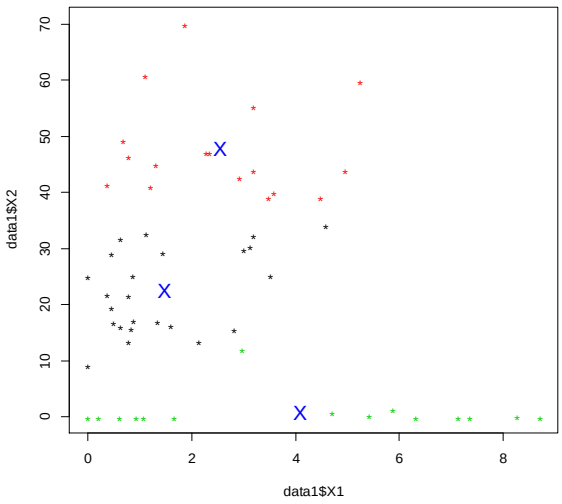

Figure 7. Classification and central point output diagram

The sensitivity analysis of the Pb-Ba glass subtype classification shows that the data validation we have obtained does not fluctuate too much.

4. Correlation analysis model

4.1. Model establishment and solution

Through data processing [9], we establish the correlation analysis model and difference analysis, analyze different types of glass cultural relics samples based on the correlation analysis model, and compare the chemical composition correlation between different types by difference analysis.

Where the sample size is n, where there are p observations for each sample size value, n = 14 in this problem, p = 16 in the high potassium data, and p = 40 in the lead-barium data.

Calculate the simple correlation between two samples separately, because the correlation coefficient between each variable and itself is 1, that is: r_{ij} .

$$\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} = \left(r_{ij}\right)_{p\times p} \tag{5}$$

Where is the simple correlation coefficient of the two variables. $\left(r_{ij}\right)_{p\times p}$

$$\left(r_{ij}\right)_{p\times p} = \frac{\sum (x_i - \bar{x})(x_j - \bar{x})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (x_j - \bar{x})^2}} \tag{6}$$

Use Python to import the data into the above model for correlation analysis, and the results are shown in the following figure 8:

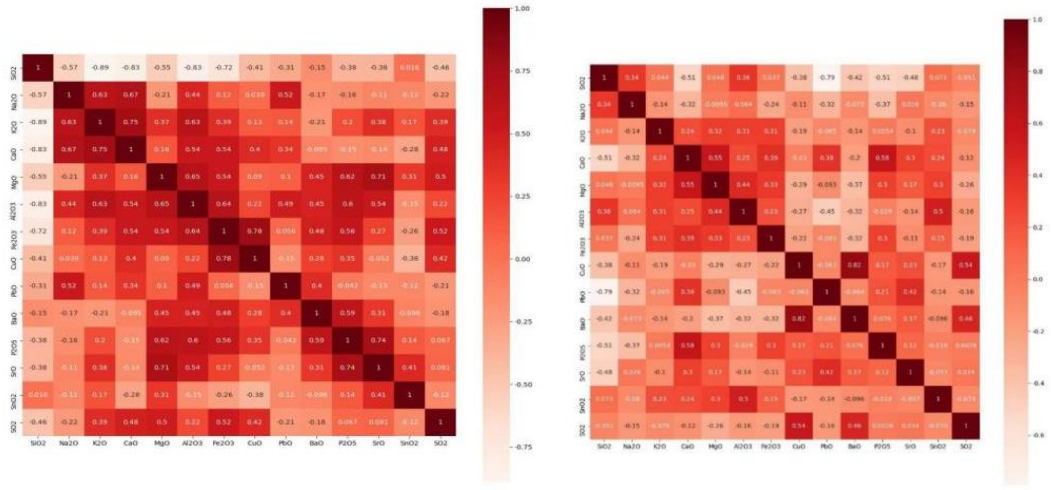

Figure 8. Correlation coefficient thermogram of high potassium and lead-barium glass

It can be seen from the correlation coefficient thermogram of high-potassium glass that SiO_2 is mainly negatively correlated with other chemical components, among which the negative correlation with K_2O is the strongest, and its negative correlation coefficient is -0.89. Na_2O is mainly positively correlated or not correlated with other chemical components, among which the positive correlation with CaO is good, and its positive correlation coefficient is 0.67; There is a strong negative correlation between CaO and SiO_2, and the negative correlation coefficient is -0. MgO and SrO have a good positive correlation, the positive correlation coefficient is 0.71; Al_2O_3 and SiO_2 have a strong negative correlation, the negative correlation coefficient is -0.83; Fe_2O_3 and CuO have a good positive

correlation, and the positive correlation coefficient is 0.78; PbO and Na_2O have a positive correlation with a positive coefficient of 0.52; There is a positive correlation between BaO and P_2O_5, and the positive correlation coefficient is 0.59; P_2O_5 and SrO have a good positive correlation, and the positive correlation coefficient is 0. 74; SnO_2 and other chemical components have a poor correlation; SO_2 and Fe_2O_3 correlate, and the correlation coefficient is 0. 52.

According to the thermodynamic diagram[10] of the correlation coefficient of Pb-Ba glass, it can be seen that there is a good negative correlation between SiO_2 and PbO, the negative correlation coefficient is -0.79; there is a positive correlation between CaO and P_2O_5, the positive correlation coefficient is 0.58; there is a correlation between CuO and SO_2, the correlation coefficient is 0.54; BaO is correlated with SO_2, and the correlation coefficient is 0.46; SrO is correlated with SiO_2, and the correlation coefficient is -0.48; Na_2O, K_2O, MgO, Al_2O_3, Fe_2O_3 and SnO_2 are poorly correlated with other chemical components;

4.2. Difference analysis

The problem requires to analyze the difference in the chemical composition relationship between different types of glass cultural relics. Through the data, we can intuitively see the data to be analyzed. The grouping variable is the classification variable, and the independent variable is the quantitative data. On this basis, the difference analysis is established, and The results are analyzed in Figure 9.

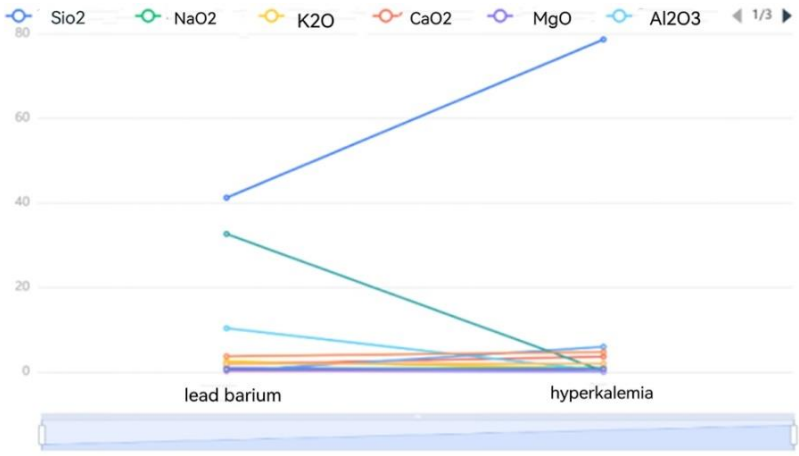

Figure 9. ANOVA comparison chart

As can be seen from the table 2, for the variable silica (SiO2), the order of magnitude of the means is High Potassium > Lead Barium. There is a significant difference between Pb-Ba and high potassium. For the variable sodium oxide (Na2O), the order of the mean values was Pb-Ba > high-K. There was no significant difference in the overall mean. For the variable potassium oxide (K2O), the order of mean values is: high potassium > lead and barium. There is a significant difference between Pb-Ba and high potassium.

Table 2. Post hoc multiple comparison results

	name	Name	Average	Average	Difference	P
Silica (SiO2)	Lead and barium	High potassium	41.213	78.629	-37.416	0.000***
Sodium oxide (Na2O)	Lead and barium	High potassium	0.966	0.521	0.445	0.368
Potassium oxide (K2O)	Lead and barium	High potassium	0.186	5.949	-5.762	0.000***

5. Conclusion

Through the analysis and evaluation of the surface weathering of glass cultural relics and the comprehensive relationship between glass type, ornamentation and colour. It can be seen that the

model in this paper has strong regularity and practicability, and can be extended to other issues, for example, it can be used to study historical fields such as the time and place of production, the source of raw materials, the production process and the social and cultural sacred flame at that time. The model established in this paper has high stability, which can solve the research problems of ancient glass relics and even the foreword problems of modern glass to a certain extent. If the model is improved, it can get better results, which is of great help to understand the social culture at that time and the trade civilization between China and foreign countries, and has a strong practical significance.

References

[1] Yu Fuxi. The Development of Glass Science and Technology in Ancient China. Shanghai: Shanghai Science and Technology Press, 2005.

[2] Huang Qishan. A study of ancient glass in Guangxi. See: Yu Fuxi, chief editor. A Study of Ancient Glass in South China: Proceedings of the 2002 Nanning Symposium on Ancient Glass in South China [C]. Shanghai: Shanghai Science and Technology Press, 2003:1020.

[3] Marne. Bose-like peaks in strained glasses and the physical unity of glasses [J]. Chinese Science Bulletin 2021 (36).

[4] Lilan Zheng. Research on cargo location allocation in steel logistics park based on improved multi-objective optimization algorithm [D]. Wuhan University of Science and Technology. 2021.

[5] Zheng Mengyao, Zhang Xiaowen, Xiang Ning, Ding Yao, Yan Yue. Study on the interfacial bonding properties of organic laminated glass. Laboratory Research and Exploration. 2021 (11).

[6] Zhao Tongyao, Ren Jinjun. Review on the application of solid-state NMR in the structural analysis of rare earth doped oxyfluoride glasses and glass-ceramics. Journal of Silicate Science. 2021 (12).

[7] Liu Yaru, Tian Yingliang, Li Junjie. Analysis of factors affecting the accuracy of surface tension measurement of glass melt by sessile drop method. Glass.2021 (12).

[8] Lu Lipeng. Glassware unearthed from Zagunluk and Shanpula tombs [J]. Turpan Studies, 2004 (2): 127 — 137.

[9] Zhang Jianlin. Scientific analysis and research on the glass unearthed from Emperor Wudi Xiao Mausoleum of the Northern Zhou Dynasty [J]. Archaeology and Cultural Relics, 2011 (1): 107 — 112.

[10] Yu Chun, Wen Shuang, Huang Dan. Variable-cluster effect test and variable selection in mixed regression model [J]. Statistics and Decision Making. 2021 (24).