WEP
Warwick
Evans
Publishing

Transactions on Materials, Biotechnology and Life Sciences
ISSN: 2960-1932, eISSN: 2960-2319 | Volume 4
NSCBM 2024

This paper analyzes and identifies the composition of ancient glass products by using data analysis method

Runhe Tang [1, *, #], Jingyi Zhou [2, #]

[1] School of Mathematics and Statistics, Lanzhou University, Lanzhou, China, 730000

[2] Cuiying Honors College, Lanzhou University, Lanzhou, China, 730000

* Corresponding Author Email: tangrh21@lzu.edu.cn

[#]These authors contributed equally.

Abstract. Glass is the precious material evidence of the ancient Silk Road trade. China's ancient glass and foreign glass, although similar in appearance, but the chemical composition is different. Ancient glass is susceptible to environmental influences and thus weathering, and the weathering process will affect the change of its chemical composition, so it is necessary to study the relationship between weathering and some characteristics of glass. First of all, this paper uses the method of Chi-square test to analyze the relationship between the surface weathering of cultural relics and the type, color and decoration of glass, so that there is a greater relationship between the surface weathering and the type. Secondly, the glass is divided into two types: lead barium and high potassium, and the bar chart of each chemical component content of weathered glass and unweathered glass is made. Next, the paper predicts their contents before weathering, and then compares them with the contents of other types and unweathered cultural relics with the same color and pattern to test the prediction results of this paper. Then, this paper uses the K-Means ++ algorithm to classify the unclassified data in Table 2. By analyzing the classification rules obtained by clustering, the classification standards for high-potassium glass and lead-barium glass are approximately obtained. Next, the same clustering algorithm is used to divide the data into high potassium glass group and lead barium glass group, and cluster the two groups of data respectively. Through the analysis of the clustering center law, the classification standard of the subclass is obtained. Finally, the correlation analysis method was used to conduct correlation analysis between the content of main compounds in the subclass classification law of each cultural relic and the distance between each cultural relic and the final clustering center, and the fluctuation of 5% was processed for the content of important compounds, and the correlation analysis was carried out again to test the change of correlation to achieve the purpose of testing sensitivity.

Keywords: Chi Square Test, Classification Summary, K-Means ++ Clustering, Dbscan Clustering.

1. Introduction

Glass is valuable physical evidence of trade on the ancient Silk Road. Although the appearance of ancient Chinese glass is similar to that of foreign glass, the chemical composition is different. Ancient glass is susceptible to environmental influences and thus weathering, and the weathering process will affect the change of its chemical composition[1-3].

Zhang Qian et al. proposed to analyze the composition of ancient glass products based on the random forest model, and find out the causes of glass surface weathering by analyzing the correlation between external factors such as glass type and surface weathering and the correlation between glass composition and surface weathering. Zhang Shuaifan et al. used gray correlation analysis to study the correlation between the chemical components of ancient glass, and identified the differences between the chemical components of glass with the overall average value and k-mean cluster analysis. Now there is a batch of Chinese ancient glass products data, the data comes from, these glasses can be divided into lead barium and high potassium two types.

In this paper, the Chi square test and K-Means ++ algorithm are used to classify and cluster the glass. After that, the method of correlation analysis is used to test the glass by observing the change of correlation. In the end, we have determined the relationship between the weathering of ancient glass and its composition content

2. The establishment and solution of the model

2.1. Establishment and solution of Chi-square test model

For the difference analysis between certain variables and certain variables, this paper adopts the Chi-square test. The Chi-square test is the deviation degree between the actual observed value and the theoretical inferred value of the statistical sample. The deviation degree between the actual observed value and the theoretical inferred value determines the size of the chi-square value.

The degree of deviation between the two is smaller; If the two values are exactly equal, the chi-square value is zero, and the surface theoretical value is exactly the same.

Chi-square test steps: The chi-2 square value is used to indicate the degree of deviation between the observed value and the theoretical value. Calculate this deviation.

The basic idea of degree is as follows:

(1) Let A represent the observed frequency of A certain category, and E represent the expected frequency calculated based on H0, the difference between A and E.

This is called the residual.

(2) Obviously, residuals can represent the degree of deviation between observed and theoretical values for a certain category, but if the residuals are simple.

In contrast, there are certain shortcomings in expressing the difference between the observed frequency and the expected frequency of each category. Because the residual error has

Positives have negatives, which cancel each other out when added, and the sum remains 0, so you can square and sum the residuals to do this.

(3) On the other hand, the residual size is a relative concept, relative to the expected frequency when the expected frequency is 10.

The residuals of 20 are very large, but the residuals of 20 are small relative to the expected frequency of 1000. With this in mind, one then divides the square of the residual by the expected frequency and sums it to estimate the difference between the observed frequency and the expected frequency.

After the above operation, we get the commonly used chi-2 square statistic, which is also called pearson Chi-2square because it was first proposed by the British statistician Karl Pearson in 1900, and its calculation formula is:

$$\chi^2 = \sum \frac{(A-E)^2}{E} = \sum^{k} \frac{(A_i - E_i)^2}{E_i} = \sum^{k} \frac{(A_i - np_i)^2}{np_i} \tag{1}$$

Among them, Ai is the observation frequency at the i level, Ei is the expected frequency at level i, N is the total frequency, Pi is the expected frequency at level i. The expected frequency Ei at level i is equal to the total frequency n x the expected probability pi at level i, K is the number of cells. When n is relatively large, The chi square distribution of k-1 (the number of parameters used to calculate Ei) degrees of freedom approximately follows the chi square distribution of the chi square statistic.

The original assumption of this paper is that there is no significant difference between ornamentation, type, color and surface weathering.

In order to intuitively observe the relationship between ornamentation, type, color and weathering, the correlation bar chart is made in this paper(See appendix).

The following is the chi-square inspection table of decoration, type and color in order[4-5].(As shown in table 1-3)(Data sourced from (mcm.edu.cn))

Table 1. Chi-square test results of the decoration.

	Chi-square test ornamentation		
	value	freedom	Progressive significance (bilateral)
Pearson chi-fang	5.304	2	0.071
likelihood ratio	7.371	2	0.025
Number of valid cases	69	0	0

Table 2. Chi-square test results of the types.

Chi-square test type			
	Value	freedom	Progressive significance (bilateral)
Pearson chi-fang	11.268	1	0.001
Continuity correction "	9.516	1	0.001
likelihood ratio	11.236	1	0.001
Fisher's exact test	0	0	0
Number of valid cases	69	0	0

Table 3. Chi-square test results of the color.

Chi-square test color			
	value	freedom	Progressive significance (bilateral)
Pearson chi-fang	13.233	7	0.067
likelihood ratio	17.288	7	0.016
Number of valid cases	69	0	0

According to the results of Chi-square test, this paper gets the conclusion:

(1) Based on surface weathering and ornamentation, the P-value of significance is 0.071, and there is no significance at the level. The null hypothesis is accepted.

Therefore, there is no significant difference between surface weathering and ornamentation.

(2) Based on surface weathering and type, the P-value of significance is 0.001, showing significance horizontally, rejecting null hypothesis.

Therefore, there is a significant difference between surface weathering and type.

(3) Based on surface weathering and color, the P-value of significance is 0.067, and there is no significance at the level. The null hypothesis is accepted.

Therefore, there is no significant difference between surface weathering and type.

2.1.1. Glass type classification process

First of all, the total content of the two types of glass is analyzed in this paper, and the following table is obtained:

After the analysis of the table, it is found that for the glass with high potassium, the difference between the change of total content after weathering and before weathering is not large, and the same is true for the glass with lead barium, and the total content of both kinds of glass increases slightly. Then, the content of some chemical components of the glass was analyzed in this paper. (As shown in Table 4).

Among them, sodium oxide, tin oxide and other components are excluded in this paper due to the lack of data. After data elimination, this paper uses a classification and summary algorithm for them.

Classification summary algorithm: classification summary is the data after the data, first in accordance with a standard classification, and then on the basis of the classification of various types of relevant data were summed, average, number, maximum, minimum and other methods of summary, this paper only the mean of the summary.(As shown in Figure 1-3) (Data sourced from (mcm.edu.cn))

Table 4. Table of total content ranges for different types of glass.

Glass type	Whether weathering or not	engraved pattern	colour	Total content range
High potassium glass	weather	B	bluish-green	99.81% ~ 100%
	No weathering	A, C	Blue-green, light blue, dark blue	97.25% ~ 100%
Lead-barium glass	weather	A, C	Blue-green, light blue, dark blue, black and purple	90.17% ~ 99.89%
	No weathering	B	Dark blue, light blue, dark green, light green, green and purple	88.41% ~ 99.98%

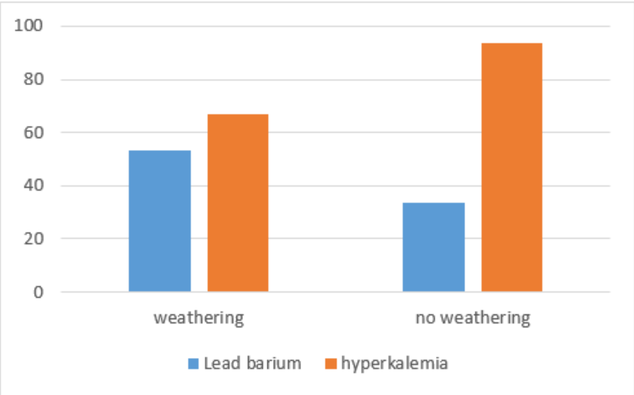

Figure 1. Summary mean of silica.

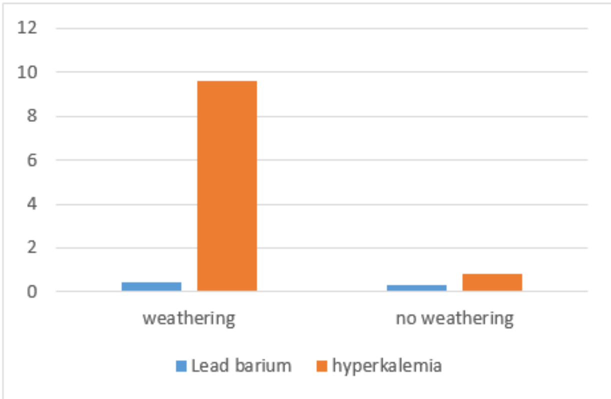

Figure 2. Summary mean of potassium oxide.

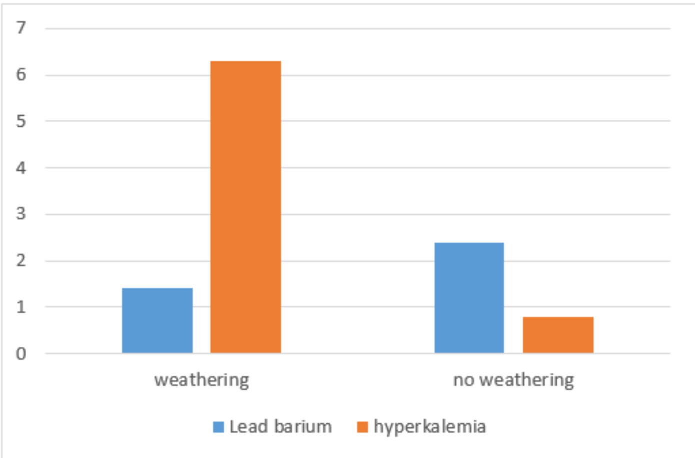

Figure 3. Summary mean of calcium oxide**.**

From the above bar chart, we can draw the conclusion:

(1) The content of silica after weathering is significantly higher for high-potassium type cultural relics than before weathering, while for lead.

The content of silica after weathering is lower than that before weathering for barium type cultural relics.

(2) For the relics of the high potassium type, the content of potassium oxide after weathering is significantly lower than that before weathering, while that of the lead barium type.

The potassium oxide content of cultural relics before and after weathering is not much different.

(3) The content of calcium oxide after weathering is significantly lower for high-potassium cultural relics than before weathering, while for lead barium.

The content of calcium oxide after weathering is higher than that before weathering.

(4) The content of alumina after weathering is significantly lower for the relics of high potassium type than before weathering, while that of lead barium type.

The content of alumina after weathering is slightly higher than that before weathering.

(5) For high-potassium cultural relics, the content of iron oxide after weathering is significantly lower than that before weathering, while for lead barium.

The same is true for the type B cultural relics.

(6) For high-potassium cultural relics, the copper oxide content after weathering is slightly lower than that before weathering, while for lead barium the copper oxide content after weathering is slightly higher than before weathering.

(7) For the relics of high potassium type, the content of phosphorus pentoxide after weathering is slightly lower than that before weathering, while for lead barium.

The content of phosphorus pentoxide after weathering was significantly higher than that before weathering.

At the same time, by analyzing the tabular data in this paper, it is found that for some types of lead-barium glass (related to color), it contains no or very little sulfur dioxide and lead oxide before weathering, but after weathering, the content of these two chemical components rises sharply.

2.1.2. Prediction of chemical composition before and after weathering

The topic asks this paper to predict the content of chemical components before weathering according to the detection data of weathering points. First, the data in the table is processed in this paper. In this paper, all the measured weathering point data are listed into a separate table to study the problem. According to the data, this paper finds the chemical content data of three severe weathering points, and then this paper uses the law obtained in previous text to predict.

According to the monitoring data of weathering points extracted from the data in this paper, the statistical law obtained in previous text will be used to predict the contents of various chemical components before weathering.(As shown in Table 5-6) (Data sourced from (mcm.edu.cn))

Table 5. Predicted pre-weathering content of each component.

engraved pattern	type	colour	Surface weathering	Cultural relics sampling point	Oxidized steel (Cu0)	Lead oxide (PbO)	Barium oxide (BaO)	Phosphorus pentoxide (P205)	Strontium oxide (SrO)	Tin oxide (SnO_2)	SOz (dioxide production)	total content
C	Lead and barium	purple	weather	26 Severe weathering point	4.45	15.92	3.45	0.88	0.60	0.00	0.00	94.35
C	Lead riveting	purple	weather	08 severe weathering point	8.14	29.45	25.62	0.55	0.65	0.00	0.00	96.55
C	Lead lock	light blue	weather	54 Severe weathering point	0.81	25.13	10.23	0.13	0.86	0.00	0.00	96.67

The figure above is the amount of various chemical components before weathering at each weathering point predicted according to the law obtained from previous text.

Table 6. The actual pre-weathering content of each component.

engraved pattern	type	colour	Surface weathering	Cultural relics sampling point	Silicon dioxide (SiO2)	Sodium oxide (Na20)	Potassium oxide (K20)	Calcium oxide (CaO)	Magnesium oxide (XgO)	Alumina (A1203)	Iron oxide (Fe2O3)
C	Lead and barium	purple	weather	26 Severe weathering point	65.67	0.00	0.00	1.03	0.00	2.35	0.00
C	Lead and barium	purple	weather	08 severe weathering point	30.64	0.00	0.00	0.19	0.00	1.31	0.00
C	Lead and barium	light blue	weather	54 Severe weathering point	54.11	0.00	0.00	0.00	1.35	4.05	0.00

engraved pattern	type	colour	Surface weathering	Silicon dioxide (Si02)	Sodium oxide and potassium oxide (Na20) (K20)	Calcium oxide (CaO)	Magnesium oxide (Mg0)	Alumina (A12O3)	Iron oxide (Fe2Os)
C	Lead and barium	purple	No weathering	31.94	.0.00 0.00	0.47	0.00	1.59	0.00
C	Lead and barium	purple	No weathering	65.91	0.00	0.00 1.60	0.89	3.11	4.59
C	Lead and barium	light blue	No weathering	55.21	0.00	0.25 0.00	1.67	4.79	0.00

engraved pattern	type	colour	Surface weathered copper oxide	Lead oxide (CuO) (PbO)	Barium oxide (BaO)	Phosphorus pentoxide (P20s)	Strontium oxide (SrO)	Tin oxide (SnO2)	Sulfur dioxide (SO2)	total content
C	Lead and barium	purple	No weathering 8.46	29.14	26.23	0.14	0.91	0.00	0.00	98.88
C	Lead and barium	purple	No weathering 0.44	16.55	3.42	1.62	0.30	0.00	0.00	98.43
C	Lead and barium	light blue	No weathering 0.77	25.25	10.06	0.20	0.43	0.00	0.00.	98.63

Analysis of prediction results: This paper compares the predicted data with the contents of various chemical components of unweathered cultural relics of the same type and color given by the title, and finds that the predicted data in this paper is close to the actual situation.

2.2. Establishment and solution of K-Means ++ algorithm

First, according to the requirements of K-Means ++ algorithm, this paper needs to process and screen the attached data[6-8].

Firstly, this paper selects the index used for clustering. As far as the feature of "ornamentation" is concerned, this paper holds that it.

It is of no practical significance to mineral classification, so this element is no longer considered in this paper. As for the mineral feature of "color", this paper believes that color has its significance for mineral classification, but color contains at least two aspects of the characteristics of light and shade, so this paper decides to cluster the two characteristics respectively, and analyzes the characteristics of that type of color as the classification law is better.

As for the indexes of "degree of weathering" and "content of various compounds", this paper believes that they are more critical for mineral classification and need to be included in the clustering standard. Secondly, for different sampling points in the data, the mineral data collected from the "severe weathering point" will be excluded in this paper, and the data from different parts of the same sampling point will be averaged as the data used in this paper for clustering.

For different degrees of weathering, this paper uses the number 1 to represent the unweathered minerals and the number 2 to represent the weathered minerals.

For different color systems, this paper uses 1 to represent the green system, 2 to represent the blue-green system, and 3 to represent the blue system, and removes the purple and black mineral data;

For different color shades, this paper uses 1 to represent the dark series, where the dark series contains black, dark blue, dark green, purple, and 2 to represent the light series.

Second, for the selected data, this paper begins to process them according to the K-Means ++ algorithm. The following table shows the data processing results using SPSS software.(As shown in Table 7-8).

Table 7. Final clustering centers for clustering of color shades.

Final clustering center		
	cluster	
	one	2
colour	2	2
Surface weathering	one	2
Silicon dioxide (Si02)	70.3	31.92
Oxidized (Xa20)	1.16	0.66
Potassium oxide (K20)	3.46	0.22
Calcium oxide (Ca0)	2.29	2.01
* Magnesium oxide (Ng0)	0.75	0.48
Alumina (A1203)	4.68	2.68
Iron oxide (Fe203)	0.86	0.55
Copper oxide (CuO)	1.49	2.59
Lead oxide (PbO)	7.97	39.16
* barium oxide (BaO)	3.1	12.75
Phosphorus pentoxide (P205)	0.74	2.85
Strontium oxide (SrO)	0.11	0.37
Tin oxide (Sn02)	0.08	0.04

(Data sourced from (mcm.edu.cn))

Table 8. Final clustering centers for color clustering.

Final clustering center		
	cluster	
	one	2
colour	2	2
Surface weathering	one	2
Silicon dioxide (Si02)	70.45	33.18
Sodium Na20)	1.19	0.76
Potassium oxide (K20)	3.57	0.25
Calcium oxide +(CaO)	2.31	2.15
Magnesium oxide (Mg0)	0.75	0.56
Alumina (A1203)	4.73	2.91
Iron oxide (Fe203)	0.74	0.64
Copper oxide (CuO)	1.52	1.45
Lead oxide (Pb0)	7.7	40.74
Barium oxide (BaO)	3.08	10.04
Phosphorus pentoxide (P205)	0.71	2.94
Strontium oxide (Sro)	0.1	0.34
Tin oxide (Sn02)	0.08	0.05
While sulfur oxide (S02)	0.16	0

By analyzing the above two groups of data and comparing the classification results with those in the original attachment.

This paper finds that the classification results after clustering are almost consistent with the classification of the original data, so this paper can be rooted.

The classification rules of the original data are analyzed according to the classification of this paper.

Analysis:

(1) As far as "color" features are concerned, this paper finds that color system and color depth have little impact on classification, so

In this paper, the feature of "color" is not included in the classification rule.

(2) As for the feature of "surface weathering degree", this paper finds that it has a greater impact on classification, so this paper

It is decided to include the feature of "surface weathering degree" into the classification rule.

(3) For the content of various compounds, this paper found that it has a greater impact on the classification, so this paper decided to start from each.

Among the compounds, several compounds with obvious characteristics and differences are screened out and included in the classification law of this paper.

Specific results (classification rules of two types of minerals)

(1) In terms of weathering degree, lead barium glass has a higher degree of weathering, while high potassium glass has a lower degree of weathering.

(2) In terms of the content of compounds contained in lead-barium glass, the content of lead oxide and barium oxide is much higher than that of high-potassium glass and the content of silica in high potassium glass, potassium oxide is much higher than that of lead barium glass.

2.3. Data clustering of two groups of glass

In the first step, this paper divides the data into two groups, namely high potassium glass group and lead barium glass group. As for the first minor problem in previous text, in order to facilitate the comparison between the classification results and those in the original attachment, we averaged different parts from the same sampling point. For accuracy, we no longer averaged different parts from the same sampling point in this paper [9-10].

In the second step, this paper uses the K-Means ++ algorithm to cluster the two groups of data respectively. The following is the clustering results. (As shown in Table 9-10) (Data sourced from (mcm.edu.cn))

Table 9. Clustering classification of high potassium glass: final clustering center.

Final clustering center		
	cluster	
	one	2
Silicon dioxide (5102)	63.2	89.6633
Sodium oxide (Na20)	1.24	0
Potassium oxide (K20)	10.05	1.99
Calcium oxide (Ca0)	5.2064	1.3267
Magnesium oxide (Mg0)	1.0973	0.4367
Alumina (A1203)	6.2991	2.7644
Iron oxide (Fe203)	2.0809	0.44
Copper oxide (CuO)	2.5227	1.4922
Lead oxide (PbO)	0.37	0.14
Barium oxide (BaO)	0.47	0.22
Phosphorus pentoxide (P205)	1.2864	0.5333
Strontium oxide (SrO)	0.04	0.01
Tin oxide (Sn02)	0	0.26
Sulfur oxide (S02)	0.11	0

Table 10. Cluster classification of lead barium glass: final cluster center.

Final clustering center		
	cluster	
	one	2
Silicon dioxide (Si02)	56.51	24.73
Sodium Na20)	2.32	0.19
Potassium oxide (K20)	0.21	0.14
Calcium oxide (CaO)	1.12	2.85
Magnesium oxide (NgO)	0.78	0.63
Alumina (A1203)	5.63	2.67
Iron oxide (Fe203)	0.76	0.6800.
Steel oxide (CuO)	1.18	2.66
Lead oxide (Pb0)	19.5	41.52
Barium oxide (BaO)	8.33	13.31
Phosphorus pentoxide (P205)	0.9	5.25
Strontium oxide (SrO)	0.24	0.44
Tin oxide (Sn02)	0.09	0.0500
Sulfur oxide (S02)	0	1.42

Through the analysis of clustering results, the following subclass classification results are obtained:

(1) For high-potassium glass, the content of various metals in class 1 high-potassium glass is significantly higher than that in class 2 high-potassium glass.

The silica content of class 2 high potassium glass is higher than that of class 1 high potassium glass, so this paper believes that it can be.

High potassium glass is divided into two categories: metal-rich and silicon-rich.

(2) For lead-barium glass, the lead oxide and phosphorus pentoxide content of class 1 lead-barium glass are significant.

The content of alumina, sodium oxide and silica in class 1 lead barium glass is higher than that in class 2 lead barium glass.

Lead barium glass, so this paper believes that lead barium glass can be divided into two categories: lead-rich phosphorus and silica-rich aluminum.

2.4. Rationality and sensitivity of classification results are analyzed

As the title requires this paper to analyze the rationality and sensitivity of the classification results. So this article decides to be right.

The sensitivity analysis of four subclasses was carried out. The specific methods are as follows: according to the division rules of the subclasses, that is, used for classification.

The content of the main compounds, for example, as far as silica is concerned, this paper selects a series of silica content different and for cultural relics with large differences, correlation analysis is performed with the distance between the cultural relic and the final clustering center to obtain the sensitive path degrees. Here are the results obtained in this article. (As shown in Figure 4-9).

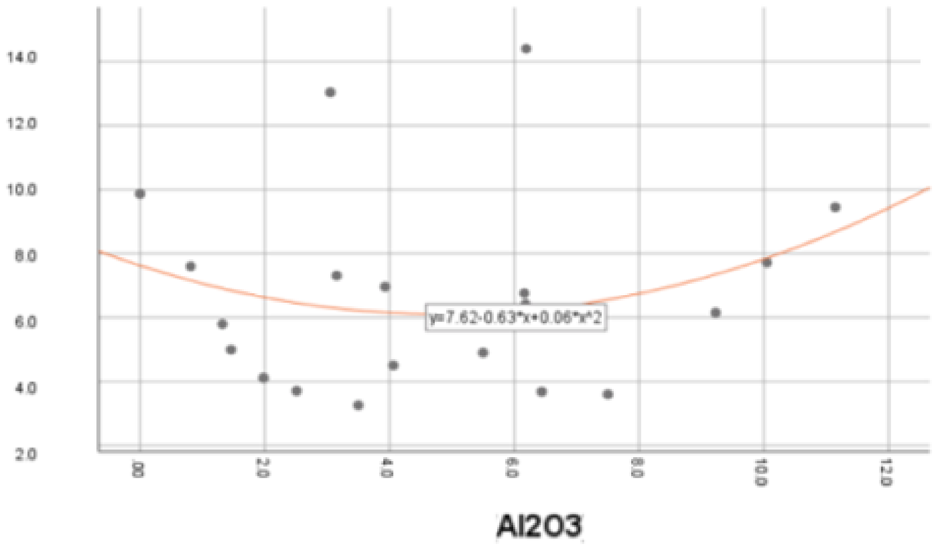

Figure 4. Subclass classification of high potassium glass: relationship between alumina content and distance of artifacts from clustering centers.

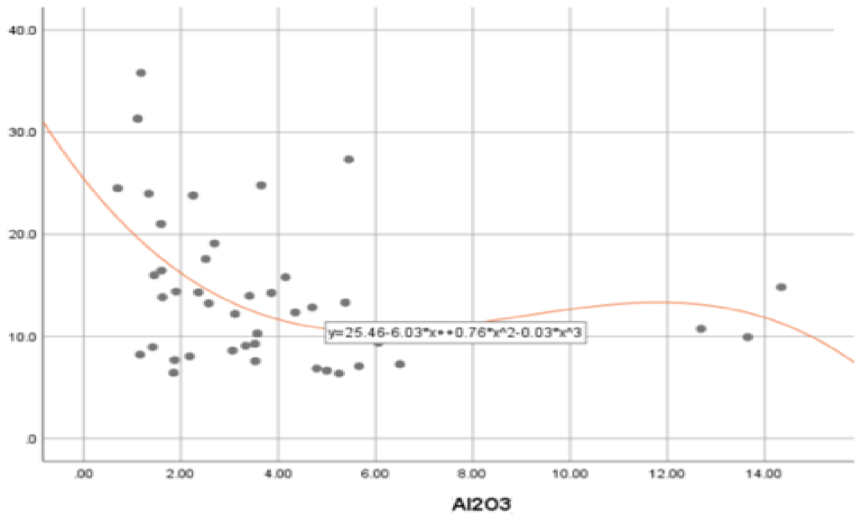

Figure 5. Subclass classification of high potassium glass: relationship between calcium oxide content and distance of artifacts from clustering centers.

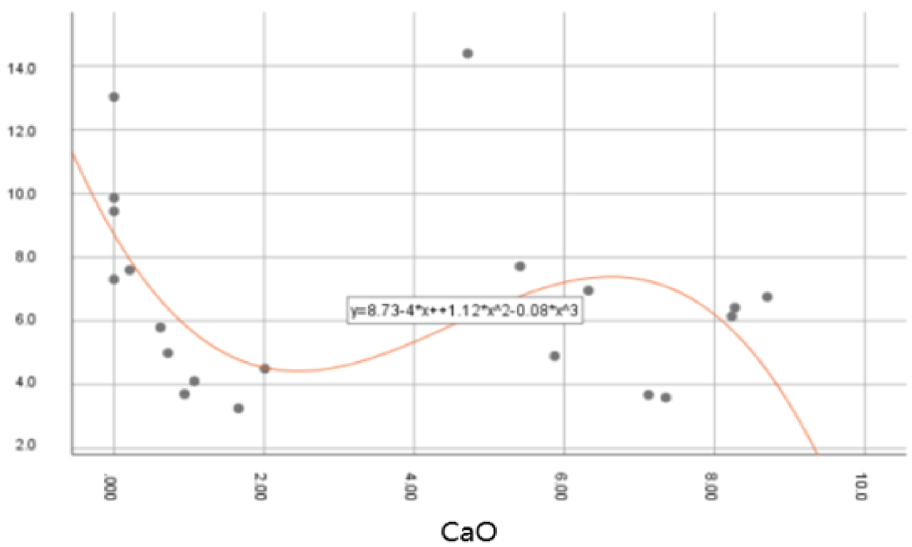

Figure 6. Subclass classification of lead-barium glass: relationship between alumina content and distance of artifacts from clustering centers.

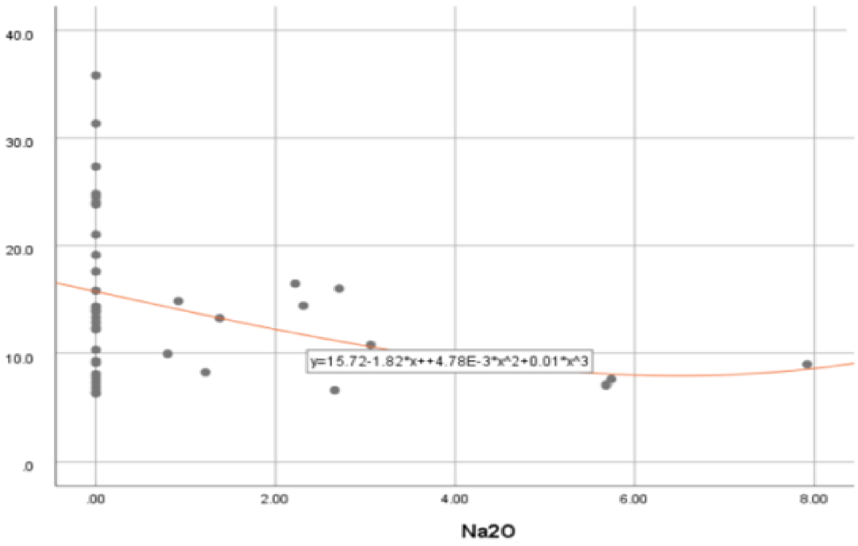

Figure 7. Subclass classification of lead-barium glass: Relationship between sodium oxide content and distance of artifacts from clustering centers.

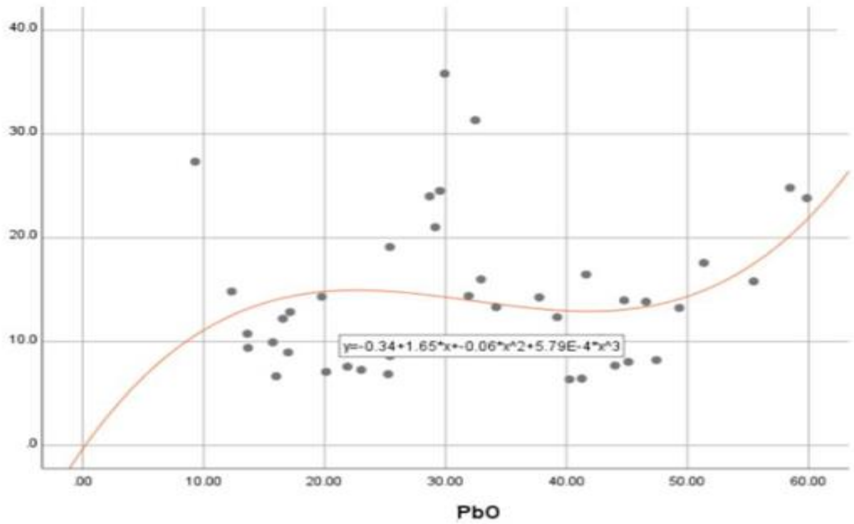

Figure 8. Subclass classification of lead-barium glass: relationship between lead oxide content and distance of artifacts from clustering centers.

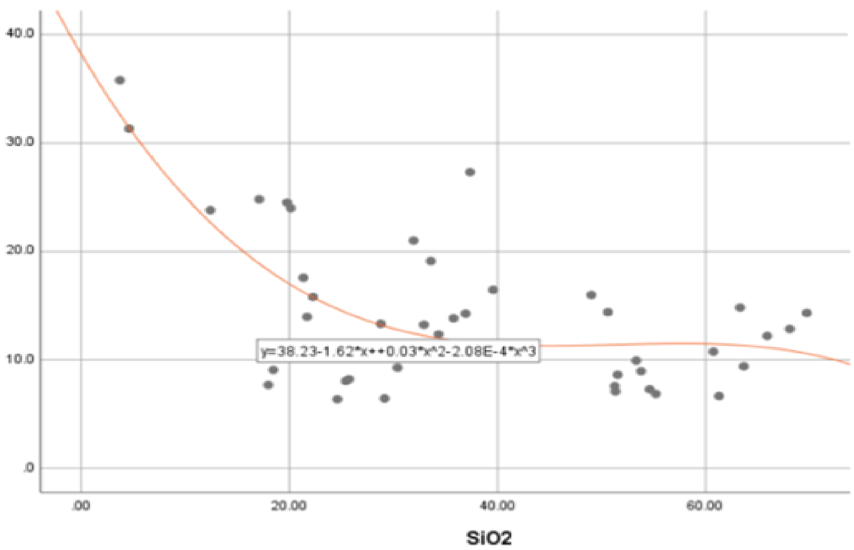

Figure 9. Subclass classification of lead-barium glass: relationship between silica content and distance of artifacts from clustering centers

Through the above data, this paper found that for the subclass classification of high-potassium glass, its effect on alumina, oxygen.

The content of calcium oxide is more sensitive, which accords with the subclass classification of high potassium glass in this paper; for lead barium glass subclass classification

In terms of the class rule, it is more sensitive to the content of lead oxide, silicon dioxide, sodium oxide and alumina, which accords with this paper.

The subclass classification of this paper.

3. Conclusion

Firstly, this paper analyzes the relationship between the surface weathering of cultural relics and the type, color and pattern of glass by Chi-square test. Secondly, the glass is divided into two types: lead barium and high potassium. Then this paper predicts their contents before weathering, and compares the contents with those of other types and unweathered cultural relics with the same color and pattern. Then this paper uses the K-Means ++ algorithm to classify the unclassified data in Table 2 and analyze the classification rule obtained by clustering. Next, the same clustering algorithm is used to classify the data into the high-potassium glass group and the lead-barium glass group, and cluster the two groups of data respectively. By analyzing the clustering center rule, the classification standard of the subclass is obtained. Finally, the correlation analysis method is used to conduct correlation analysis between the content of main compounds in the subclass classification law of each cultural relic and the distance between each cultural relic and the final clustering center, so as to test the sensitivity.

The conclusion drawn in this article can predict the composition content of ancient glass unearthed in archaeology, which is of great significance to archaeology.

In this paper, the linear regression model is used to predict the chemical composition. Due to the simple prediction model, the experimental error is relatively large. In the next step, a more complex prediction model can be used, such as the nonlinear prediction model to fit, so as to predict more accurately.

References

[1] Cai Guanghui, Wu Zhimin. Study on the efficacy and AUC of several types of normality tests under perfect and imperfect two-stage sorted set sampling [J]. Systems Science and Mathematics, 2023, 43 (1): 17.

[2] Qian Hongyu, Yue Siqing, Xu Lei, et al. A discrimination method for ancient glass systems based on multiple linear regression [J]. Gansu Science and Technology, 2024 (002): 040.

[3] Sun Haodong. Chemical composition analysis and research of ancient glass products based on logistic regression [J]. Computer enthusiast, 2023 (1): 19-21.

[4] Liu Zhen SPSS Statistical Analysis (6th Edition) (Classic Edition) Science and Computer Science in College and Secondary School [M]. Electronic Industry Press, 2023.

[5] Liu Yonghong, Niu Zheng, Wang Changyao. Research and application of decision tree classification method based on MODIS data [J]. Journal of Remote Sensing, 2005, 9 (4): 8.

[6] Zhang Yufang, Mao Jiali, Xiong Zhongyang. An improved K-means algorithm [J]. Computer Applications, 2003, 023 (008): 31-33, 60.

[7] Tang Meiyan, Li Zhongying. Analysis of the influencing factors of electricity consumption based on chi square test and logistic regression: A case study of household electricity consumption in Qinzhou City. Statistics and Consulting, 2022 (4): 22-26.

[8] Yang Junchuang, Zhao Chao Review of K-Means clustering algorithm research [J]. Computer Engineering and Applications, 2019, 55 (23): 9.

[9] Zhao Le, Zhang En, Qin Leiyong, et al. Multi party privacy protection k-means clustering scheme based on blockchain [J]. Computer Applications, 2022, 42 (12): 12.

[10] Yang Nan, Song Hang, Li Weiwen. Mining and Impact Analysis of Two Types of Glass Weathering Laws Based on Statistical Testing [J]. Progress in Applied Mathematics, 2022.