Composition analysis and identification model of glass artifacts

Huilin Pan #,*, Fuyuan Chen#, Wei Hang

Gansu University of Political Science and Law, Gansu province, Lanzhou,730070 China

*Corresponding author:ytytytgood@126.com

#These authors contributed equally.

Abstract: This paper analyzes the correlation and variability between surface weathering, type, color, decoration, and chemical composition of glass artifacts. Based on the interactions of these five factors, the classified glass artifacts are studied and modeled, and the unclassified glass artifacts are further investigated.

Keywords: Correlation Analysis; Chi-Square Test Analysis; Linear Regression (Least Squares); Hierarchical Cluster Analysis; Decision Tree.

1. Introduction

In ancient times, glass was used as a precious item for trade between China and the West along the Silk Road, greatly promoting the exchange and mutual appreciation between the East and the West.[1~3] Ancient glass products were made in China by absorbing foreign technology and taking materials according to local conditions. Ancient glass is susceptible to weathering due to its long-term underground burial environment and the complexity of its environment.[4~6] As a result of weathering, the surface of glass products became dull and less transparent due to the weathering layer. Some research results have been conducted on the weathering phenomenon of glass and its mechanism, which show that the degree of weathering of glass is closely related to its chemical composition, glass type, the external environment (temperature, humidity, etc.), and weathering time[7~10].

In this paper, we analyze the correlations and differences between the weathering of the surface of classified glass artifacts and the type, decoration, and color of glass, and predict the chemical composition of glass artifacts before weathering. Glass is classified and subclassified, and the results are analyzed for reasonableness and sensitivity. The correlation relationships and variability between the different types of chemical compositions were analyzed.

2. Methods

The surface weathering of glass artifacts was analyzed for their glass type, decoration, and color. In the first step, the data were analyzed by the correlation Spearman method using spss, and the correlation analysis of cross-tabulation; in the second step, the difference chi-square test was analyzed by SPSSpro, and finally, the data were analyzed according to the p-value of the chi-square test with the correlation analysis of Spearman and cross-tabulation, and from the final data analysis, it was proved with sufficient certainty that: there is no significant relationship between the presence or absence of weathering and ornamentation and color; there is a significant relationship between the presence or absence of weathering and type The final data analysis shows that there is no a significant relationship between the presence or absence of weathering and decoration and color; there is a significant relationship between the presence or absence of weathering and type.

Combined with the type of glass, the statistical pattern of the content of chemical composition with and without weathering on the surface of the artifact samples is analyzed. From the data in the form, it is clear that the type of glass can be divided into high potassium and lead-barium categories, so we need to fix a variable, i.e., fix the type variable, to analyze the relationship between ornamentation, color, and the presence or absence of weathering. We have to determine the prediction target (i.e., chemical composition content before weathering), the dependent variable, and the independent variable first. The model is then determined by linear correlation, and finally, a linear regression equation is derived that can predict the chemical composition content before weathering.

Based on the relationship between each chemical composition of high potassium glass and lead-barium glass and its type, the basis for classifying different types was summarized and summarized by drawing a box plot. Using the hierarchical clustering algorithm in the unsupervised learning algorithm, we aim to figure out the pattern of indicators by learning from unlabeled training samples, to classify the glass artifacts for unknown categories in the future. In this clustering process, we use the number of clusters to classify the samples several times.

Based on the hierarchical clustering of classified glass artifacts, we analyze the chemical composition of unclassified glass artifacts, identify the categories, and analyze the sensitivity of the classification results. We use machine learning, using the decision tree algorithm in machine learning, under the python programming implementation, scipy. cluster. hierarchy clustering function to generate clustering tree function linkage, the parameter method indicates the clustering method, that is, the method of determining the value of the new sample each time the sample is synthesized into a new sample. The parameter metric indicates the distance measure, which is the euclidean distance. The subclassification is based on problem two, identifying the type to which it belongs; the sensitivity analysis is performed based on the enthalpy of information entropy size in the decision tree algorithm.

To analyze the correlation between the chemical composition of the glass samples of different categories, we use correlation analysis. Firstly, we need to import the panda's library in python and use the merge() function to divide and merge the different categories of glass, i.e., the artifacts of the high potassium category are divided into one category and the artifacts of the lead-barium category are divided into one category; secondly, we use the subplots function in the matplotlib library and the heatmap function in the seaborn library to establish a heatmap of the correlation between the chemical composition of high potassium glass and the chemical composition of lead-barium glass; finally, the analysis of the chemical compositions was carried out by observing the correlation heatmap. Difference analysis was used. The pair plot function in the seaborn library was used to build histograms of the variability of the chemical composition of high potassium glass and the chemical composition of lead-barium glass, and the variability of the correlation relationship between each chemical composition was observed and analyzed.

3. Results and Discussion

3.1 Glass chemical composition prediction

After analyzing the data we can see that the four indicators, surface weathering, glass type, ornamentation, and color, are all definite class variables, and the existence of a statistically significant relationship between the four indicators was tested, and the correlation was significant when $p < 0.05$. Surface weathering and type at the 0.05 level (two-tailed) were shown to be significantly correlated. The two showed a negative correlation.

The distribution of multiple variables and correlation analysis were analyzed simultaneously using cross-tabulations, and the three frequencies of surface weathering with ornamentation, type, and color were analyzed. The strongest correlation was obtained between surface weathering and ornamentation C, and glass artifacts with ornamentation C were more prone to weathering; the strongest correlation was obtained between surface weathering and lead-barium, and glass artifacts with more lead-barium content were more prone to weathering; the strongest correlation was obtained between whether the surface weathering was light blue and the color, and glass artifacts with light blue color were more prone to weathering.

It was assumed that there was no significant variability between surface weathering and type, decoration, and color. The pre-processed files were analyzed by cardinality test using SPSSPRO software. From the chi-square test table, we can conclude that: the p-value of surface weathering and type is 0.004, which is <0.05, so the original hypothesis is rejected and there is a significant difference between surface weathering and type; the p-value of surface weathering and decoration is 0.055, which is >0.05, so the original hypothesis is accepted and there is no significant difference between surface weathering and decoration; the p-value of surface weathering and the color is 0.183, which is > 0.05, so the original hypothesis is accepted, and there is no significant difference between surface weathering and color.

For the two types of high potassium and lead-barium, information on the presence or absence of weathering, color, ornamentation, and total chemical content of each was counted and plotted in Table 1. The independent and dependent variables were determined based on the assay data of the differentiation points. The independent variable was the content of 14 chemical components and the total content, and the dependent variable was the presence or absence of weathering on the surface.

Table 1 Glass Type

Type	With or without differentiation	Color	Ornamentation	Total chemical composition
High Potassium	Weathered	Blue-Green	B	99.8%-100%
	No weathering	Blue-green, light blue, dark blue	A、C	97.3%-100%
Lead Barium	Weathered	Blue-green, light green, light blue, dark green, purple, black	A、C	90.1%-99.89%
	No weathering	Dark green, light green, dark blue, light blue, green, purple	B	88.4%-99.9%

Before performing a regression analysis, we first need to understand the correlation between the independent and dependent variables to subsequently determine the type of regression model. So we will first perform a correlation analysis. Assumptions: There is no significant linear relationship between the presence or absence of weathering on the surface and the chemical composition, and the regression coefficient is 0.

Using SPSSPRO software, a linear regression (least squares) model prediction was performed. From the analysis of the results of the F-test, it can be obtained that the significant P-value is 0.000***, which presents significance at the level, so the original hypothesis that the regression coefficient is 0 is rejected, so the model meets the requirements and there is a significant linear relationship between the chemical composition and the presence or absence of surface weathering.

Establishing linear regression equation. After the model was established, the model needed to be tested. The fit of the model R^2=0.526, and the model meets the requirements. The equation of the model is as follows.

$$a = y = 0.703 + 0.024 * (SiO_2) + 0.068 * (Na_2O) + (-0.045) * (K_2O) + 0.044 * (CaO) + (-0.397) * (SrO) + 0.065 * (SO_2) + (-0.016) * Total\ ingredient\ content + 0.087 * (SnO_2) + 0.01 * (MgO) + 0.056 * (P_2O_5) + (-0.098) * (Fe_2O_3) + 0.001 * (BaO) + 0.05 * (Al_2O_3) + 0.034 * (PbO) + 0.085 * (CuO) \tag{1}$$

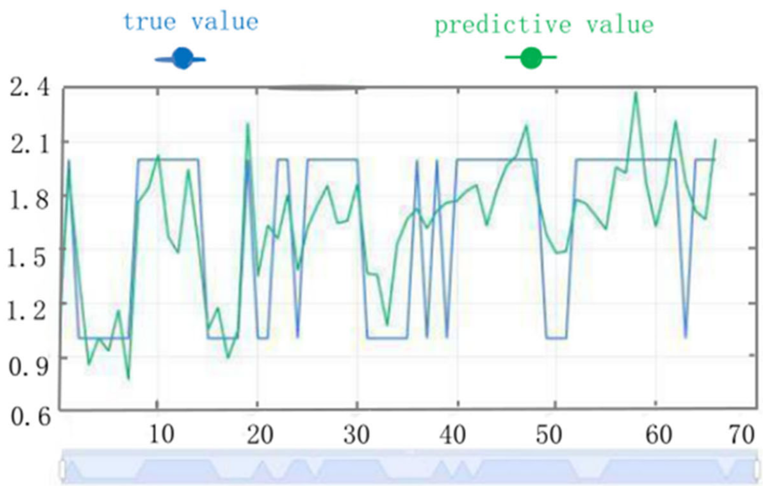

Figure 1 Fitting effect graph

Figure 1 is a fitting effect graph. According to the corresponding regression equation, we can predict that the chemical composition content of high potassium glass before weathering is

The chemical composition content before weathering is silica, potassium oxide, calcium oxide, aluminum oxide, sodium oxide, etc.; the chemical composition content before weathering of lead-barium glass is silica, lead oxide, barium oxide, sulfur dioxide, etc.

3.2 Glass classification

We compare the distribution characteristics of multiple data sets by plotting box plots, such as in figure 2.

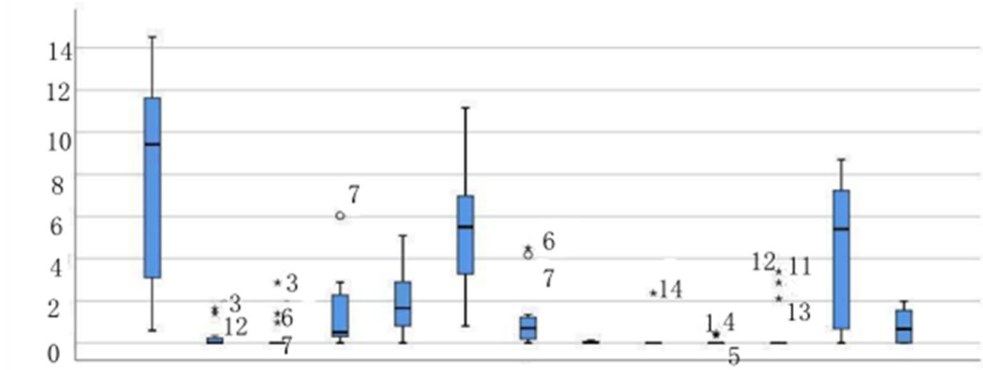

Figure 2 High potassium box plot

As can be seen from the figure: the most content of potassium oxide (K_2O) in high potassium, aluminum oxide (Al_2O3), sodium oxide (Na_2O) and content of more, and copper oxide (CuO) and magnesium oxide (MgO) content is less, and the above five chemical components are no outliers, indicating that they are all necessary substances to form high potassium glass. Lead oxide (PbO), barium oxide (BaO), sulfur dioxide (SO_2), sodium oxide (Na_2O), and phosphorus pentoxide (P_2O_5) have more outliers, which can be used as a basis for distinguishing the classification of high potassium glass and lead-barium glass. Whether the five chemical substances with more outliers can be used as the basis for classifying glass requires further analysis in comparison with the chemical composition in the lead-barium glass.

Similarly lead barium glass contains the most lead oxide (PbO), more barium oxide (BaO) and phosphorus pentoxide (P_2O_5), and less of all other chemicals. Among the chemical components with more content, lead oxide (PbO) contains not only more but also no outliers, and lead oxide (PbO) has more outliers.

Therefore, through the comprehensive box plot of high potassium and lead-barium, we can know that potassium oxide (K_2O) and lead oxide (PbO) are the basis for distinguishing between high potassium glass and lead-barium glass. Those with high potassium oxide (K_2O) content are high potassium glasses, and those with high lead oxide (PbO) content are lead-barium glasses.

With the unsupervised learning algorithm, the markers of the training samples are unknown, and we aim to figure out the pattern of the indicators by learning from the unmarked training samples, to classify the glass artifacts for unknown categories in the future. Among the many unsupervised learning algorithms, we choose the hierarchical clustering algorithm. The clustering results are shown in Figure 3~6.

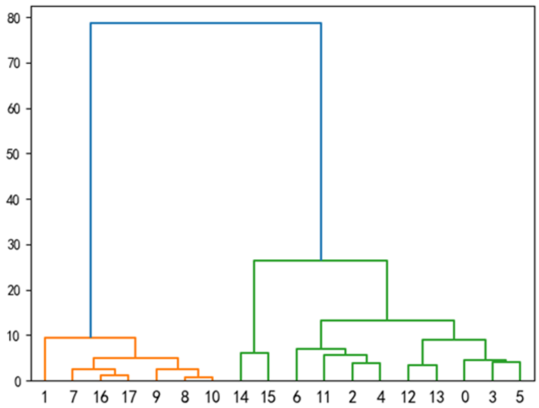

Figure 3 High potassium class clustering tree

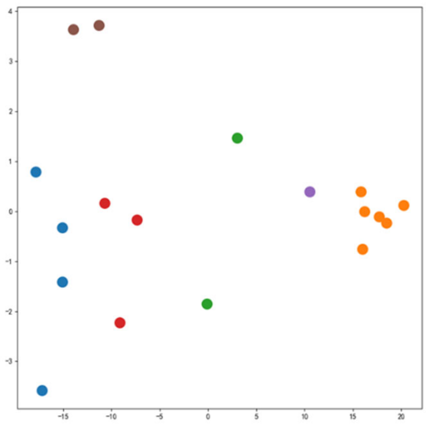

Figure 4 Two-dimensional visualization of the data distribution of high potassium into 6 categories

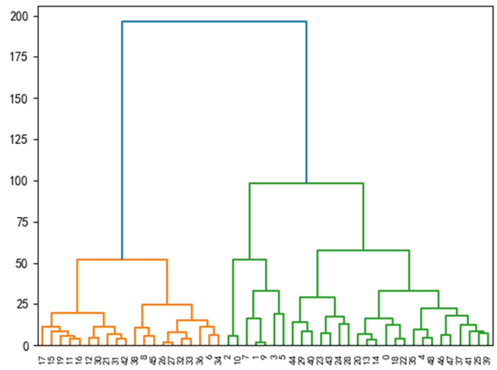

Figure 5 Cluster tree of lead-barium classes

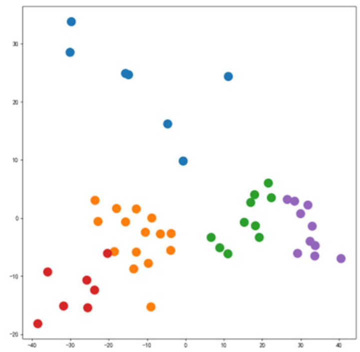

Fig. 6 Two-dimensional visualization of the data distribution of lead and barium classified into 5 categories

3.3 Decision tree algorithm

For unknown categories of glass artifacts, we have to analyze them and identify their types based on the chemical composition they contain. We have to perform machine learning using a decision tree algorithm for classification. Figure 7 is the decision tree.

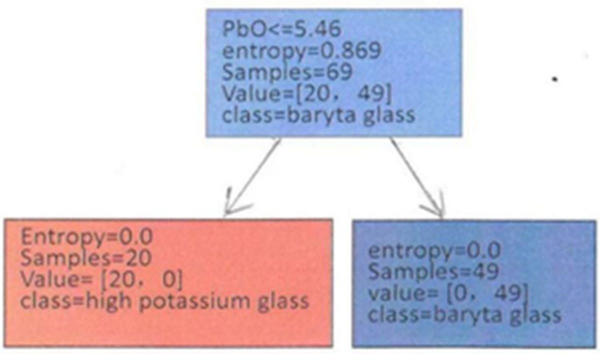

Figure 7 Decision tree

Using the sklearn module, the method of hierarchical cluster analysis based on sub-classification, using python to build a decision tree model; learn calculates the information gain based on the information entropy, that is, the difference between the information entropy of the parent node and the information entropy of the child node. So according to the judgment of information enthalpy entropy, decide how to deal with the differentiated data; the maximum depth max_depth is 4, limit greater than 4, pruning after prediction; use the feature data of hierarchical classification to classify the data, specify the training set as 0.8, test set as 0.2; finally use matplotlib.pyplot library for data visualization.

According to the figure shows: under the standard of lead oxide <= 5.46, under the total number of samples, the branching classification process is carried out to get its subclass 1 high potassium glass and subclass 2 lead barium glass; when the information entropy is smaller, the less information, the more certain the event is, so when the information entropy of the two subclasses is 0, it can be analyzed effectively based on its sensitivity size, so the modeling is established effectively.

3.4 Correlation analysis

To analyze the relationship between the chemical compositions of different categories of glass artifact samples, we use correlation analysis, i.e., the creation of a thermogram. Heat map, also known as correlation coefficient map. The principle is that the magnitude of the correlation between variables can be judged according to the magnitude of correlation coefficients corresponding to the colors of different squares in the heat map.

We implement it by python code, see the appendix for the specific code. Firstly, we need to import the panda's library in python and use the merge function in it to divide and merge different categories of glass, that is, the artifacts of the high potassium category are divided into one category, and those of lead-barium category is divided into one category; secondly, we use the subplots function in the matplotlib library and the heatmap function in the seaborn library to establish The heatmap of the chemical composition of high potassium glass and lead-barium glass. The heat maps of the chemical compositions are in Figures 8 and 9.

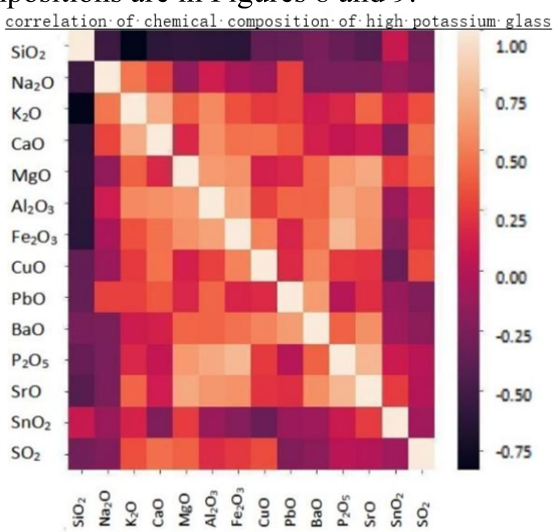

Figure 8 Thermal diagram of the chemical composition of high potassium glass

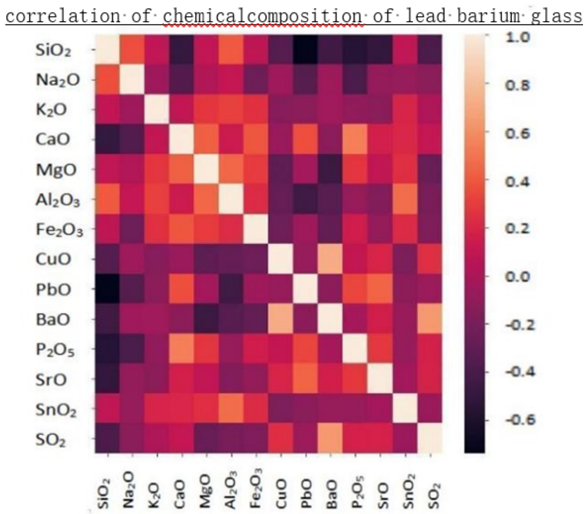

Figure 9 Heat diagram of the chemical composition of lead barium glass

Interpretation of the results: The scale on the right side of the thermogram shows the color shades corresponding to different correlation coefficients. It can be seen from both graphs that the correlation between silica and other chemical components in the thermogram is darker, i.e., the correlation between silica and other chemical components is lower, confirming the statement of the title that the main component of the glass is silica, so the correlation with other chemical components is low. From Figure 10, it can be seen that the correlations of chemical components such as potassium oxide, calcium oxide, and magnesium oxide are high in high potassium glass; from Figure 11, it can be seen that the correlations of barium oxide and copper oxide and sulfur dioxide are high in lead-barium glass, and the correlations of lead oxide and strontium oxide are high.

For the comparison of the correlations of chemical composition between different categories, we use the analysis of variance. It is implemented by python code, which is shown in the appendix, simply by using the pair plot function in the seaborn library to build histograms of the chemical composition of high-valent glass and the chemical composition variability of lead-barium glass.

The histogram of chemical composition variability is as Tables 10 and 11.

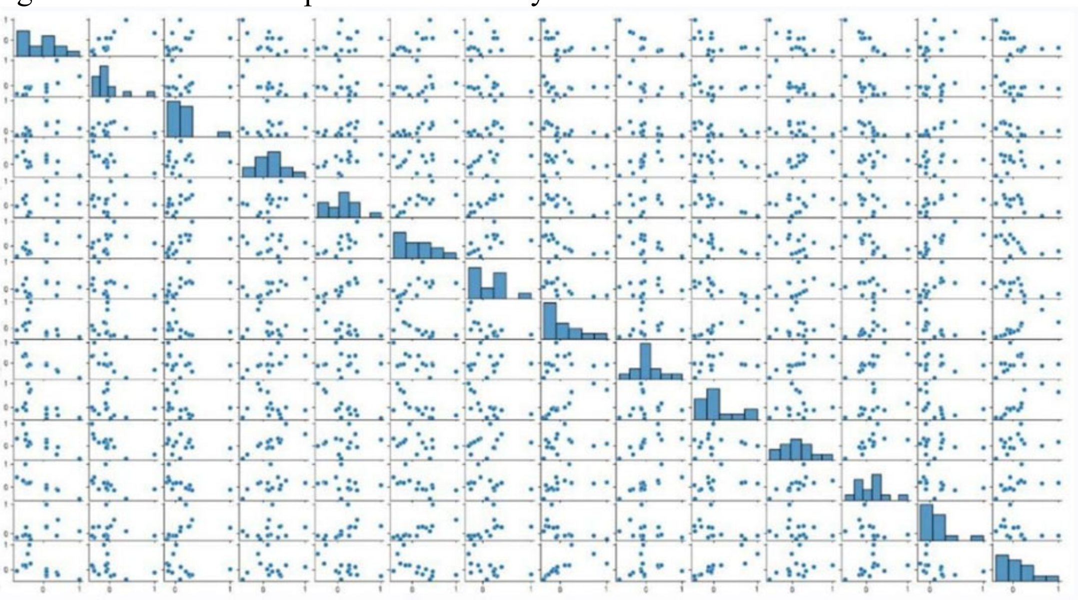

Figure 10 Histogram of the variability of the chemical composition of high potassium glass

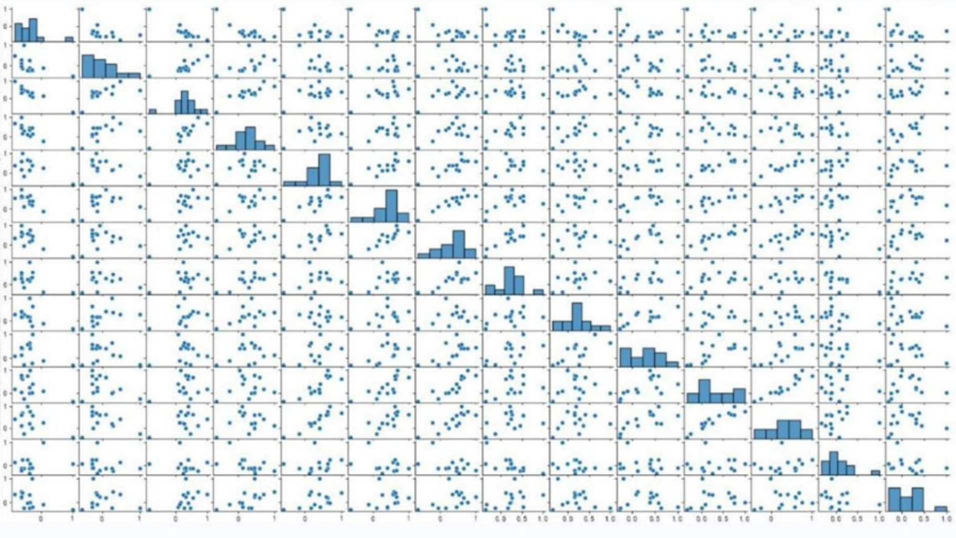

Figure 11 Histogram of the variability of the chemical composition of lead-barium glass

Interpretation of results: From the above two histograms, it can be seen that whether it is high potassium glass or lead-barium glass, its silica has a large difference from other chemical components, it is obvious that the main component of the illuminated glass is silica, and other chemical components are less, so its difference is larger.

4. Conclusion

The main conclusions of this paper are as follows

(1) There is a significant relationship between the presence or absence of weathering and type.

(2) The model was determined by linear correlation, and finally, a linear regression equation was derived to predict the chemical composition content before unweathering.

(3) After several classifications, high potassium was classified into five types, and lead barium into six types.

(4) When the information is first, the smaller the amount of information, the more certain the event to carry out the sensitivity risk.

(5) There is a correlation between potassium oxide and other chemical components in high-potassium glass, and there is a correlation between lead oxide and barium oxide and other chemical components in lead-barium glass.

References

[1] Liu Huawei The outstanding craftsmanship of glass -- "Chinese style" in European glass art [J] World of Cultural Relics, 2018 (7): 4

[2] Wu Yanpeng, Yizhuo. Penglai: the starting point of the Maritime Silk Road [J] Towards the World, 2020 (S01): 2

[3] Wang Di. From the Perspective of the Silk Road: The Origin of the Chinese Imagery Speculated in French Decorative Art in the 18th Century [J] Popular Literature and Art: Academic Edition, 2018 (3): 2

[4] Fu Xiangshen, Wang Qi, Yang Zhenbei, et al Experimental study on compressive and shear strength of weak expansive soil improved by machine-made sand and sodium silicate [J] Journal of Civil and Environmental Engineering (Chinese and English), 2022 (044-004)

[5] Ding Chen, none It makes glass art "colorful" [J] Chinese Collection, 2022 (3): 2

[6] Talk about Shengguang, Qin Chunlei Guide: Colorful: Glass Art in the Communication between the Ancient East and West ^ Ming Dynasty [J] Art Appreciation, 2022 (3): 4

[7] Xu Jian, Peng Longgui Study on the control method of glass fiber crystalline transition and its inhibition mechanism on concrete alkali-aggregate reaction [J] Concrete World, 2022 (000-006)

[8] Chu Yadong, Xu Dongxia, Wang Zongwei, et al Study on brazing technology and joint properties of high silicon aluminum alloy (CE11)/soda-lime glass [J] Weapon Materials Science and Engineering, 2022 (045-004)

[9] Zheng Jiexi, Gong Kuiyuan, Zhang Hongyang, et al The action mechanism and research progress of fluxes and adhesives on glass lining [J] Silicate Bulletin, 2022, 41 (8): 9

[10] Chen Yan, Yu Kuaihai, Li Jiaqi, et al Study on the particle impact resistance of micro-textured columns on the surface of photovoltaic glass [J] Spacecraft Environmental Engineering, 2022, 39 (4): 8