

Chemical composition analysis and type identification of ancient glass products

Jiaman Li, Mengyao Zhao

Northwest Normal University, Lanzhou 730070, Gansu, China

Abstract. In this paper, the chemical composition analysis and type identification of ancient glass products are studied. Firstly, the qualitative analysis is changed into quantitative analysis, the type of glass, pattern and color are treated quantitatively, and the Chi-square test model is established. On this basis, the Chi-square cross thermal map is output to analyze the correlation degree of samples. The results of effect quantification are obtained. In order to study the division method and results of the internal chemical components of high potassium glass and lead barium glass, the principal component analysis method was adopted to reduce the dimension of the chemical components. Then train the data and build the model. This paper discusses the differences of correlation between chemical constituents of different types of cultural relics samples from the perspective of statistics.

Keywords: Chi-square test, Grey prediction, K-means cluster analysis, Logistic regression model, Canonical correlation analysis.

1. Questions to ask

In question 1, Quantitative analysis and prediction of the chemical composition of glass products based on the weathering, type, ornamentation and color data of the glass surface.

In question 2, problem is to make a quantitative analysis of the classification rules of the two kinds of glass by using the attached data, divide the two categories according to chemical components and give the classification methods and results, and analyze the rationality and sensitivity of the classification results.

In question 3, Combined with problem 2, Form 3 is used to analyze the composition of glass relics of unknown types, and the sensitivity analysis of the results is given.

In question 4, problem is to analyze the relationship between the chemical constituents of different types of glass cultural relics samples, and to show the differences of their associations.

2. Problem analysis

2.1 Analysis of Question 1

In order to analyze the relationship between the surface weathering of these glass relics and its glass type, ornamentation and color, it is necessary to quantify the properties of the cultural relics samples and establish a chi-square test model. The evaluation index of the chemical composition content of cultural relic samples before weathering is selected to establish a grey prediction model. Based on the established index system, the chemical composition content of cultural relic samples before weathering is evaluated and calculated.

2.2 Analysis of Question 2

Based on the data, the classification rules of high-potassium glass and lead-barium glass were analyzed, and the appropriate chemical composition was found for each type of glass and subclass was divided. Firstly, principal component analysis is used to reduce the dimension of each chemical composition of the two kinds of glass. Next, K-means cluster analysis is used to obtain the number of clusters of two kinds of glass, so as to obtain the number of subclasses of each kind of glass and the basis for division.

2.3 Analysis of Question 3

Through the data results obtained in problem 1 and Problem 2, the data were trained in problem 3 to establish a logistic regression model (stepdown method), and then the data was imported into SPSS Pro software to solve the established logistic regression model. Finally, the prediction and evaluation result table of the test data was output.

2.4 Analysis of Question 4

It is required to analyze the chemical composition of glass cultural relics samples of different categories, analyze the correlation between them, and compare the differences of the chemical composition correlation between different categories. Firstly, correlation analysis was used to analyze each chemical component of high potassium glass and lead barium glass, and the correlation matrix between them was obtained, so as to study the correlation between each chemical component. After the correlation is obtained, independent T-test is carried out for each chemical component of the glass cultural relic samples, so as to study the differences between the chemical components of the two types of glass.

3. Model building and solving

3.1 Creation and solution of Problem 1

3.1.1 Correlation analysis between surface weathering of cultural relics and indexes

1) Establishment of Chi-square test model

In order to study the relationship between the surface weathering of glass relics and its glass type, ornamentation and color, a Chi-square test model was established for the above variables. The chi-square test process is shown in the figure below:

2) Solving the Chi-square test model

Through Python programming and SPSS data analysis, the significant difference of chi-square test was analyzed. When P value is less than 0.05, it is significant. When P value is greater than 0.05, there is no significant effect. The final analysis results of chi-square test are shown in Table 1 below:

Table 1. Analysis results of Chi-square test

Name	No weathering	weathering
C (ornamentation)	13	15
A (ornamentation)	11	9
B (ornamentation)	0	6
High potassium (type)	12	6
Barium lead (type)	12	24
Blue-green (color)	6	9
Light Blue (color)	8	12
Purple (color)	2	2
Dark green (color)	3	4
Navy (color)	2	0
Light green (color)	2	1
Black (color)	0	2
green	1	0

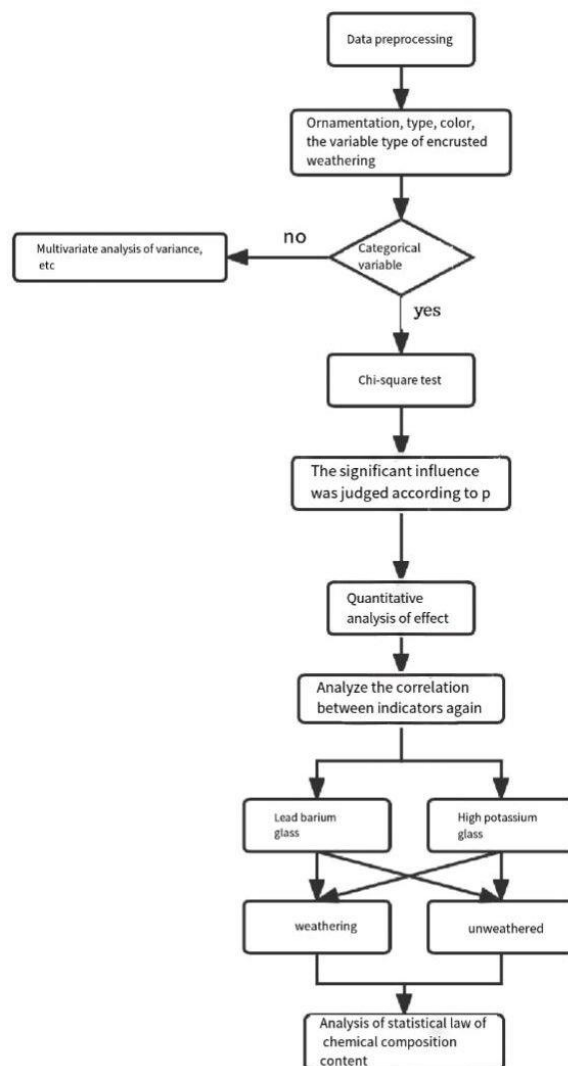


Figure 1. Chi-square test flow chart

In the above table, ***, ** and * represent the significance level of 1%, 5% and 10% respectively. From the analysis results of Chi-square test in the above table, it can be seen directly that the P value of the surface weathering of glass cultural relics and its glass type is 0.020**, less than the critical value of 0.05, indicating that there is a significant difference between the surface weathering of glass cultural relics and its glass type. The P value of the surface weathering of glass relics and its ornamentation is 0.056*, which is greater than the critical value of 0.05, indicating that there is no significant difference between the surface weathering of glass relics and its ornamentation. The P value of the surface weathering of glass cultural relics and its color is 0.507, which is greater than the critical value of 0.05, indicating that there is no significant difference between the surface weathering of glass cultural relics and its color.

By visualizing the above results, the chi-square cross thermal map can be obtained. The darker the color, the larger the value of the cross cross lexicon is. The lighter the color, the smaller the value of the cross contingency table.

Figure 2 shows the relationship between the surface weathering of glass relics and the type of glass.

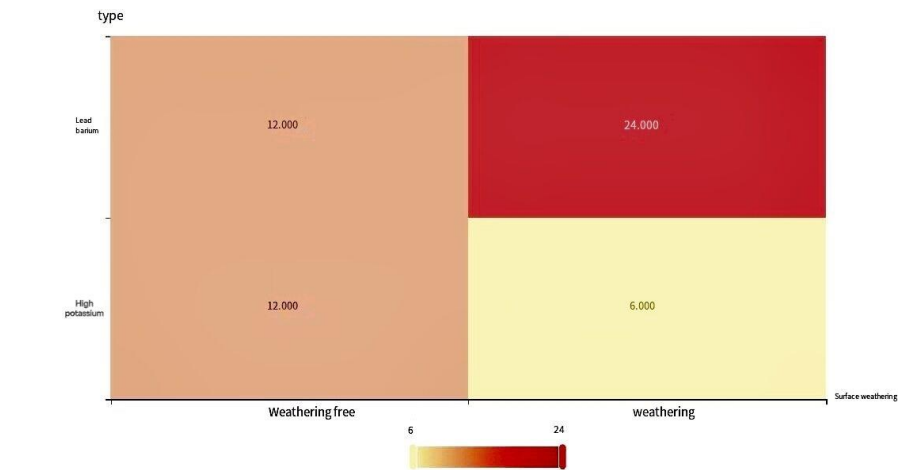


Figure 2. Surface weathering - type heat map

Figure 3 shows the relationship between the surface weathering of glass relics and its ornamentation.

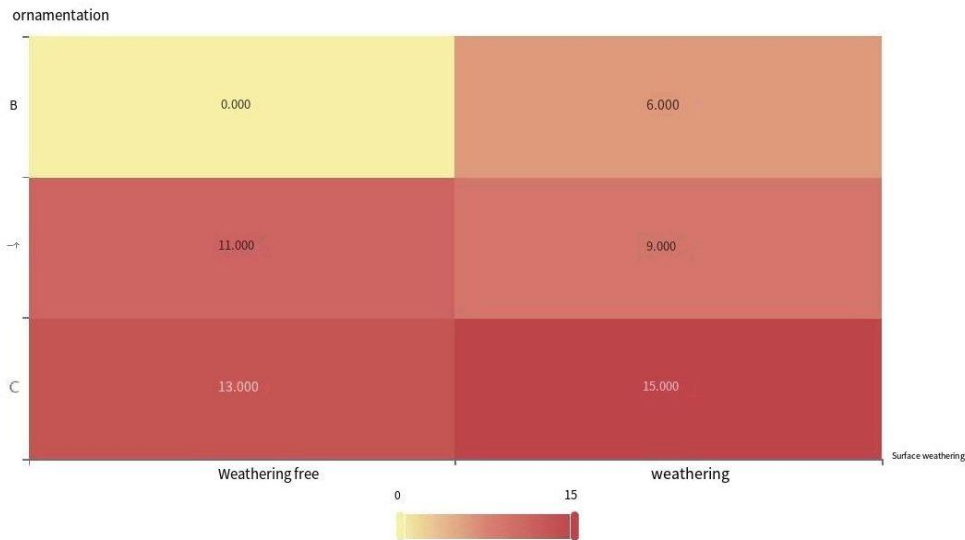


Figure 3. Surface weathering - ornamentation heat map

Figure 4 shows the thermal map relationship between weathering and color of glass relics.

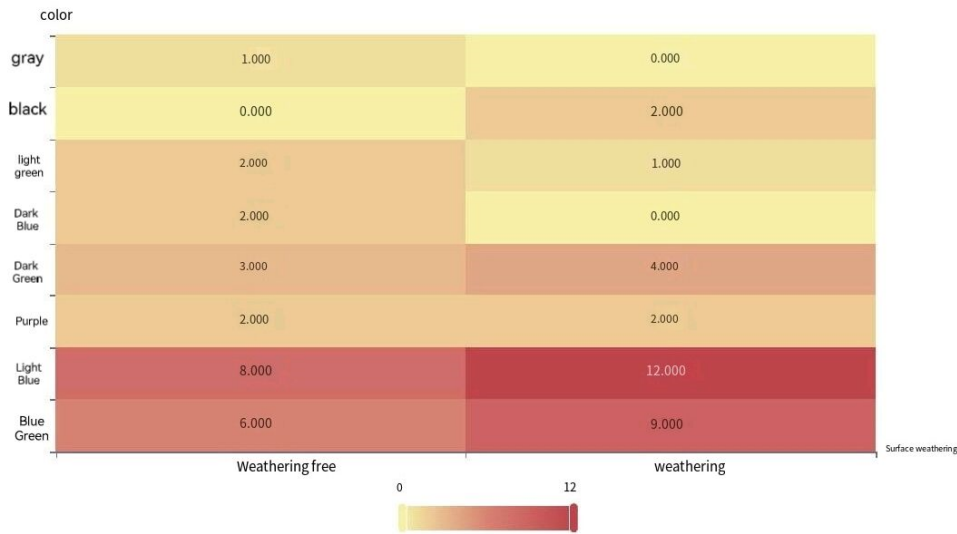


Figure 4. Surface weathering - color heat map

Based on the above quantitative effect analysis, the following conclusion is drawn: the Cramer's V value of glass type is 0.316, so the surface weathering of glass relics is moderately different from

that of glass type. The Cramer's V value of ornamentation is 0.326, so the surface weathering and ornamentation of glass relics are moderately different. The Cramer's V value of the color is 0.341, so the surface weathering and color of the glass relics are moderately different.

3.1.2 Statistical law of chemical composition content

According to the above analysis of the relationship between the surface weathering of glass cultural relics and their glass type, ornamentation and color, the average value of each chemical composition of the cultural relics sample before and after weathering was obtained, and the statistical law of the content change before and after weathering of the same chemical component was obtained, as shown in Figure 5 below :

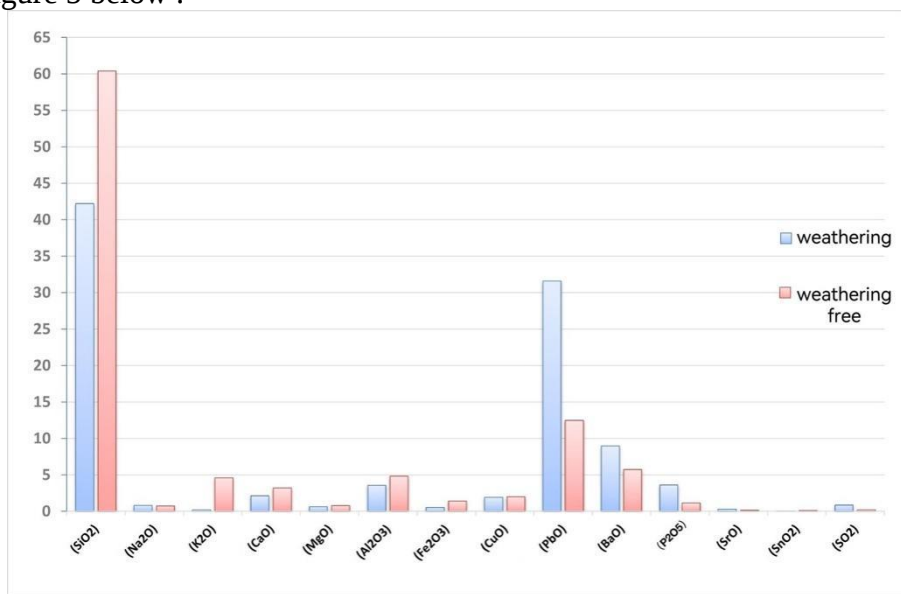


Figure 5. Variation of chemical composition of cultural relic samples before and after weathering

Furthermore, SPSS Pro software was used to obtain the analysis results of classification and summary (mean), that is, the statistical rule of the chemical composition content of high potassium and lead barium after weathering together, as shown in Figure 6 below.

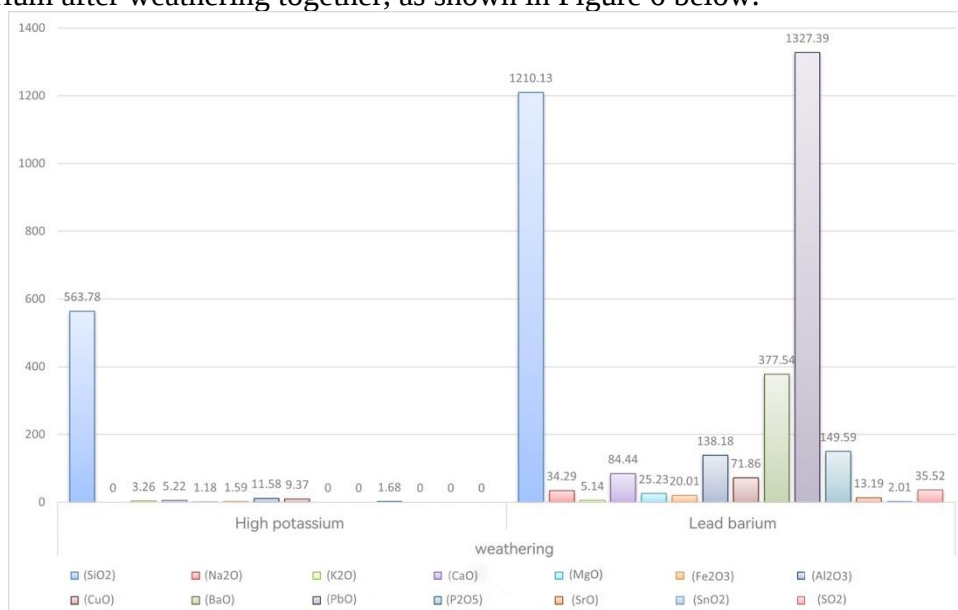


Figure 6. Chemical constituents of high potassium and Pb barium after weathering

3.1.3 The establishment of grey model

On the basis of the known significant relationships and statistical rules obtained above, we select the evaluation index of the chemical composition content of cultural relic samples before weathering

to obtain a quantifiable index body system. Grey prediction model is used to evaluate and calculate the chemical composition content of cultural relic samples before weathering based on the index system.

Let $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ be the original non-negative data column, and we add it once to get the new generated data column $x^{(1)}$,

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$$

Among them, the $x^{(1)}(m) = \sum_{i=1}^m x^{(0)}(i)$, $m = 1, 2, \dots, n$ Order, $z^{(1)}$ is the sequence, $x^{(1)}$: The immediate approximation of the means generates a sequence, that is, $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$, where:

$$z^{(1)}(m) = \delta x^{(1)}(m) + (1 - \delta)x^{(1)}(m - 1), m = 2, 3, \dots, n \text{ and } \delta = 0.5$$

We called $x^{(0)}(k) + az^{(1)}(k) = b$ is GM(1,1)'s basic form of the model. ($k = 2, 3, \dots, n$), Among them, b Indicates the amount of gray action, $-a$ indicates the coefficient of development. Below we introduce the matrix form:

$$u = (a, b)^T, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

so, GM(1,1) model $x^{(0)}(k) + az^{(1)}(k) = b$ can be expressed as,

$$Y = Bu$$

The least squares method is used to obtain a, b an estimate of the parameter

$$\hat{u} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (B^T B)^{-1} B^T Y$$

Next, treat the sequence as the dependent variable and the $x^{(0)}$ sequence as the $y, z^{(1)}$ independent variable x , and perform regression.

The established gray premodel is solved by Matlab, and some of the prediction data obtained.

3.2 Establishment and solution of problem 2

5.2.1 Principal component analysis dimensionality reduction treatment chemical index

1) Standardized treatment

Assuming that there are n samples and p indicators, a sample matrix x of $n \times p$ size can be formed, the mean and standard deviation are calculated by column, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ and $S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$ the standardized data is calculated $X_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$. The original sample matrix is normalized to:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} = (X_1, X_2, \dots, X_p)$$

2) Calculate the covariance matrix for standardized samples thereinto

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) = \frac{1}{n-1} \sum_{k=1}^n X_{ki} X_{kj}$$

3) Calculate the eigenvalues and R eigenvector eigenvalues of, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ (R is a semipositive definite matrix, and $\text{tr}(R) = \sum_{k=1}^p \lambda_k = p$) The eigenvectors are:

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

4) To calculate the principal component contribution rate and the cumulative contribution rate, set the contribution

5) The principal components of high-potassium glass and lead-barium glass were calculated, and the eigenvalues, corresponding eigenvectors and contribution rate of the correlation coefficient matrix of high-potassium glass were listed in the following table 4.

Table 2 PCA results for high-potassium glass

Feature vectors	a_1	a_2	a_3	a_4	a_5	a_6
SiO₂	-0.397	-0.176	-0.0845	-0.066	-0.034	-0.079
Na₂O	0.098	0.492	-0.136	0.311	-0.272	0.150
K₂O	0.302	0.276	0.183	0.246	0.167	0.232
CaO	0.276	0.435	0.074	-0.085	0.051	-0.015
MgO	0.311	-0.271	0.224	0.096	0.191	-0.364
Al₂O₃	0.374	0.009	-0.024	0.127	-0.284	-0.320
eigenvalue	5.642	2.627	1.749	1.631	0.959	0.663
Contribution rate	0.403	0.188	0.125	0.117	0.068	0.047
Cumulative contribution rate	0.403	0.591	0.716	0.832	0.901	0.948

Similarly, we can also obtain the eigenvalues, corresponding eigenvectors and contribution rate of the correlation coefficient matrix of lead-barium glass.

3.2.2 K-means clustering determines subclasses

1) Determine the number of clusters

The coefficient of polymerization of the system to obtain high potassium glass is shown in Figure 7: Aggregation coefficient

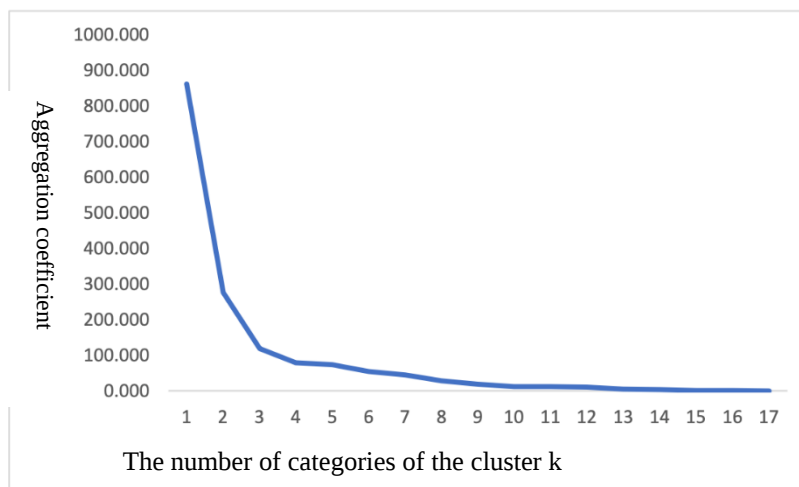


Figure 7 Polymerization coefficient of high potassium glass

The resulting systematic polymerization coefficient of lead-barium glass is shown in Figure 8

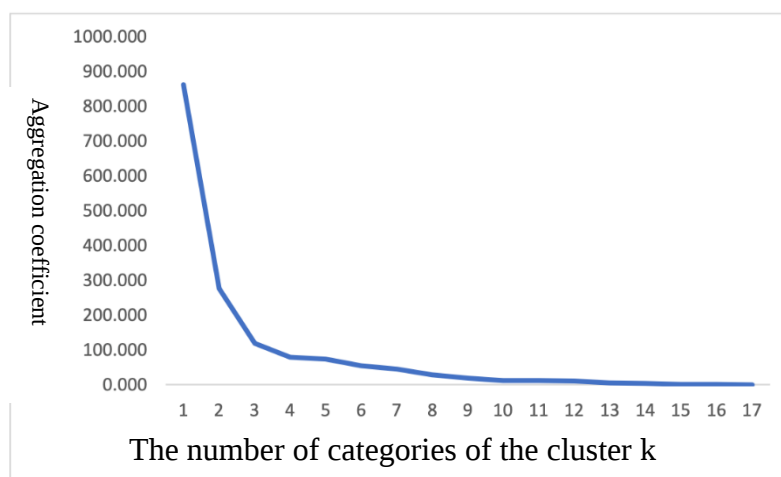


Figure 8 Polymerization coefficient of lead-barium glass

2) K-means cluster analysis

K-means clustering was performed on the data of high-potassium glass, and the number of categories K was set to 4, and the resulting partial data is shown in Table 3 below:

Table 3 Clustering results of high-potassium glass:

Case number	Relic number	clustering	distance
1	7	3	1.256
2	9	3	2.175
3	10	3	3.893
4	12	3	1.365
5	22	3	1.002
6	27	3	1.237

Similarly, K-means clustering was performed on lead barium glass, and the number of categories K was set to 3, and the results obtained were shown in Table 4:

Table 4 Prediction results of lead barium glass:
Case number Relic number clustering distance

1	2	1	5.485
2	8	3	4.346
3	8	3	11.823
4	11	1	2.55
5	19	1	2.295
6	23	2	8.455
7	25	2	8.604

3.3 Establishment and solution of the third problem

On the basis of the foreshadowing of the first two questions, with the help of the above data, in order to identify the category of glass artifacts at the end of the knowledge, through the training set data, we establish the logistic regression (gradient descent) model as follows:

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$z = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n = \sum_{i=0}^n \theta_i x_i = \theta^T x$$

3.3.2 Solution of logistic regression model

Import the combined data from question one into the SPSS Pro software and output part of the model parameter results .Finally, some of the data for the model's prediction results are obtained .

3.3.3 Sensitivity analysis of classification results

The sensitivity of the state or output changes of the model established to test the changes in the parameters or surrounding conditions of this system, that is, when a certain parameter in the model changes, the model fluctuates with its parameters, so sensitivity analysis is carried out.

The confusion matrix presented in the form of a heat map is shown in Figure 9 below:

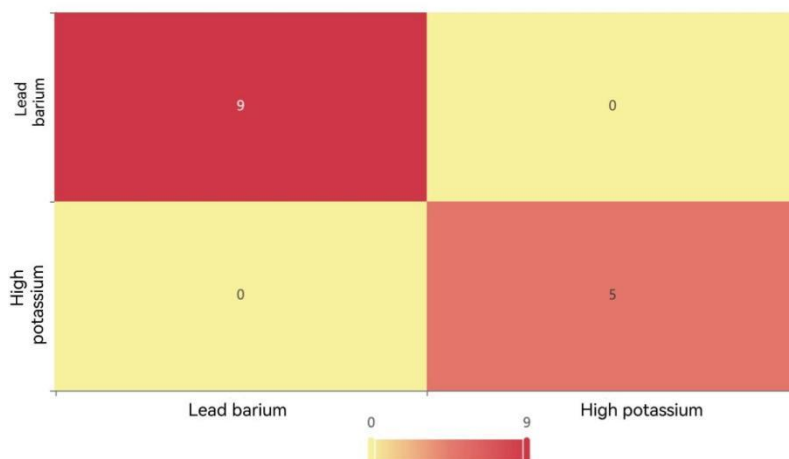


Figure 9 Test Data Confusion Matrix Heat Diagram

The data indicators that can be quantified from the above chart analysis can measure the predictive effect of logistic regression. Among them, the evaluation data index of the cross-validation set can continuously adjust the hyperparameters to obtain a reliable and stable model.

3.4 Establishment and solution of problem four

3.4.1 Correlation analysis model solves chemical composition associations

In order to study the correlation between multiple indicators between variables and better reveal the intrinsic relationship between variables, we establish a typical correlation analysis model:

Let's assume that the two sets of variables are:

$$X^1 = (X_1^1, X_2^1 \dots X_p^1), X^2 = (X_1^2, X_2^2 \dots X_q^2)$$

Next, we select several representative composite variables from two sets of variables U_i 、 V_i so that each composite variable is a linear combination of the original variable, i.e.

$$U_i = a_1^{(i)} X_1^{(1)} + a_2^{(i)} X_2^{(1)} + \dots + a_p^{(i)} X_p^{(1)} \equiv \mathbf{a}^{(i')} \mathbf{X}^{(1)}$$

$$V_i = b_1^{(i)} X_1^{(2)} + b_2^{(i)} X_2^{(2)} + \dots + b_q^{(i)} X_q^{(2)} \equiv \mathbf{b}^{(i')} \mathbf{X}^{(2)}$$

The first set of variables needs to $\text{var}(U_1) = \text{var}(V_1) = 1$ find and a^1 b^1 two sets of coefficients under the conditions of , such that $\rho(U_1, V_1)$ maximum. In order for the information of the

second group to be more effective, it is necessary to ensure that the information of the two groups is not relevant, i.e.

$$\text{cov}(U_1, U_2) = \text{cov}(V_1, V_2) = 0$$

And because the correlation coefficient is dimensional, so . $\rho(U_1, V_1) = \rho(aU_1, bV_1)$

Through Python programming and SPSS Pro analysis, the positive and negative directions and correlation degree of the correlation coefficient were analyzed, and then the correlation coefficients of unweathered high potassium and non-weathered lead barium were solved, respectively.

If it is significant, there is a correlation between the variables, and vice versa, there is no correlation between the variables. Finally, the heat map generated by the correlation coefficient is as follows:

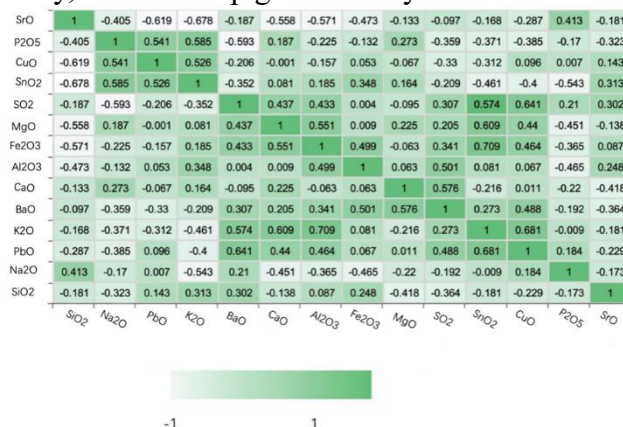


Figure 10 .High potassium heat map with correlation coefficient

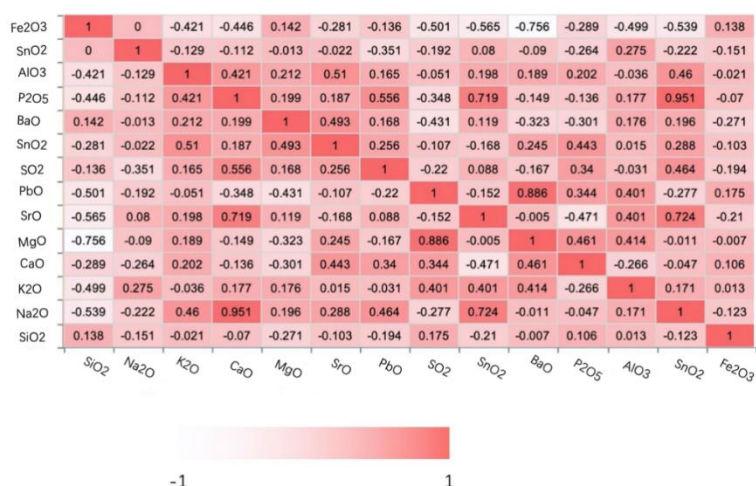


Figure 11. Correlation coefficient lead-barium heat diagram

3.4.2 Independent t-test solution

The results of the t-test analysis are shown in Table 5 below:

	Lead barium ($n = 49$)	High potassium ($n = 18$)	t	p
(SiO_2)	38.88 ± 18.65	76.64 ± 14.47	-7.764	0.000**
(Na_2O)	0.90 ± 1.81	0.46 ± 1.09	0.968	0.337
(K_2O)	0.17 ± 0.28	6.40 ± 5.31	-4.976	0.000**
(CaO)	2.05 ± 1.63	3.84 ± 3.31	-2.205	0.039*
(MgO)	0.65 ± 0.63	0.79 ± 0.71	-0.776	0.441
(Al_2O_3)	3.67 ± 3.01	5.06 ± 3.08	-1.665	0.101
(Fe_2O_3)	0.66 ± 0.95	1.38 ± 1.57	-1.832	0.081
(CuO)	1.88 ± 2.47	2.16 ± 1.49	-0.444	0.659
(PbO)	33.35 ± 14.95	0.27 ± 0.51	15.464	0.000**
(BaO)	10.49 ± 8.33	0.40 ± 0.84	8.364	0.000**
(SO_2)	0.80 ± 3.14	0.07 ± 0.16	1.627	0.110
(SnO_2)	0.06 ± 0.21	0.13 ± 0.56	-0.782	0.437
(SrO)	0.35 ± 0.26	0.03 ± 0.04	8.202	0.000**
(P_2O_5)	3.29 ± 3.91	1.03 ± 1.28	3.567	0.001**

Where. * $p < 0.05$ ** $p < 0.01$.

From the above table, it can be seen that different types of samples are for sodium oxide (Na_2O), magnesium oxide (MgO), alumina (Al_2O_3), iron oxides (Fe_2O_3), copper oxide (CuO), sulfur dioxide (SO_2), tin oxide (SnO_2) total 7 Terms do not exhibit significance ($p > 0.05$).

Specific analyses such as:

The glass type showed a level of significance of 0.01 for silica (SiO_2) ($t = -7.764$, $p = 0.000$), and the specific comparison difference showed the mean value of lead barium (38.88), which would be significantly lower than the average for high potassium (76.64).

In summary, different types of samples were used for sodium oxide (Na_2O), magnesium oxide (MgO), alumina (Al_2O_3), iron oxide (Fe_2O_3), copper oxide (CuO), sulfur dioxide (SO_2), tin oxide (SnO_2) total 7 Terms do not show significant differences in other types of samples for silica (SiO_2), potassium oxide (K_2O), Calcium oxide (CaO), lead oxide (PbO), barium oxide (BaO), strontium oxide (SrO), phosphorus pentoxide (P_2O_5) totals 7 Items show significant differences.

4. Evaluation of the model

The model describes the problem to a certain extent accurately and cleverly, and it is quite simplified and easy to understand. In view of the small sample size given by this problem, in the first

question, we choose a gray prediction model with less modeling information and convenient calculation, which greatly improves the modeling accuracy. However, at the same time, there are many complex factors in the model, and it cannot be comprehensively considered, which will cause certain inconsistencies with the reality.

References

- [1] Huang Qishan Research on Han Dynasty glass products in Beibu Gulf area of Guangxi[A]. Guangxi Museum Association, Guangxi Zhuang Autonomous Region Museum. Museum Collections Build a Bridge of Communication-Proceedings of the First Academic Symposium of Guangxi Museum Association and the Seventh Academic Symposium of Guangxi Zhuang Autonomous Region Museum [C]. Guangxi Museum Association, Guangxi Zhuang Autonomous Region Museum: Guangxi Zhuang Autonomous Region Museum, 2014:166-190
- [2] Gan Fuxi, Zhao Hongxia, Li Qinghui, Li Ling , Cheng Huansheng Scientific and technological analysis and research of Warring States glass products unearthed in Hubei Province [J] Jiangnan Archaeology, 2010(02): 108-116+151+0
- [3] Wen Rui, Zhao Zhiqiang, Ma Jian, Wang Jianxin Composition analysis of glass beads excavated from the Renzigou site group of Balikunshi in Xinjiang[J]. Spectroscopy and Spectral Analysis, 2016, 36(09): 2961-2965
- [4] Sun Quansen, Zeng Shengen, Wang Pingan , Xia Deshen Theory of Typical Correlation Analysis and Its Application in Feature Fusion [J]. Chinese Journal of Computers, 2005(09): 1524-1533
- [5] Wen Lihua. Grey system theory and its application [D]. Harbin Engineering University, 2003
- [6] Wang Jun, Wang Shitong, Deng Zhao Hong Some problems in cluster analysis research [J]. Guangdong Chemical Industry, 2021, 48(21): 85- 87+92.
- [7] Mo Shisong, Wang Jinglong, Pu Xiaolong, et al Advanced Mathematical Statistics (2nd Edition)[M]. Beijing: Higher Education Press, 2006
- [8] Liu Yue, Hao Shuxin, Song Jie, et al Research on rapid cleaning and automatic classification and aggregation method of outpatient emergency data in time series analysis of the relationship between air pollution and disease[J]. Health Res (4): 109-115