Composition analysis and identification of ancient glass products

Hongbiao Zeng *

College of Computer Science, Guangdong University of Science and Technology, Dongguan, China

* Corresponding Author Email: 370823182@qq.com

Abstract. Ancient glass products were the main commodities exchanged between ancient China and the West. In this paper, focusing on the analysis and identification of ancient glass products, based on the annex data mining analysis, Pearson correlation analysis, chi-square analysis and one-way ANOVA, a cluster analysis model was established to derive the classification rules of high potassium glass and lead-barium glass, and a K-Means model was established to classify the samples.

Keywords: Glass; One-way ANOVA; Cluster analysis; K-Means model; Correlation analysis.

1. Introduction

"Silk Road" is an important bridge between the economic and cultural exchanges between China and the West, its initial role is to transport the silk produced in ancient China, and later glass as an important item for trade between the two sides. After the introduction of glass products into China, China's glass on the basis of its improved production components, so the appearance of our ancient glass and foreign glass products are similar, but its chemical composition is not the same. After ancient glass is weathered by the environment, its composition content will change, thus affecting the judgment of its category.

Glass is also an indispensable material for us in life, windows of various buildings, light bulbs, and the glass water bottles and glasses we usually use, etc. are all glass. As early as the Warring States period in China, a special kind of glass, lead-barium glass, was discovered in China [1], but there is no indication how it was made, and even now, some people are still studying the formula of lead-barium glass [2], and the research on the chemical composition of glass does not stop nowadays when technology is developing rapidly and glass is being used more widely [3-4].

And with the development of artificial intelligence, more and more technologies are incorporated into machine learning, and cases of machine learning abound in life, such as the epidemic that is happening, and also many people are studying through machine learning methods, Zhiang Zhang [5] et al. used daily electricity consumption data of 87 regions in Zhejiang Province for two years, and conducted K-means cluster analysis on the time series of the region, thus quantify the impact of COVID-19 on electricity consumption patterns in Zhejiang through clustering evaluation indicators, dynamic time distortion, and finally explain the impact of COVID-19 through economic and social parameters, such as Sadeghi Banafsheh [6] et al. divided the time into 24 semimonthly periods and COVID-19 patient and mortality data into five groups according to each semimonthly period to perform a hierarchical cluster point analysis, and finally concluded that COVID-19 mortality can be a good representation of the country's resources and the ability of the system to re-operate in managing larger epidemics, as Lee Eunyoung Emily [7] et al. did by performing K-means cluster analysis on 105 patients divided into three groups according to cells within the body, and then unbiased cluster analysis based on 105 patients into three groups, and finally gave an opinion to improve the clinical outcome of COVID-19. Not only that, machine learning methods are also useful in various fields, for example, in diet Silva Vanderlei Carneiro (2022) [8] and others used K-means clustering algorithms, complemented by algorithms such as SVM, to predict the dietary patterns of more than 10,000 healthy Brazilian adults, thus classifying them into 2 categories with a high degree of accuracy, providing health professionals with a scientific argument for analyzing This provides a

scientific argument for health professionals to analyze dietary data. At the social electricity level, Liao Changyi [9] et al. by classifying 30 provinces and K-means clustering into 5 clusters based on CO2 emission characteristics, finally concluded that for most provinces industrial electricity emissions contribute the most to CO2, also providing insights into decarbonization paths for the provincial electricity sector. In terms of student learning, Yu Li [10] and others proposed a fuzzy C-mean clustering algorithm using two-dimensional and three-dimensional clustering to assess students' performance based on their test scores, and finally concluded that the experimental results could better provide educators with teaching methods for students, and even more, some optimized as well as promoted the cluster analysis algorithm, such as Ruxin Zhao [11] and others pointed out that the algorithm execution process is easy to fall into the local optimal solution, thus SMSHO algorithm is proposed to increase the diversity of the algorithm population to improve the algorithm global search ability, and finally SMSHO is verified to have faster convergence speed as well as higher accuracy through the data set, such as Weifeng Wang [12] et al. The implementation method, procedure, and single and multi-exponential of the panel cluster analysis of R software operation steps were investigated.

In this paper, cluster analysis is used for the given chemical composition of the glass to make a better classification for the glass classification.

2. Problem Analysis

2.1. Analysis of Question 1

The first sub-question: belongs to the mathematical problem of correlation analysis, which requires the analysis of the relationship between the weathering of the glass surface and its glass type, ornamentation and color; the card test is used because the variables are non-parametric tests. According to the data given in Annex Form 1, the cardinality test is used to determine the variability among the indicators and to observe the relationship between the data of these indicators.

The second sub-question: belongs to the mathematical problem of data analysis, which requires inductive analysis of the proportion of each chemical composition content with and, on the basis of the glass category, statistics of its surface with and without weathering chemical composition. The third sub-question, on the basis of the second question, calculates the loss rate of weathering chemical composition of different types of glass, so as to calculate its chemical composition content before weathering.

2.2. Analysis of Question 2

Sub-question 1: The question asks for an analysis of the patterns that exist in the chemical composition content of high potassium glass and lead-barium glass, based on their classification according to these two types of glass. This requires that we first distinguish high potassium glass from lead-barium glass from Annex 1. Operation A one-way ANOVA was used to conclude that silica content had a significant effect on the classification of the glasses. Therefore, the pattern was analyzed using silica content as the main indicator.

The second sub-question requires the selection of suitable chemical elements for subclassification, while in the first sub-question it is stated that silica content has a significant effect on the classification of glasses. Therefore, cluster analysis was applied to delineate four tables (see Annex 2).

The third sub-question: the requirement to verify the reasonableness and sensitivity of the results of the second sub-question, using the data of the division results of the second sub-question brought into the classification law of the first sub-question, can be obtained in full compliance with its law therefore, to prove that the division results are reasonable. And then by adding data to determine whether the resulting results occur in large fluctuations, if not, the sensitivity is weak.

3. Model building and solving

3.1. Problem 1 Model building and solving

3.1.1. Cardinality test modeling

First, we start by sorting and summarizing the data in Annex Form 1, as shown in Table 1.

Table 1. Classification statistics of weathered and unweathered glass according to each index.

Title	Name	Surface weathering(%)		Total
		No weathering	Weathering	
Ornamentation	A	11(45.83)	11(32.35)	22(37.93)
	B	0(0.00)	6(17.65)	6(10.34)
	C	13(54.17)	17(50.00)	30(51.72)
Total		24	34	58
Color	light green	2(8.33)	1(3.33)	3(5.56)
	Pale Blue	8(33.33)	12(40.00)	20(37.04)
	Dark Green	3(12.50)	4(13.33)	7(12.96)
	Deep Blue	2(8.33)	0(0.00)	2(3.70)
	Purple	2(8.33)	2(6.67)	4(7.41)
	Green	1(4.17)	0(0.00)	1(1.85)
	Blue-Green	6(25.00)	9(30.00)	15(27.78)
	Black	0(0.00)	2(6.67)	2(3.70)
Total		24	30	54
Type	Lead Barium	12(50.00)	28(82.35)	40(68.97)
	High Potassium	12(50.00)	6(17.65)	18(31.03)
Total		24	34	58

A chi-square test was then established to analyze the relationship between surface weathering and its glass type, ornamentation and color. The chi-square test is a non-parametric test that focuses on comparing two and more sample rates (composition ratios) and the analysis of variance of two categorical variables, calculated as follows.

$$X^2 = \frac{H - T^2}{T} \tag{1}$$

Where: A is the actual statistical value of each indicator, H is the theoretical value of each indicator, and X^2 is: a measure of the degree of difference between theory and practice

$$p = \int_{X^2}^{\infty} f(x)dx \tag{2}$$

p is the significance value.

3.1.2. Solution of the chi-square test model

The results of the chi-square test between weathering and glass type, decoration and color on the surface of glass products were calculated according to equation (2) using SPSS software, and the results are shown in Table 2.

Table 2. Results of the chi-square test between surface weathering and glass type, decoration, and color

Title	Name	Surface weathering(%)		Total	X^2	*p*
		No weathering	Weathering			
Ornamentation	A	11(45.83)	11(32.35)	22(37.93)	4.957	0.084
	B	0(0.00)	6(17.65)	6(10.34)		
	C	13(54.17)	17(50.00)	30(51.72)		
Total		24	34	58		
Color	light green	2(8.33)	1(3.33)	3(5.56)	6.287	0.507
	Pale Blue	8(33.33)	12(40.00)	20(37.04)		
	Dark Green	3(12.50)	4(13.33)	7(12.96)		
	Deep Blue	2(8.33)	0(0.00)	2(3.70)		
	Purple	2(8.33)	2(6.67)	4(7.41)		
	Green	1(4.17)	0(0.00)	1(1.85)		
	Blue-Green	6(25.00)	9(30.00)	15(27.78)		
	Dark	0(0.00)	2(6.67)	2(3.70)		
Total		24	30	54		
Type	Lead Barium	12(50.00)	28(82.35)	40(68.97)	6.880	0.009**
	High Potassium	12(50.00)	6(17.65)	18(31.03)		
Total		24	34	58		
* $p<0.05$ ** $p<0.01$						

From the table, we can conclude the classification pattern of high potassium glass and lead-barium glass: the p-values between surface weathering of glass and its decoration and color are 0.084 and 0.507 respectively, which are both greater than 0.05. Therefore, there is no correlation between surface weathering of glass and its decoration and color. In contrast, the p-value between glass surface weathering and its type was 0.009, showing a significance at 0.01 level, so there was a significant correlation between surface weathering and its type.

The p-value between glass surface weathering and its decoration is 0.084, which is greater than 0.05, so there is no correlation between surface weathering and decoration. The p-value between glass surface weathering and its color is 0.507, which is greater than 0.05, so there is no correlation between surface weathering and color.

The p-value between glass surface weathering and its type was 0.009, showing a significance at the 0.01 level, so there was a significant correlation between surface weathering and type.

3.1.3. Constructing and solving the component turnover rate model

Step 1: In this question, the paper first averages the chemical composition contents corresponding to different types of glass weathering and unweathering, respectively, as in Table 3.

Table 3. Average values of chemical composition content corresponding to different types of weathered and unweathered glass.

Type	Degree of weathering	SiO_2	Na_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
High Potassium	Unweathered	67.98	0.70	5.33	1.08	6.62	1.93	2.45	0.41	0.60	1.40	0.04	0.20	0.10
High Potassium	Weathering	93.96	0.00	0.87	0.20	1.93	0.27	1.56	0.00	0.00	0.28	0.00	0.00	0.00
Lead Barium	Unweathered	54.66	1.68	1.32	0.64	4.46	0.74	1.43	22.08	9.00	1.05	0.27	0.05	0.16
Lead Barium	Weathering	24.91	0.22	2.70	0.65	2.97	0.58	2.28	43.31	11.81	5.28	0.42	0.07	1.37

Step 2: Visualization of the data in Table 3, yields the following results:

(1) High potassium glass, silica content will increase due to the influence of weathering. NaO, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide, iron oxide, copper oxide, lead oxide, barium oxide, phosphorus pentoxide, strontium oxide, tin oxide, sulfur dioxide content ratio will be reduced by the influence of weathering.

(2) lead barium glass, calcium oxide, magnesium oxide, copper oxide, lead oxide, barium oxide, phosphorus pentoxide, strontium oxide, tin oxide, sulfur dioxide content will increase due to the influence of weathering. The proportion of silicon dioxide, potassium oxide, sodium oxide, aluminum oxide, iron oxide, and content will decrease due to the influence of weathering.

Step 3: Calculate the loss rate of each chemical component with the following formula.

$$R = \frac{X_u - X_f}{X_u} \times 100\% \tag{3}$$

For the explanation of the variables in Eq. (3): is the composition loss rate, is the chemical composition content of unweathered glass, and is the chemical composition content of weathered glass. In summary, we can obtain

$$X_u = \frac{X_f}{1 - R} \tag{4}$$

According to equation (4), the chemical composition content before weathering can be predicted.

3.2. Solution and Analysis of Problem 2

3.2.1. One-way ANOVA modeling and solving

First, a one-way ANOVA was performed on the variables in Annex 4. One-way ANOVA was performed using SPSS software to obtain Table 4.

Table 4. Normality test results.

Variable Name	Sample size	Median	Average value	Standard deviation	Kurtosis	S-W test	K-S test
Silicon dioxide (SiO2)	67	51.33	49.022	24.316	-0.818	0.963(0.042**)	0.102(0.457)
Sodium oxide (Na2O)	67	0	0.786	1.653	6.106	0.554(0.000***)	0.429(0.000***)
Potassium oxide (K2O)	67	0.2	1.847	3.879	3.008	0.526(0.000***)	0.402(0.000***)
Calcium oxide (CaO)	67	1.66	2.532	2.325	0.214	0.879(0.000***)	0.154(0.076*)
Magnesium oxide (MgO)	67	0.64	0.683	0.65	-0.099	0.883(0.000***)	0.226(0.002***)
Aluminum oxide (Al2O3)	67	3.41	4.041	3.067	2.782	0.844(0.000***)	0.141(0.125)
Iron oxide (Fe2O3)	67	0.32	0.849	1.179	5.541	0.741(0.000***)	0.236(0.001***)
Copper oxide (CuO)	67	1.07	1.954	2.242	5.447	0.750(0.000***)	0.213(0.004***)
Lead oxide (PbO)	67	23.02	24.463	19.513	-0.951	0.929(0.001***)	0.148(0.097*)
Barium oxide (BaO)	67	6.65	7.779	8.425	2.736	0.806(0.000***)	0.189(0.015**)
Phosphorus pentoxide (P2O5)	67	1.04	2.684	3.544	1.389	0.764(0.000***)	0.262(0.000***)
Strontium oxide (SrO)	67	0.23	0.262	0.267	0.808	0.875(0.000***)	0.164(0.049**)
Tin oxide (SnO2)	67	0	0.078	0.337	34.673	0.249(0.000***)	0.502(0.000***)
Sulfur dioxide (SO2)	67	0	0.603	2.698	27.846	0.233(0.000***)	0.469(0.000***)

Analysis of Figure 1 shows that the significance of silicon dioxide (SiO_2) is significantly weaker than the other independent variables. According to the title, it is known that quartz sand is the main raw material of glass, and the main chemical composition is silicon dioxide (SiO_2), and potassium glass is fired with grass wood ash as a flux. Meanwhile, according to the literature, grass wood ash will be decomposed into carbon dioxide and K_2O due to heating, which leads to high K_2O content, and similarly, in lead-barium glass the content of PbO and BaO in lead barium glass is also high. Therefore, we use the average value of silica (SiO_2) content in each category as the judgment index.

Taking high potassium glass as an example, first, the average value of silica content ($\overline{x_1}$) and the average value of potassium oxide content ($\overline{x_2}$) for the sampling points of the high potassium class of artifacts were calculated using Equation (5).

$$x_1 = \frac{1}{T}\int_0^T x_1(t)dt \tag{5}$$

From the formula (5): the calculation can be obtained $\overline{x_1}$ = 76.6438888888889; $\overline{x_2}$ = 6.401666667. The maximum and minimum values of silica content are 96.77 and 59.01, respectively, and the maximum and minimum values of potassium oxide are 14.52 and 0, respectively, using the EXCEL table.

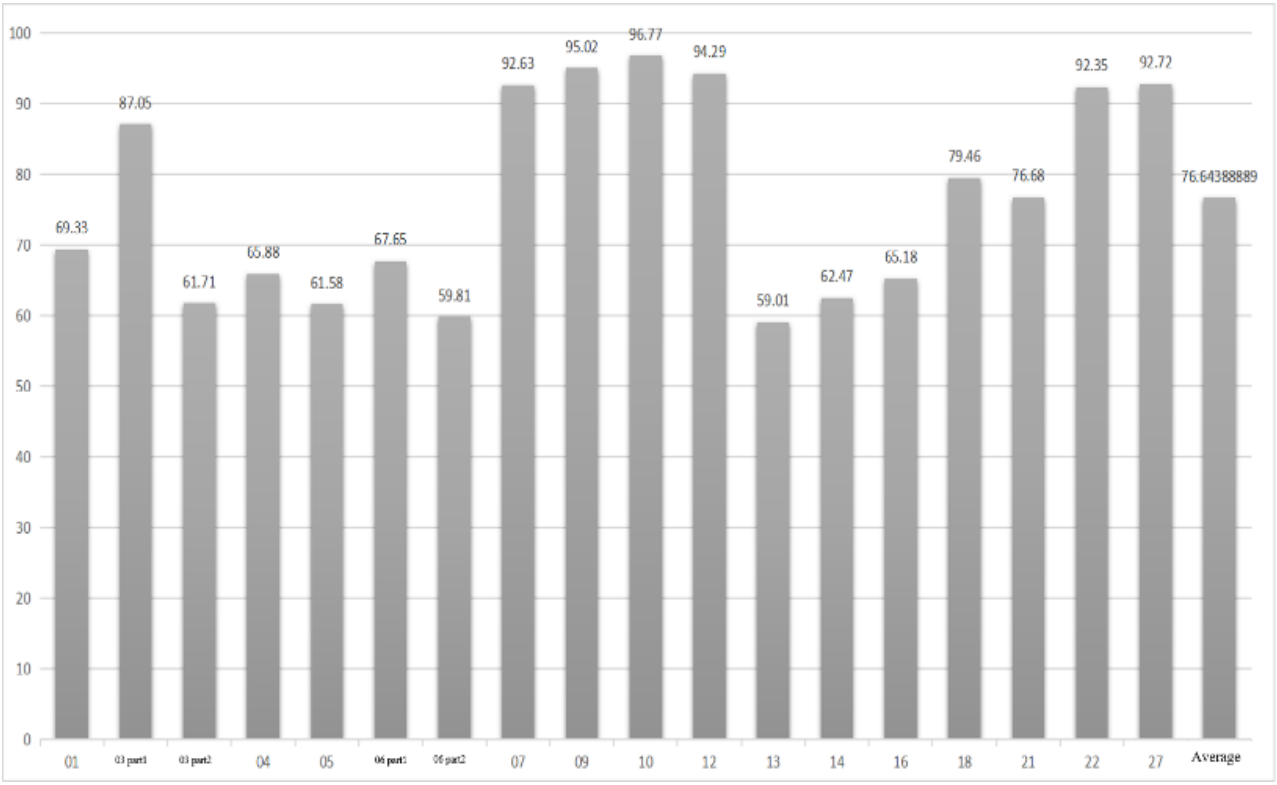


Figure 1. The average value of silica content and silica content of each artifact sampling site.

The average value of silica content was used to compare with the silica content of each artifact sampling site in Annex 1, and it can be obtained from Figure 1 that there are 9 artifact sampling sites with silica content higher than the average value; there are 9 artifact sampling sites with silica content lower than the average value. From this, it is concluded that the classification pattern of high potassium glass is related to potassium oxide, which is higher (higher or equal to the average value of potassium oxide content) than the average value of silica at the sampling sites of cultural relics. Potassium oxide content is low (lower than the average value of potassium oxide content) at the sites sampled above the average value of silica content.

Similarly, the average value of silica content in lead-barium glass artifacts is calculated to be 42.84262, which can be obtained from Annex 1, and the silica content of 20 groups of artifact sampling points in lead-barium class is higher than the average value of silica content of lead-barium

class artifacts, and the silica content of 22 groups of artifact sampling points is lower than the average value of silica content of lead-barium class artifacts. It can be concluded that the law of lead-barium glass classification is related to the sum of barium oxide content and lead oxide content. Lower than the average silica content of cultural relics sampled in the lead oxide content and the sum of barium oxide content is low (lower than the average of the sum of barium oxide and lead oxide content). The sum of lead oxide content and barium oxide content is high (higher than the average value of the sum of barium oxide and lead oxide content) in the sampling sites of relics with higher-than-average silica content.

3.2.2. Cluster analysis model building

Cluster analysis is a quantitative method of multivariate statistical analysis for quantitative classification of multiple samples (or indicators). Suitable chemical components need to be selected as subclass classification criteria for each category in the high potassium glass and lead barium glass artifact sampling sites. It is then necessary to classify each chemical composition quantitatively, and therefore it is necessary to describe the degree of similarity between chemical compositions quantitatively. Therefore, for the quantitative variables of chemical composition Minkowski distances are required.

$$d_q(x,y) = \left[\sum_{k=1}^{p}|x_k - y_k|^q\right]^{\frac{1}{q}}, q > 0 \tag{6}$$

$$d_2(x,y) = \left[\sum_{k=1}^{p}|x_k - y_k|^2\right]^{\frac{1}{2}} \tag{7}$$

Where, is the sample observation from the total dimension?

There is a strong correlation between the content of each chemical element, and the correlation of silica content is relatively high, while it is known that the main raw material of glass is quartz sand according to the title, and the main chemical composition of quartz sand is silica. Therefore, we consider the clustering analysis by the content of silica. The clustering tree pattern is shown in Figure 2.

3.2.3. Rational modeling

Annex 2, most of the sampling points of high potassium glass silica content and potassium oxide component relationship are in line with the classification law described in 3.2.1, below the average value of silica heritage sampling points of potassium oxide content is high (higher than or equal to the average value of potassium oxide content). Higher than the average value of silica content of potassium oxide content of the artifact sampling sites is low (lower than the average value of potassium oxide content). Therefore, the classification results can be judged as reasonable.

3.2.4. Sensitivity modeling

From 3.2.1, it can be obtained that in the classification law of high potassium glass types, the silica and potassium oxide content is closely related. This classification pattern is also found in lead-barium glass. With this, the sensitivity analysis of the cluster analysis model of 5.2.2 is used in this paper. The sensitivity of the results to the subclassification of chemical composition was demonstrated.

Six groups of high potassium types, three groups of weathering and three groups of no weathering were added to the classified data (see Annex 4), and the silica content of each control group was set to 105 for cluster analysis according to the classification law of high potassium glass and lead-barium glass.

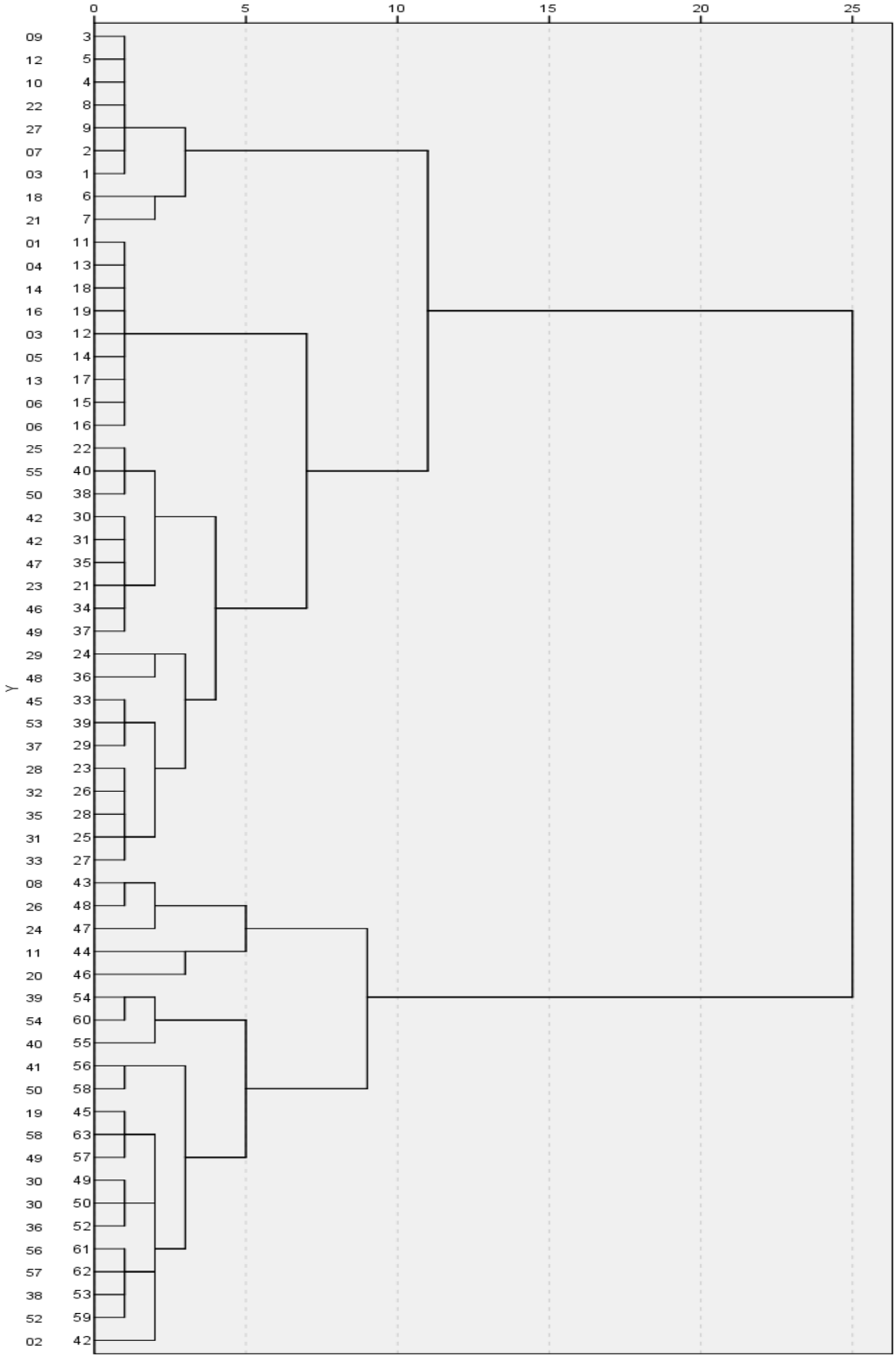


Figure 2. Indicator clustering tree diagram.

4. Conclusions

In this study, the correlation between weathering and type, decoration, color and chemical composition of glass surface was better analyzed, and the methods of chi-square test and Pearson

correlation coefficient were applied to make the analysis results more scientific and reasonable. The index data were screened to avoid the possible errors brought by irrelevant or wrong data on the results. The visualization of the data of various indexes enables a more intuitive observation of the relationship between various indexes and increases the reasonableness and sensitivity of the results.

The model developed in this paper uses methods that can be applied to other areas, such as the cluster analysis we used in solving Problem 2, which can also be used to classify plants and animals and to classify genes to gain insight into the intrinsic structure of populations.

References

[1] Wang Rong, Xu Yiwen, Zhou Xubo. Study on the related problems of lead-barium imitating-jade glass in the Warring States period of China [J]. Archaeological and Anthropological SciencesVolume 14, Issue 8. 2022.

[2] Qian Ma, Dennis Braekmans, Andrew Shortland, A. Mark Pollard. The Production and Composition of Chinese Lead-Barium Glass through Experimental Laboratory Replication [J]. Journal of Non-Crystalline SolidsVolume 551, 2021. PP 120409-.

[3] Niilisk Ahti, Haak Arvi, Ritslaid Peeter, Mäesalu Ain, Kikas Jaak. Chemical composition of potassium-glass vessels from the 14th-15th centuries found in Tartu, Estonia [J]. Journal of Archaeological Science: ReportsVolume 38, 2021.

[4] M. N. Andreev, A. A. Drozdov,V. D. Krasnobrov. Complex Research on the Composition and Microstructure of Historical Glass [J]. Moscow University Chemistry BulletinVolume 74, Issue 6. 2019. PP 330-338.

[5] Zhang Zhiang, Cheshmehzangi Ali,Ardakani Saeid Pourroostaei. A Data-Driven Clustering Analysis for the Impact of COVID-19 on the Electricity Consumption Pattern of Zhejiang Province, China [J]. Energies 2021, 14, 8187.

[6] Sadeghi Banafsheh, Cheung Rex C Y,Hanbury Meagan. Using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020[J].BMJ Open 2021.

[7] Lee Eunyoung Emily, Song Kyoung Ho, Hwang Woochang, Ham Sin Young, Jeong Hyeonju, Kim Jeong Han, Oh Hong Sang, Kang Yu Min, Lee Eun Bong, Kim Nam Joong, Chin Bum Sik, Park Jin Kyun.Pattern of inflammatory immune response determines the clinical course and outcome of COVID-19: unbiased clustering analysis[J].Scientifc Reports, Volume 11, Issue 1. 2021. PP 8080-8080.

[8] Silva Vanderlei Carneiro, Gorgulho Bartira, Marchioni Dirce Maria, Araujo Tânia Aparecida de, Santos Itamar de Souza, Lotufo Paulo Andrade, Benseñor Isabela Martins. Clustering Analysis and Machine Learning algorithms in the prediction of dietary patterns: Cross-sectional results of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) [J]. Journal of Human Nutrition and DieteticsVolume 35, Issue 5. 2022. PP 883-894.

[9] Liao Changyi, Wang Saige, Zhang Yiyi, Song Dan, Zhang Chaohai. Driving forces and clustering analysis of provincial-level CO2 emissions from the power sector in China from 2005 to 2015[J]. Journal of Cleaner ProductionVolume 240, Issue C. 2019. PP 118026-118026.

[10] Yu Li, Jin Gou, Zongwen Fan. Educational data mining for students' performance based on fuzzy C-means clustering [J]. The Journal of EngineeringVolume 2019, Issue 11. 2019. PP 8245-8250).

[11] Ruxin Zhao, Yongli Wang, Gang Xiao, Chang Liu, Peng Hu, Hao Li.A selfish herd optimization algorithm based on the simplex method for clustering analysis[J].The Journal of Supercomputing2021. PP 1-71.

[12] Weifeng Wang, Yugui Lu. Application of Clustering Analysis of Panel Data in Economic and Social Research Based on R Software [J]. Academic Journal of Business & ManagementVolume 3.0, Issue 10.0. 2021.