Research on glass classification based on linear regression model

Ru Wu [1, *], Yilin Zhang [2], Xukang Zhao [3]

[1] School of Management, Zhengzhou university, Zhengzhou, Henan, China, 450000

[2] School of Computer Sciece and Artificial Intelligence, Zhengzhou university, Zhengzhou, Henan, China, 450000

[3] School of Water conservation science and Engineering, Zhengzhou university, Zhengzhou, Henan, China, 450000

* Corresponding author: marswuru@163.com

Abstract. This paper studies the rule of chemical composition of ancient glass products and the identification of glass types. Through Spearman correlation analysis, the establishment of multiple linear regression model and Fisher linear discriminant model, the relationship and classification rule of chemical composition of different types of glass are determined, and the linear regression model of chemical composition content before and after weathering and the identification method of glass types are obtained. The results show that the mathematical model proposed in this paper can effectively identify glass types and provide guidance for the identification of glass components.

Keywords: Spearman correlation analysis, Multiple linear regression, Fisher linear discriminant model.

1. Introduction

In the Silk Road, glass is a precious symbol of this trade channel. The early glass was introduced into China from Western Asia and Egypt. The ancient glass in China was made locally. Although its appearance is similar to that of glass products, its chemical composition is quite different.

The protection of glass relics can be roughly divided into two categories: one is to control the environment of glass preservation to minimize its damage and decay; The other is to take measures to prevent the decay of glass when necessary to prevent further deterioration [1-3]. To effectively protect glass relics, it is necessary to determine the chemical composition of glass. Ancient glass is susceptible to weathering due to the influence of burial environment. In the weathering process, the internal elements further exchange a lot with the environmental elements, resulting in a slight change in their composition proportion, which affects the correct judgment of their categories.

In view of this, this paper studies the chemical composition of ancient glass products and glass types, which provides guidance for the identification of glass relics.

2. Prediction model of chemical composition content of glass relics before weathering

2.1. Relationship between surface weathering, glass type, decoration and color of glass relics

The Spearman correlation coefficient [4,5] of the corresponding data is obtained through analysis and calculation. According to the relationship between the Spearman correlation coefficient, the P value and R value between the decoration and weathering are analyzed, and the P value is greater than 0.05. From this, we know that there is no correlation between the decoration and weathering. Similarly, the P value between the color and weathering is greater than 0.05, and the P value between the glass type and weathering is less than 0.05, so there is correlation, at this time, the R value is -0.3444, so it is easy to know whether the ornamentation is related to weathering. To sum up, it can be concluded that the decoration and color are not related to weathering, so there is no clear correlation between them. For the glass type and weathering, it can be preliminarily judged that they are related, but the correlation is not strong, and there is a certain correlation between them. We can also conclude from the data that 73.47% of the lead barium glass will be weathered, and 33.33% of the high potassium glass will be weathered (Table 1, Figure 1).

Spearman correlation coefficient:

$$r = 1 - \frac{6\sum_{i=2}^{n} d_i^2}{n(n^2-1)} \tag{1}$$

Where d_i represents the rank difference of each pair of observations (x, y), and n represents the number of observation pairs.

Table 1. Spearman correlation coefficient

Symbol	Value			
R	1.0000	0.1190	0.4520	-0.0373
	0.1190	1.0000	0.4513	-0.3444
	0.4520	0.4513	1.0000	0.1057
	-0.0373	-0.3444	0.1057	1.0000
P	1.0000	0.3736	0.0004	0.7812
	0.3736	1.0000	0.0004	0.0081
	0.0004	0.0004	1.0000	0.4297
	0.7812	0.0081	0.4297	1.0000

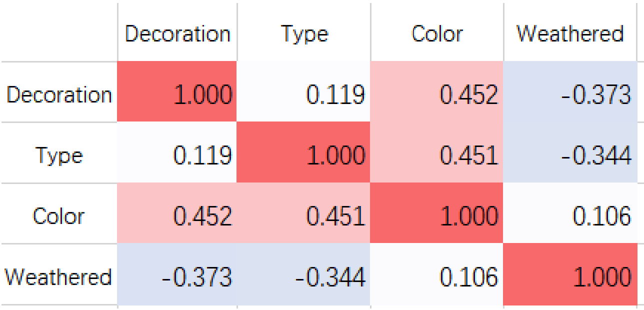

Figure 1. Thermal diagram of Spearman correlation coefficient

2.2. Whether the glass type cultural relics have the statistical law of chemical composition content of weathering

In order to obtain the statistical law of whether there is weathering chemical composition on the sample surface, this section establishes a multiple regression model [6-8], obtains the regression coefficient, compares the confidence factor, analyzes the relationship between the content of chemical composition and whether it is weathered, so as to speculate the content of chemical composition before weathering.

We choose silicon dioxide, sodium oxide, potassium oxide, calcium oxide and magnesium oxide as the main variables. At the same time, because some components will have strong correlation between the pre and post dummy variables in the linear regression analysis, we choose to discard the data of some components to ensure the accuracy of the data. With the silicon dioxide content, potassium content, silicon dioxide content, lead content and barium content in the high potassium glass as dependent variables, this paper sets the dummy variables A_i(i=1,2), B_i(i=1,2......7), D_i(i=1)) Establish a linear regression model:

$$y_1 = \beta_0 + \sum_{i=1}^{14} \beta_i x_{i,} + \sum_{i=1}^{2} a_i A_{i,} + \sum_{i=1}^{7} b_i B_{i,} \sum_{i=1}^{1} c_i D_{i,} \tag{2}$$

$$y_2 = \omega_0 + \sum_{i=1}^{14} \omega_i x_{i,} + \sum_{i=1}^{2} d_i A_{i,} + \sum_{i=1}^{7} e_i B_{i,} \sum_{i=1}^{1} f_i D_{i,} \tag{3}$$

y_1,y_2,y_3,y_4,y_5 respectively represent the content of corresponding chemical components, θ_0, θ_1, β_0, β_1, α_1 represent the regression coefficient, A_i represents the pattern type, B_i represents the color, D_i represents whether the surface is weathered, and x_i represents the content of other chemical components. At this time, we use stata to calculate the estimated regression coefficients (Table 2, 3):

Table 2. Regression Analysis of Silica Content in High Potassium Glass

Variable	Regression coefficient	Variable	Regression coefficient
MgO A1	4.449 0.000	P_2O_5 A2	-2.128 62.501***
A3	-1.356	D2	-31.771**
D1	0.000	cons	64.457***

Table 3. Regression analysis of potassium content in high potassium glass

Variable	Regression coefficient	Variable	Regression coefficient
Na_2O MgO	1.590 1.860	SiO_2 Al_2O_3	-0.259 -1.084
D1	28.996**	D2	0.000
cons	1.934		

$$y_1 = 64.457 + 62.501A_1 - 37.771D_1 \tag{4}$$

$$y_2 = 1.934 + 25.184A_1 + 28.996D_1 \tag{5}$$

$$y_3 = 114.995 - 43.116A_1 - 46.669A_3 \tag{6}$$

$$y_4 = 67.399 - 0.729x_1 - 2.19x_{13} \tag{7}$$

$$y_5 = 24.768 - 0.221x_1 + 0.958x_8 - 0.510x_{11} \tag{8}$$

Through the BP test using Stata to check whether there is heteroscedasticity, it is found that when calculating the regression coefficient of barium oxide content in lead barium glass, the P value is less than 0.05, indicating that the original error is rejected at the 95% confidence level, that is, we believe that there is heteroscedasticity in the disturbance term. In order to solve the heteroscedasticity problem, the c4 regression model is regressed with OLS+robust standard error.

$$y_5 = 64.457 + 62.501A_2 - 31.771D_2 \tag{9}$$

3. Classification model of glass chemical composition

3.1. Classification rules of high potassium glass and lead barium glass

First, we try to reduce the dimensions of the data by factor analysis of the given data. The analysis results show that factor analysis is not applicable. At the same time, due to sample imbalance, logical regression is not applicable. Therefore, we found that the given data met the sher linear discriminant analysis, and the results were reasonable (Table 4, 5).

Table 4. Classification Results

Type= Pb-Ba			Forecast group member information		total
		.00	.00	1.00	
original	count	.00	14	2	16
		1.00	1	37	38
		Group not classified	1	12	13
	%	.00	87.5	12.5	100.0
		1.00	2.6	97.4	100.0
		Group not classified	7.7	92.3	100.0

Table 5. Classification Function Coefficient

Type= Lead barium glass					
	.00	1.00		.00	1.00
SiO_2	33.941	34.091	CuO	19.618	19.162
Na_2O	37.185	37.861	PbO	33.651	33.909
K_2O	38.116	37.733	BaO	41.795	41.715
CaO	38.019	37.327	P_2O_5	43.844	43.832
MgO	48.036	49.537	SrO	5.255	12.094
Al_2O_3	28.895	29.104	SnO_2	24.571	27.204
Fe_2O_3	41.923	42.488	SO_2	16.602	16.941
Features	-1681.369			-1693.542	

3.2. Classification method

The hierarchical clustering method is used to divide the chemical components, and the number of clusters is determined according to the correlation between the main components of the two types of glass and the content range.

SPSS software was used for system Q clustering [9,10], and the data pedigree corresponding to all data was drawn (Figure 2):

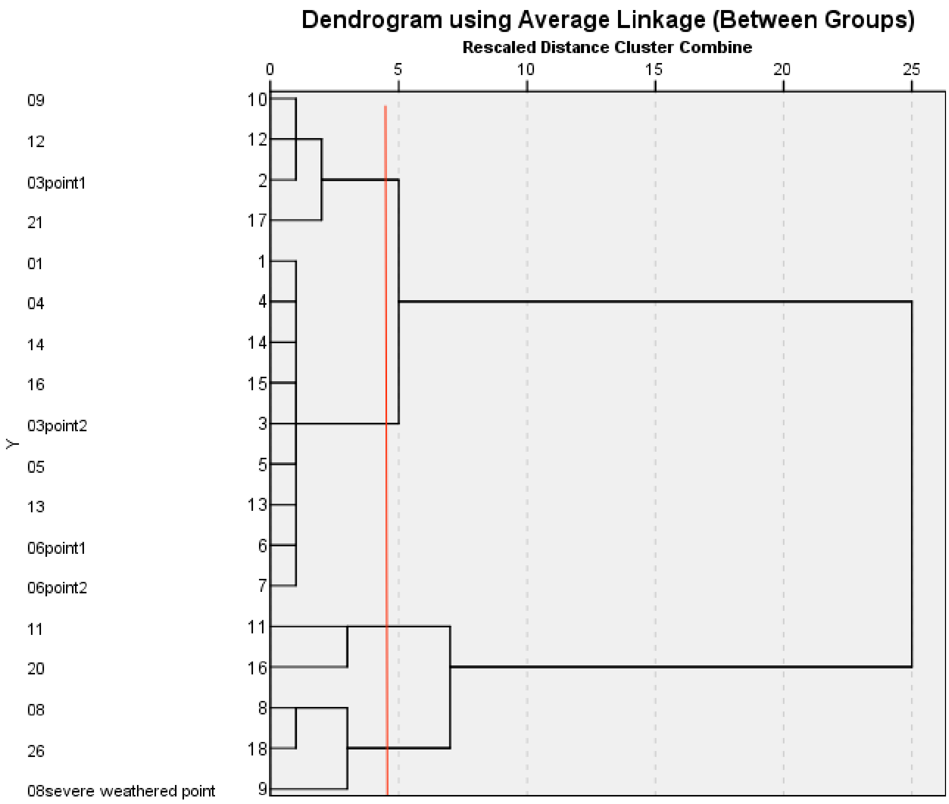

Figure 2. Genealogy of high potassium glass

We make scatter plots for several main chemical compositions of the two types of glass, preliminarily determine the number of clusters in combination with the content and relationship between different components, and make reasonable explanations for these data results (Figure 3-6).

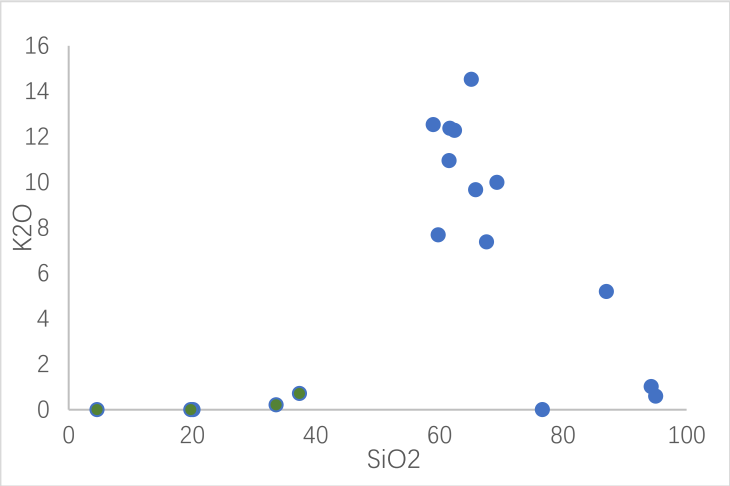

Figure 3. Scatter Chart (1)

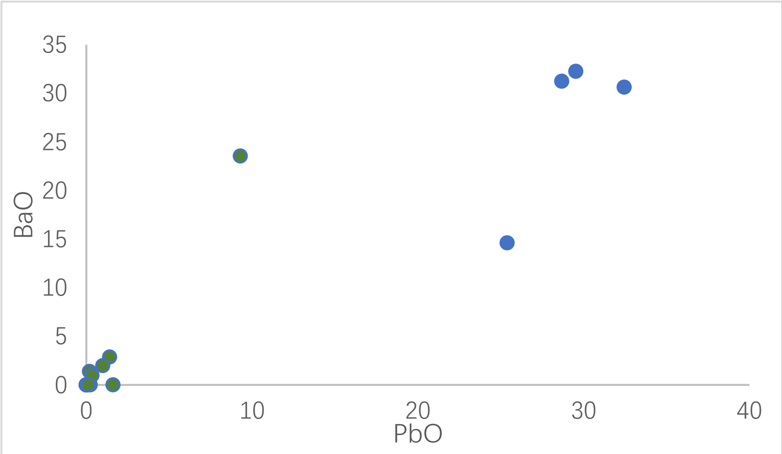

Figure 4. Scatter Chart(2)

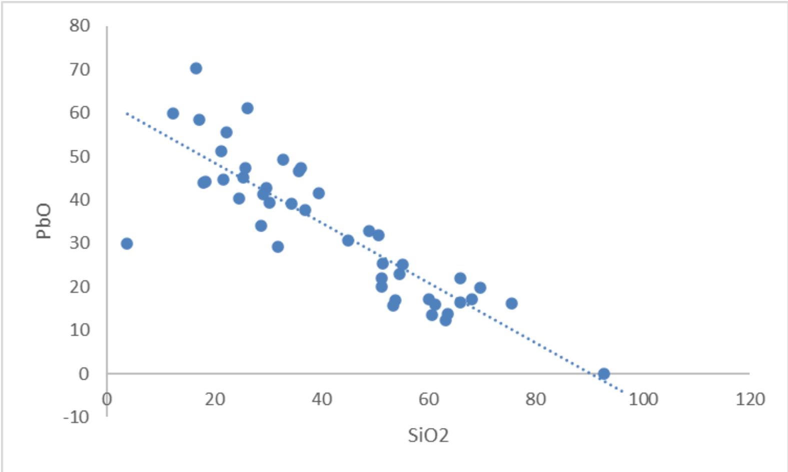

Figure 5. Scatter Chart (3)

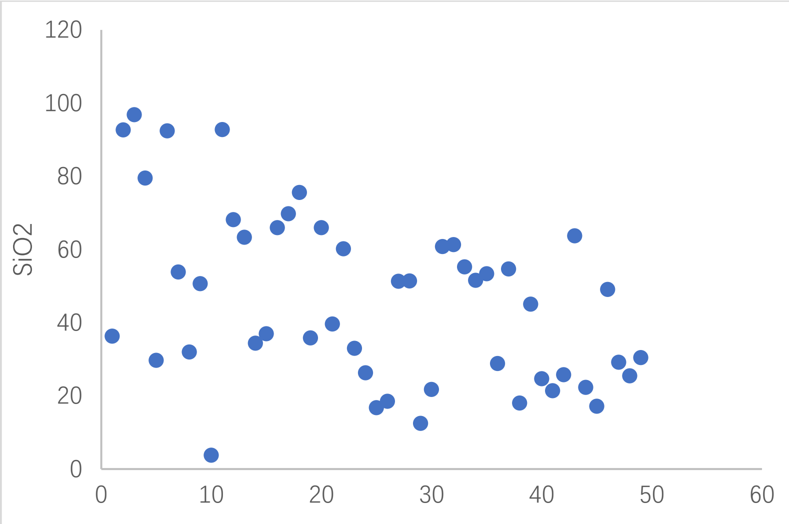

Figure 6. Scatter Chart (4)

We can see from the scatter diagram that silicon dioxide is special and will affect the acid-base property, so it should be taken as the main basis. At the same time, for high potassium glass, the potassium content is high. According to the content of silicon dioxide and potassium oxide, it can be divided into two categories (Figure 3), and the content of lead oxide and barium oxide is almost the same (Figure 4). Therefore, it is more reasonable to select these four types of chemical components and divide them into four categories. At the same time, for lead barium glass, there is almost a negative correlation between silicon dioxide and lead oxide (Figure 5). The content of silicon dioxide can be roughly divided into four layers (Figure 6), so we divide it into four categories.

We use the elbow rule to roughly estimate the optimal number of clusters through graphs, and make a broken line graph of aggregation coefficient based on tables (Figure 7, 8).The abscissa does not drop obviously at four places, which is consistent with the above analysis, so the number of subcategories is determined to be 4.

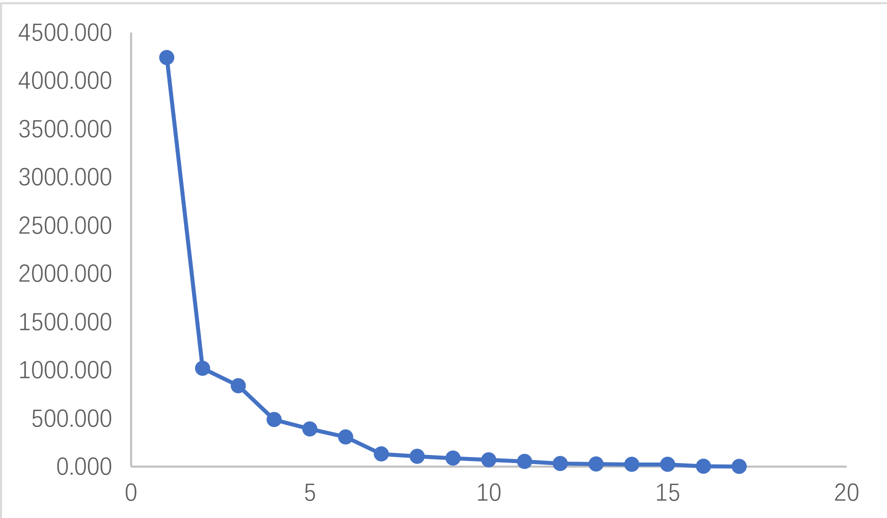

Figure 7. Polymerization coefficient broken line diagram of high potassium glass

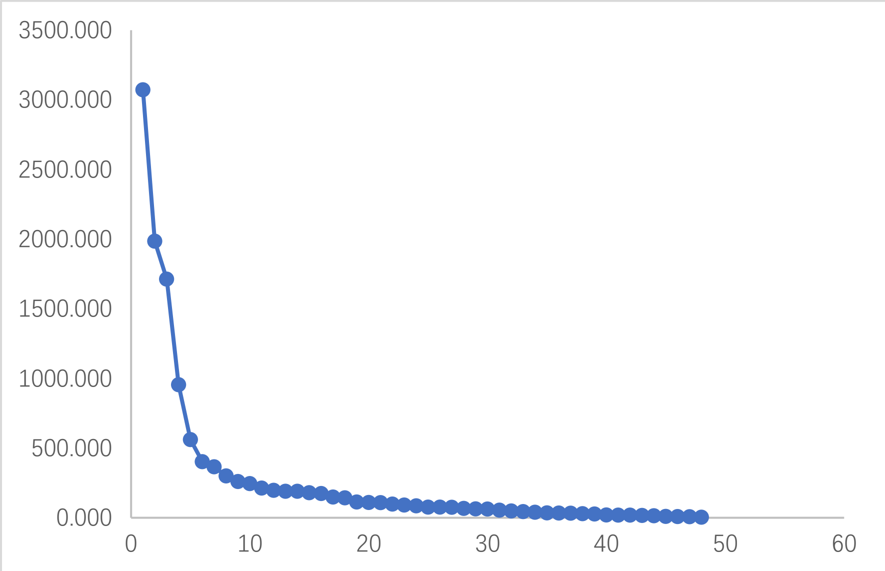

Figure 8. Polymerization coefficient broken line diagram of lead barium glass

4. Class discrimination model of glass relics

We use the model established in Section 3 to substitute the data into it, and the results are shown in Table 8:

Table 8. Prediction Results

Number	Silicon type	High K probability	Pb-Ba probability
A1	Pb-Ba	0.13993	0.86007
A2	Pb-Ba	0.058	0.942
A3	Pb-Ba	0.00419	0.99581
A4	Pb-Ba	0.31156	0.68844
A5	Pb-Ba	0.00052	0.99948
A6	Pb-Ba	0.30922	0.69078
A7	Pb-Ba	0.25505	0.74495
A8	high K	0.96196	0.03804

From the calculation results, we can analyze that seven groups of probability belong to Pa-Ba glass, and the other group belongs to high potassium glass.

5. Conclusion

(1) The mathematical model established in this paper can effectively identify glass components.
(2)It's reasonable that the two types of glass are each divided into four categories.
(3) Texture and color are not related to weathering.

References

[1] Xue Lu Protection and Restoration of Glass Cultural Relics [C]//. Proceedings of the Seventh Annual Academic Conference of China Association for Cultural Relics Protection Technology, 2012: 89 - 92.

[2] Gan Fuxi. The Road of Glass and Jade -- Also on Sino foreign Cultural and Technical Exchanges of Silicate Cultural Relics before the Pre Qin-Period [J]. Journal of Guangxi University for Nationalities (Natural Science Edition), 2009, 15 (04): 6 - 17.

[3] Liu Jingbao. Special Glass for Cultural Relics Protection [J]. Henan Science and Technology, 1990 (05): 17.

[4] Yu Qun, Huo Xiaodong, He Jian, Li Lin, Zhang Jianxin, Feng Yuyao. Prediction of the trend of power outage accidents in China's power grid based on Spearman correlation coefficient and system inertia [J]. Chinese Journal of Electrical Engineering: 1-12 [2022-10-01].

[5] Yun Tao, Zhang Jinqiannan, Li Shanshan, Gong Yue. Research on the analysis and evaluation method of expert review behavior based on Spearman coefficient and multi-layer perceptron [J]. Science and Technology Bulletin, 2022, 38 (05): 107 - 112.

[6] Yang Guochuang, Yin Tong, Wang Jinyue, Gao Yan, Sun Zeyu. Research on large area coordinate transformation method based on multiple regression grid correction model [J]. Journal of Navigation and Positioning: 1-8 [2022-10-01].

[7] Liu Yingcheng, Guo Rui, Liu Rongchuan, Qiao Honglei. Optimization model of calcareous soil consolidation characteristics based on multiple regression model [J]. Science and Technology Innovation, 2022 (16): 67 - 70.

[8] Wang Qian, Sun Daqiang. Prediction of health resource allocation in Tianjin based on multiple regression grey model [J]. China Chronic Disease Prevention and Control, 2022, 30 (08): 592 - 596.

[9] Wang Yanan, Xie Yanqi, Xie Liqin, Chen Wei. Decomposition analysis of China's urban carbon emissions based on LMDI model and Q-cluster [J]. Environmental Science Research, 2019, 32 (04): 539 - 546.

[10] Wang Weitao. Research on the application of Q-type clustering method in the CBA regular season 2016~2017 [J]. Liaoning Sports Science and Technology, 2017, 39 (05): 58 - 62.