Statistical Rule and Correlation Study of Chemical Composition of Glass Products Based on Machine Learning

Yaodong Zhang

19511426889@163.com

Shanxi University of Finance and Economics, Taiyuan, Shanxi Province, China

Abstract. In this paper, the relationship between surface weathering of glass products and glass-related elements and relationship between various chemical components were analyzed by data preprocessing and mathematical model. Firstly, a correlation analysis model was established, Pearson correlation coefficient was used. The significance test and correlation coefficient were used to conclude that the surface weathering was related to the type of glass and had a strong correlation, but had no correlation with tattoo and color. Secondly, establish the multiple linear regression model, and the statistical rule of the chemical composition content of the two kinds of glass before weathering was obtained by using the least square method. Finally, the chemical composition contents of two types of glass were taken as variables, and establish a correlation analysis model. Significance test was used to determine whether there was a correlation between the variables. The correlation coefficient was used to determine the strength of the relationship between the two variables, and the correlation relationship between the chemical composition of the same type of glass products was found.

Keywords: Surface Weathering, Glass Type, Significance Test, Correlation Analysis, Pearson Correlation Coefficient, Multiple Linear Regression

1. Introduction

1.1 Problem Background

Ancient glass is susceptible to the influence of the surrounding environment resulting in weathering phenomenon. Weathering phenomenon makes the elements inside the glass products and elements in the environment carry out a large number of exchanges, which has a great impact on the protection and identification of cultural relics.[1]Therefore, it is of great significance to analyze and identify the composition of ancient glass products.

1.2 Related Research

Yang Lei et al. used hierarchical clustering algorithm to cluster high-potassium and lead-barium glass, and further classified the sub-categories according to the changes of the corresponding chemical composition content of each category before and after weathering.[2]According to the content of various glass chemical components, Xu Siqi et al. established a decision tree model to study and analyze.[3] Huang Huiting et al. studied the weathering laws of ancient glass products and classified them according to the relevant detection data of ancient glass products

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345351

in China.[4] Zhang Qian et al. obtained the composition of glass products by analyzing the correlation of external factors of glass and conducting independent sample Mann-Whitney test on glass.[5] Chen Zhihao et al. used Spearman correlation coefficient and difference Chi-square test to analyze the correlation and difference between surface weathering and its influencing factors, and established hierarchical clustering models to classify different kinds of glass products into sub-categories.[6] Li Lingling used cluster analysis, log-linear analysis and other statistical methods to analyze and identify glass components.[7] Yin Yulong classified glass relics according to the glass classification model, the composition and identification of ancient glass products were studied by means of optimal scale method and Fisher multiple classification criteria.[8] Langlais C et al. used a theoretical and experimental approach in order to predict the thermal resistance of fibrous insulants, and to determine the influence of the chemical composition of the bulk glass, of the insulant morphology (porosity, specific surface, anisotropy) and of the temperature of use (T < 600°C).[9] Jantzen C M thought the relationship between glass viscosity and electrical resistivity was shown to relate to glass polymerization. [10]

To sum up, regarding the relevant research on the composition of ancient glass products, scholars have constructed classification models to classify different types of glass and make subcategories, and also studied the influence of some external factors on the composition of glass. However, there were few researches on the relationship between surface weathering and factors and interaction between components.

2. Data Source

The data came from the detection of 58 kinds of glass cultural relics, and the results of color, tattoo, type, and surface weathering of glass cultural relics were statistically summarized. The chemical composition content of different sampling points of cultural relics was detected and statistically summarized, invalid data and wrong data were eliminated.

3. Model Establishment

3.1.Correlation Analysis

The sample data is obtained by conducting n random experiments on random variables. First, whether there is a statistically significant relationship between X and Y is tested. The significance coefficient P reflects the probability of an event. Generally, $P<0.05$ is considered to be statistically significant. Pearson correlation coefficient is calculated as follows:

$$\rho_{X,Y}=\frac{Cov\ (X,Y)}{\sigma_X \cdot \sigma_Y}=\frac{\sum_{i=1}^{n}(x_i-E(X))(y_i-E(Y_a))}{\sqrt{\sum_{i=1}^{n}(x_i-E(X))^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-E(Y_a))^2}} \tag{1}$$

Where E(X) and E(Y) represent their respective averages, a=1, 2, and 3.n stands for number of trials.σ_X and σ_Y are the standard deviations of X and Y, respectively. Cov（X,Y）Represents the total body covariance. Pearson correlation coefficient ranges from [-1,+1], where negative numbers represent negative correlation, positive numbers represent positive correlation, and 0 represents no correlation. The closer the correlation coefficient is to 0, the weaker the correlation is. The closer it is to -1 or +1, the stronger the correlation is.

3.2.Multiple Linear Regression Analysis Model

Now we get n independent observations, and we get $(y_i, x_{i1}, \ldots, x_{im}), i=1,2,\cdots,n, n>m$

$$\begin{cases} y_i=\beta_0+\beta_1 x_{i1}+\cdots+\beta_m x_{im}+\varepsilon_i \\ \varepsilon\sim N(0,\sigma^2), i=1,\cdots,n \end{cases} \tag{2}$$

$$X=\begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, Y=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$\varepsilon=(\varepsilon_0,\varepsilon_1,\ldots,\beta_m)^T, \beta=(\beta_0,\beta_1,\cdots,\beta_m)^T$, therefore

$$\begin{cases} Y=X\beta+\varepsilon \\ \varepsilon\sim N(0,\sigma^2 E_n) \end{cases} \tag{3}$$

E_n is the identity matrix of order n. The estimates $\hat{\beta}_j$ are selected $\beta_j=\hat{\beta}_j$, $j=0,1,2\cdots m$ so that the sum of squares of error $Q=\sum_{i=1}^n \varepsilon_i^2=\sum_{i=1}^n (y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_m x_{im})^2$ is minimized at the time. Ream $\frac{\partial Q}{\partial \beta_j}=0$, $j=0$, 1, $2\cdots n$,

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0}=-2\sum_{i=1}^n (y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_m x_{im})=0 \\ \dfrac{\partial Q}{\partial \beta_0}=-2\sum_{i=1}^n (y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_m x_{im})x_{ij}=0 \end{cases}, j=1,2,\cdots,m \tag{4}$$

It is organized into the following,

$$\begin{cases} \beta_0+\beta_1\sum_{i=1}^n x_{i1}+\beta_2\sum_{i=1}^n x_{i2}+\cdots+\beta_m\sum_{i=1}^n x_{im}=\sum_{i=1}^n y_i \\ \beta_0\sum_{i=1}^n x_{i1}+\beta_1\sum_{i=1}^n x_{i1}^2+\beta_2\sum_{i=1}^n x_{i1}x_{i2}+\cdots+\beta_m\sum_{i=1}^n x_{i1}x_{im}=\sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \beta_0\sum_{i=1}^n x_{im}+\beta_1\sum_{i=1}^n x_{im}x_{i1}+\beta_2\sum_{i=1}^n x_{im}x_{i2}+\cdots+\beta_m\sum_{i=1}^n x_{im}^2=\sum_{i=1}^n x_{im}y_i \end{cases} \tag{5}$$

The matrix form of the normal system of equations is $X^T X\beta=X^T Y$, when the X column is full rank, $X^T X$ is an invertible square matrix, so

$$\hat{\beta}=(X^T X)^{-1}X^T Y\# \tag{6}$$

Substitute $\hat{\beta}$ back to the original model to get an estimate of y

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_m x_m \tag{7}$$

The fitting value is $\hat{Y}=X\beta$, $e=Y-\hat{Y}$ and the residual sum of squares is $Q=\sum_{i=1}^n e_i^2=\sum_{i=1}^n (y_i-\hat{y}_i)^2$. Finally, the hypothesis test of the regression model is carried out, and the original hypothesis is $H_0:\beta_j=0$ $(j=1,\cdots,m)$.

$$F=\frac{U/m}{Q/(n\text{-}m\text{-}1)} \sim F(m,n\text{-}m\text{-}1) \quad (8)$$

Where U is the regression sum of squares. There is an upper α quantile $F_a\ (m,n\text{-}m\text{-}1)$ at the significance level, if $F<F_a(m,n\text{-}m\text{-}1)$, the null hypothesis is accepted; Otherwise refuse.

4. Empirical Analysis

4.1.The Relationship Among Surface Weathering of Glass Relics and Glass Type, Tattoo and Color

Table 1. Table of phase relations

	Tattoo	Type	Colour	Surface Weathering
Tattoo	(0.000 * 1.000 * *)	0.423 (0.001 * * *)	0.363 (0.007 * * *)	0.176 (0.202)
Type	0.423 (0.001 * * *)	(0.000 * 1.000 * *)	(0.001 * 0.424 * *)	0.316 (0.020 * *)
Colour	0.363 (0.007 * * *)	(0.001 * 0.424 * *)	(0.000 * 1.000 * *)	0.115 (0.406)
Surface Weathering	0.176 (0.202)	0.316 (0.020 * *)	0.115 (0.406)	(0.000 * 1.000 * *)

Table 1 shows the parameter results of the model. It can be seen from the significance level that the surface weathering glass products is statistically different from type, and there is a correlation between the two, and there is no correlation between the surface weathering of glass products and the pattern and color. It can be seen that the surface weathering of glass products has the strongest correlation with the glass type.

4.2.Statistical Rule of Composition Content of Cultural Relics before Weathering

Table 2. Results of linear regression analysis

Linear regression analysis results n=45									
	Nonnormalized coefficient		Standardization coefficient	t	p	VIF	R²	Adjust R²	F
	B	Standard error	Beta						
Constant	50.928	13.04	-	3.906	0.000 * * *	-	0.577	0.468	F=5.305 P=0.000***
Tattoo	22.55	6.11	0.589	3.691	0.001 * * *	2.109			
Col _2	7.3	7.402	0.132	0.986	0.331	1.49			
Col _3	3.407	8.684	0.045	0.392	0.697	1.105			
Col _4	13.375	7.402	0.243	1.807	0.079 *	1.49			
Col _5	13.252	9.84	0.176	1.347	0.187	1.418			
Col _6	11.146	8.249	0.169	1.351	0.185	1.297			
Col _7	31.479	11.217	0.346	2.806	0.008 * * *	1.258			
Col _8	4.446	15.543	0.035	0.286	0.777	1.236			
Surface Weathering	19.368	5.665	0.468	3.419	0.002 * * *	1.552			
Dependent Variable: Silicon dioxide $(SiO)_2$									

Table 2 shows the analysis results of this model. As can be seen from the results, P value is 0.000***, the significance at the level, rejecting the null hypothesis that the regression coefficient is 0, the model meets the requirements for the collinearity of variables, VIF is all less than 10, and the model has no multicollinearity problem. Formula is as follows: y = 50.93 + (22.55) + (7.30) * * grain col _2 + 3.41 * col _3 + 13.38 * col _4 + 13.25 * col _5 + (11.15) * col _6 + (31.48) * col _7 + _8 4.45 * col_8+19.37* Surface weathering. Lead barium glass weathering front is obtained by model in the chemical composition content of the statistical rule is: y = 38.01 + 2.44 * grain _2. 0 + (6.69) * grain _3. 0 + (7.24) * col _2 + (6.86) * col _3 + 7.69 * col _4 + 31.32 * surface weathering. Where y refers to the content of Silicon dioxide.

4.3.Correlation Between Chemical Composition of Different Glass Relics Samples

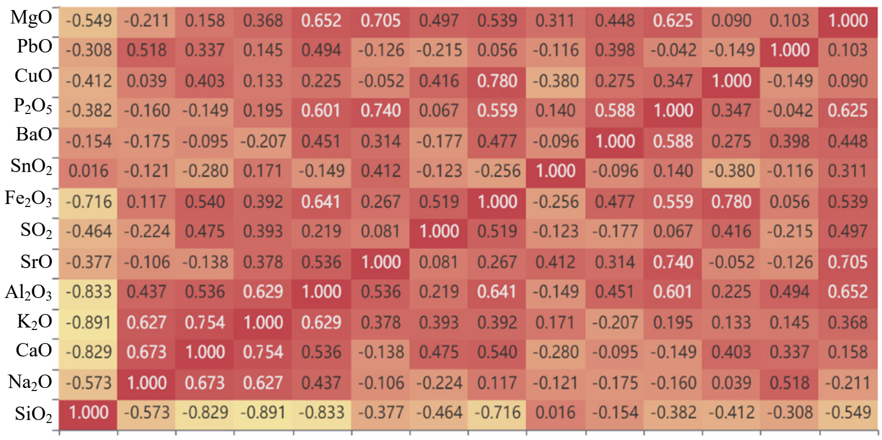

Figure 1. Heat map of correlation coefficient of high potassium glass

Figure 1 shows the value of phase relationship in the form of thermal map. The value of correlation is represented by color depth. The darker the color, the greater the correlation. According to the thermal map, the correlation between its chemical composition can be obtained: silicon dioxide in high-potassium glass, and most of the oxides, such as aluminium oxide, potassium oxide, calcium oxide, iron oxide, sodium oxide are correlated with each other, and the correlation is strong. Other oxides such as potassium oxide with aluminium oxide, potassium oxide with calcium oxide, aluminium oxide, calcium oxide with aluminium oxide, iron oxide, magnesium oxide with aluminium oxide, iron oxide, phosphorus pentoxide, strontium oxide with phosphorus pentoxide, alumina, phosphorus pentoxide with alumina, iron oxide, barium oxide, copper oxide with iron oxide, sulfur dioxide and iron oxide, there is strong correlation with each other.

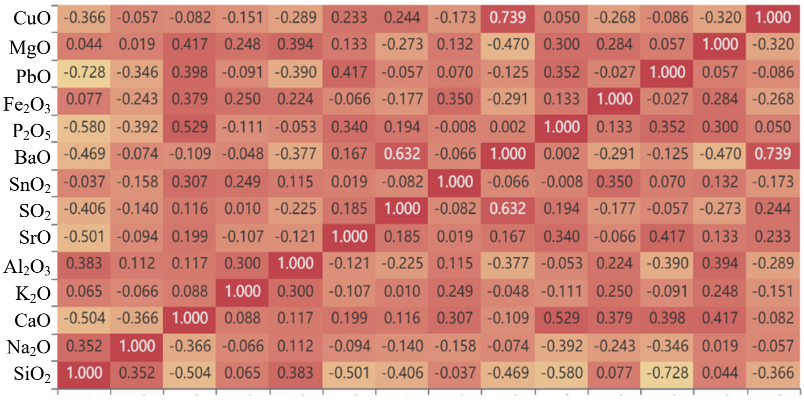

Figure 2. Heat map of correlation coefficient of lead barium glass

Figure 2 shows the value of the phase relationship in the form of a thermal map. The size of the value is represented by the color depth. According to the thermal map, the correlation between its chemical composition can be obtained: silicon dioxide in lead barium glass, and a small amount of oxides such as lead oxide, phosphorus pentoxide, calcium oxide are correlated, and the correlation is strong. Other oxides in the category of lead barium glass, there are fewer oxides such as calcium oxide and phosphorus pentoxide, sulfur dioxide and barium oxide, barium oxide and copper oxide have a strong correlation.

5. Conclusions

1.The surface weathering of glass products is correlated with the type, while they are not correlated with the pattern and color, the surface weathering of glass products is most correlated with the type.

2.The statistical law of the chemical composition content of high-potassium glass before weathering is: Y = 50.93 + (22.55) + (7.30) * * grain col _2 + 3.41 * col _3 + 13.38 * col _4 + 13.25 * col _5 + (11.15) * col _6 + (31.48) * col _7 + 4.45 * col _8 + 19.37* Surface weathering. Lead barium glass weathering before the statistical regularity of chemical composition content is: y = 38.01 + 2.44 * grain _2. 0 + (6.69) * grain _3. 0 + (7.24) * col _2 + (6.86) * col _3 + 7.69 * col _4 + 31.32 * surface weathering. Where y refers to the content of Silicon dioxide.

3.There is a strong correlation between silica and most oxides in high potassium glass. Other oxides have a strong correlation. There is a strong correlation between silica and a small amount of oxides in lead-barium glass. There are few other oxides have a strong correlation.

References

[1] Wang Jie, LI Mo, MA Qinglin, ZHANG Zhiguo, ZHANG Meifang, WANG Julin.Weathering of an octoprismatic lead-barium glassware in the Warring States Period [J]. Glass and Enamel,2014,42(02):6-13.

[2] Yang Lei, Wei Tong, ZONG Yiheng.Mathematical model for composition analysis and

identification of ancient glass products [J]. Journal of Beijing Institute of Printing and Technology, 2019,31(09):
[3] Xu Siqi, Xie Jian, Lin Han et al.Analysis and identification of weathered glass products based on decision tree model [J]. Computer Knowledge and Technology, 2019,19(25):
[4] Huang Huiting, Li Chunming, Liu Siyu et al.Composition analysis and identification of ancient glass products based on composition data [J]. Mathematical Modeling and its Application, 2019,12(02):
[5] Zhang Qian, LV Kunshan, ZHAO Xianqing et al.Composition analysis of ancient glass products based on random forest classification model [J]. Henan Science and Technology,2023,42(09):
[6] Chen Zhihao, Ji Jingmin.Composition analysis and identification of ancient glass products based on hierarchical clustering model [J]. Modern Information Technology,2023,7(08):
[7] Li Lingling. Analysis and identification of components of ancient glass products based on statistical analysis [J].China New Technology and New Products,2023,(02):
[8] Yin Yulong. Composition analysis of ancient glass products by correlation prediction [J]. Contemporary Chemical Industry Research,2023,(01):
[9] Langlais C, Guilbert G, Banner D, et al. Influence of the chemical composition of glass on heat transfer through glass fibre insulations in relation to their morphology and temperature of use[J]. Journal of Thermal Insulation and Building Envelopes, 1995, 18(4): 350-376.
[10] Jantzen C M. First principles process-product models for vitrification of nuclear waste: Relationship of glass composition to glass viscosity, resistivity, liquidus temperature, and durability[R]. Westinghouse Savannah River Co., Aiken, SC (United States), 1991.