Analysis of chemical composition of ancient glass products based on multiple linear regression

Jiankun Fang *, Zhaoyang Li, Shihao Cao

Department of Computer and Information Technology, Beijing Jiaotong University, Beijing, China

* Corresponding author: 20722004@bjtu.edu.cn

Abstract. During the Silk Road period, glass was introduced to China, and China began to produce ancient glass with Chinese characteristics by absorbing and developing its technology. However, ancient glass products is very vulnerable to weathering in the buried environment, and many components of glass change accordingly during the weathering process. Analyzing the changes of various factors and components of ancient glass before and after weathering is of great significance to the exploration and study of ancient glass. Therefore, this paper obtains the relationship between weathering on the surface of glass artifacts and glass type, decoration and color by establishing a multiple linear regression model, and obtains the statistical law of chemical composition content with and without weathering on the surface of glass samples through data analysis: weathering on the surface of glass artifacts has a strong correlation with glass type and decoration, but a weak correlation with glass color. The predicted chemical composition content before weathering was carried out by the weighted average method, and the prediction table as well as the scatter plot were obtained. From the data, it can be seen that most of the total chemical composition of each predicted sample fluctuated above and below 100%, and the average value of all samples was obtained as 98.05%, which meets the requirement of valid data.

Keywords: Ancient glass products, Multiple linear regression model, Weighted average method.

1. Introduction

Glass was introduced to China from the West early on as one of the trade commodities along the Silk Road. By studying and developing techniques for making glass, China began to produce ancient glass with Chinese characteristics [1]. However, ancient glass is very susceptible to weathering if it is buried in some specific environment. Their internal elements would exchange heavily for elements contained in the buried environment, thus changing many components of the glass. Therefore, it becomes particularly important to analyze the changes in glass characteristics and composition before and after weathering through data, and the conclusions are of great significance for the prediction and study of weathered glass composition.

2. Multiple linear regression model building and solving

There are 67 valid samples in this paper, including 16 high potassium glasses and 51 lead-barium glasses. For these samples, there are 15 indicators, including the presence or absence of weathering, ornamentation, color and 14 chemical composition contents. However, some samples had some missing indicators. Some data are shown in Table 1.

Table 1. Some data of ancient glass samples

Number	1	2	3	4	5
Type	High Potassium	Lead Barium	High Potassium	High Potassium	High Potassium
Whether weathering	No	Yes	No	No	No
Ornamentation	C	A	A	A	A
Color	Blue Green	Light Blue	Blue Green	Blue Green	Blue Green
SiO_2	69.33	36.28	87.05	61.71	65.88
Na_2O	/	/	/	/	/
K_2O	9.99	1.05	5.19	12.37	9.67
CaO	6.32	2.34	2.01	5.87	7.12
MgO	0.87	1.18	/	1.11	1.56
Al_2O_3	3.93	5.73	4.06	5.5	6.44
Fe_2O_3	1.74	1.86	/	2.16	2.06
CuO	3.87	0.26	0.78	5.09	2.18
PbO	/	47.43	0.25	1.41	/
BaO	/	/	/	2.86	/
P_2O_5	1.17	3.57	0.66	0.7	0.79
SrO	/	0.19	/	0.1	/
SnO_2	/	/	/	/	/
SO_2	0.39	/	/	/	0.36

For some of the indicators in the index, this paper first transforms each definite class of variables, as shown in Table.2.

Table 2. Variable description of some indicators

Variables	Variable Description							
Weathering or not	No				Yes			
	1				0			
Category	High Potassium				Lead Barium			
	1				0			
Ornamentation	A		B				C	
	1		2				3	
Color	Light Green	Light Blue	Dark Green	Deep Blue	Purple	Green	Blue Green	Black
	0	1	2	3	4	5	6	7

2.1. Multiple linear regression model building

Table 3. Explanation of ancient glass variables

Variable Type	Variables	Variable Name
Variable Indicators	Dependent variable	With or without weathering
	Independent variable	Category
		Ornamentation
		Color

The data provided can be divided into variables as shown in Table.3. In order to test whether there is weathering of cultural relics, this paper constructs the model of Function (1):

$$y_i = \alpha + \Sigma\beta_n \times Emblazonry_n + \Sigma\lambda_n \times Type_m + \Sigma\delta_n \times Color_j + \varepsilon_i \quad (1)$$

Function (1) in y_i indicates whether the ith artifact in the sample is weathered, $Emblazonry_n$ is the independent variable of ornamentation, this paper sets ornamentation A as the control group, the remaining 2 ornaments as independent variables, so n = 1, 2, if the ith sample of artifacts from the kth ornamentation, $Emblazonry_k = 1$, the rest of $Emblazonry_n$ ($n \neq k$) all take 0; if the ith If the ith sample is ornament A, all $Emblazonry_n$ take 0. $Type_m$ (m=1) and $Color_j$ (j=7) have approximately the same meaning as above, Type is the independent variable of type, setting lead-barium type as the control group, and Color is the independent variable of color, setting dark blue color as the control group. Where ε is an unobservable and completely random perturbation term.

The first joint significance test of the regression coefficients before the three independent variables yields $p = 0.000 < 0.01$, so the original hypothesis is rejected at 99% confidence interval and β_i is considered not all 0 [2]. The significance t-test for the distribution of the obtained regression coefficients β_i shows that the regression coefficients are significant at the 99% confidence level, the regression coefficients of the tattoos are significant at the 90% confidence level, while none of the colors are significant. In order to make the results more accurate, the heteroskedasticity White's test and the multicollinearity VIF test are proposed to be performed on the data [3-4].

Heteroskedasticity test: First, the residuals are plotted against the fitted Figure 1 Intuitively, it is difficult for us to see the presence of heteroskedasticity.

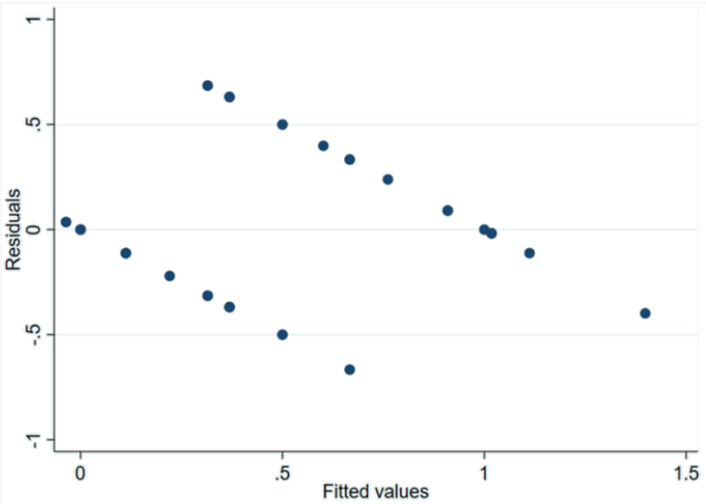

Figure 1. Scatter plot of OLS model residuals versus fitted values

For more accurate identification, White's test was performed on the OLS model and the result $\chi^2(15) = 33.23$, $p = 0.0044 < 0.01$, so the original hypothesis was rejected at 99% confidence level and the heteroskedasticity was considered to exist.

The two main methods to solve the heteroskedasticity problem are OLS+ robust standard error and generalized least squares GLS. Stock and Watson recommend the use of the "OLS + robust standard error" method in most cases, so this method is used to solve the effect of heteroskedasticity.

The regression results using OLS + robust standard error methods are shown in Table.4 below:

Table 4. Results of OLS-based regression analysis

	Weathering		
Tattoo: A	0.000	Color: dark green	-0.030
	(.)		(-0.079)
Tattoo: B	-1.018***	Color: dark blue	0.000
	(-19.795)		(.)
Tattoo: C	-0.257**	Color: Purple	0.155
	(-2.050)		(0.355)
Type: lead barium	0.000	Color: Green	0.655*
	(.)		(1.949)
Type: High Potassium	0.797***	Color: blue-green	-0.381
	(10.145)		(-1.182)
Color: light green	0.322	Color: Black	-0.601*
	(0.710)		(-1.889)
Color: light blue	-0.232		
	(-0.674)		
Constant	0.601*	R-squared	0.5530
	(1.889)		

Robust standard errors in parentheses ***p<0.01, **p<0.05, *p<0.1

Firstly, the joint significance test for the regression coefficients before the three independent variables was obtained p = 0.000 < 0.01, so the original hypothesis was rejected at 99% confidence interval that β_i was not all 0. Then a significance t-test was conducted for the distribution of the obtained regression coefficients β_i. The results in Table.4 show that the regression coefficient of type is significant at the 99% confidence level, the regression coefficient of ornamentation is significant at the 95% confidence level, and the regression coefficient of a part of color is significant at the 90% confidence level.

2.2. VIF test for multicollinearity

To test for the presence of multicollinearity this paper uses the variance inflation factor VIF. The Function (2) is shown.

$$\mathrm{VIF}_i = \frac{1}{1-R^2_{1-k/i}} \tag{2}$$

In Function (2), $R^2_{1-k/i}$ is the goodness of fit obtained by regressing the i-th independent variable as the dependent variable on the remaining k- independent variables [5-6]. The regression model VIM=max (VIF_1, VIF_2, VIF_3, VIF_4, VIF_5) is defined, and if VIM>10, it is considered that there is multicollinearity.

Bringing the data into Function (2) yields the test results as shown in Table.5 below:

Table 5. VIF test for multicollinearity

Variable	VIF	1/VIF
COLOR2	7.65	0.130779
COLOR7	6.82	0.146708
COLOR3	5.02	0.199025
COLOR5	3.55	0.282047
COLOR1	2.95	0.339165
COLOR8	2.00	0.499753
EMBLAZONRY3	1.97	0.507919
TYPE2	1.87	0.534524
EMBLAZONRY2	1.69	0.592837
COLOR6	1.68	0.596776
Mean VIF	3.52	

From Table.5, we can obtain VIF = 3.52 < 10, so it is considered that there is no multicollinearity.

2.3. Analysis of relationships using backward stepwise regression

The final regression results were standardized using backward stepwise regression to obtain Table.6:

Table 6. Standardized backward regression results

Weathering	Coef.	Std. Err.	t	P>t	Beta
EMBLAZONRY2	-1.018029	.0514293	-19.79	0.000	-.6438584
EMBLAZONRY3	-.2567611	.1252562	-2.05	0.047	-.2581836
TYPE2	.7973274	.0785933	10.14	0.000	.7564112
COLOR6	.6554248	.3362131	1.95	0.058	.1778257
COLOR8	-.6013363	.3183382	-1.89	0.066	-.228543
_cons	.6013363	.3183382	1.89	0.066	

The linear regression model Function (3) was obtained from Table.6:

$$y = -1.018 \times Emblazonry_1 - 0.2567611 \times Emblazonry_2 + 0.7973274 \times Type_1 \\ +0.6554248 \times Color_5 - 0.6013363 \times Color_7 - 0.601 \tag{3}$$

Function (3) where $Emblazonry_1 = 1$ when and only when the artifact is decorated with B, $Emblazonry_2 = 1$ when the decoration is C, $Type_1 = 1$ when the category is high potassium, $Color_5 = 1$ when the color is green, $Color_7 = 1$ when the color is black, and 0 in all other cases. when the calculation result y is greater than 0.5, the glass artifact is not weathered, and less than 0.5 when the glass artifacts weathering. From Function (3) can be known the relationship between the presence or absence of weathering of glass artifacts and the type, decoration and color. For example, lead-barium glass is more susceptible to weathering than high-potassium glass, and glass is less susceptible to weathering when the color is green.

3. Analysis and prediction of the statistical pattern of chemical composition content of ancient glass products

3.1. Overall analysis of the chemical composition content of glass products

Firstly, for the statistical analysis of the content of 14 chemical components before and after weathering of high potassium and lead-barium glass, some data are shown in Table.7 and Table.8.

Table 7. Analysis of the chemical composition of high potassium glass before and after weathering

Variable Name	Weathering condition	Sample size	Average value	Standard deviation	Minimum value	Maximum value
SiO_2	No weathering	12	67.984	8.755	59.010	87.050
	Weathering	6	93.963	1.734	92.350	96.770
Na_2O	No weathering	12	0.695	1.287	0.000	3.380
	Weathering	6	0.000	0.000	0.000	0.000
K_2O	No weathering	12	9.331	3.920	0.000	14.520
	Weathering	6	0.543	0.445	0.000	1.010
Total content	No weathering	12	98.178	1.132	96.060	100.000
	Weathering	6	99.610	0.417	98.810	100.000

Table 8. Analysis of the chemical composition of lead barium glass before and after weathering

Variable Name	Weathering condition	Sample size	Average value	Standard deviation	Minimum value	Maximum value
SiO_2	No weathering	23	54.660	11.829	31.940	75.510
	Weathering	26	24.913	10.605	3.720	53.330
Na_2O	No weathering	23	1.683	2.372	0.000	7.920
	Weathering	26	0.216	0.557	0.000	2.220
K_2O	No weathering	23	0.219	0.310	0.000	1.410
	Weathering	26	0.133	0.240	0.000	1.050
Total content	No weathering	23	97.756	2.304	88.410	99.980
	Weathering	26	96.690	2.890	90.170	99.890

According to Figure 2, the following conclusions were drawn: for high potassium glass, color does not affect the determination of the presence or absence of weathering chemical content on the surface of the artifact sample; if the surface decoration of high potassium glass is A or C, the sample is free of weathering chemical content; if the surface decoration of high potassium glass is B, the sample contains weathering chemical content.

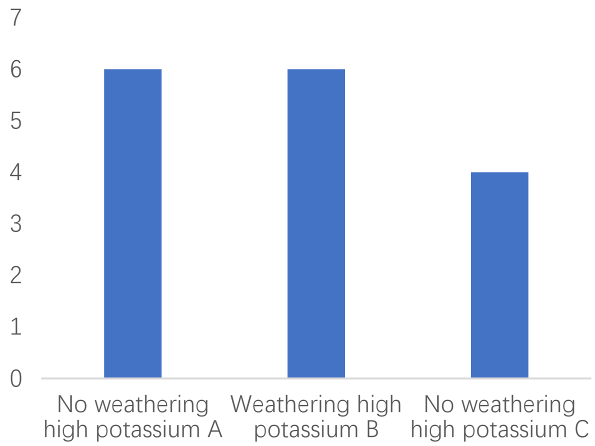

Figure 2. Weathering of high potassium glass with different ornamentation

As shown in Figure 3, after counting all the light blue or blue-green colored lead barium glass with motif C, it was found that all the samples in 8 groups were weathered glass. Therefore, it can be concluded that for the light blue or blue-green colored lead barium glass with motif C, it contains weathering chemical composition

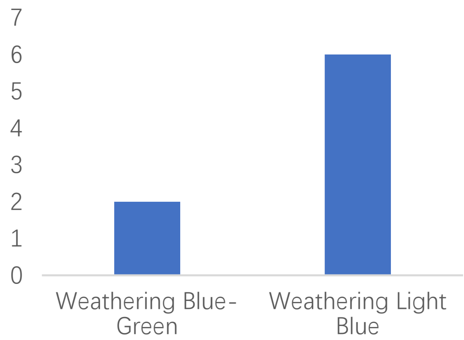

Figure 3. C grain decorated lead barium glass is blue-green and light blue weathering

According to Figure 4, For the light or dark green color lead barium glass with ornamentation C, if the silica content is greater than 40%, lead oxide content is less than 40%, and barium oxide content is less than 10%, the sample may not contain weathering chemicals; if the silica content is less than 40%, lead oxide content is greater than 40%, and barium oxide content is maintained at about 10%, the glass product may contain weathering chemicals.

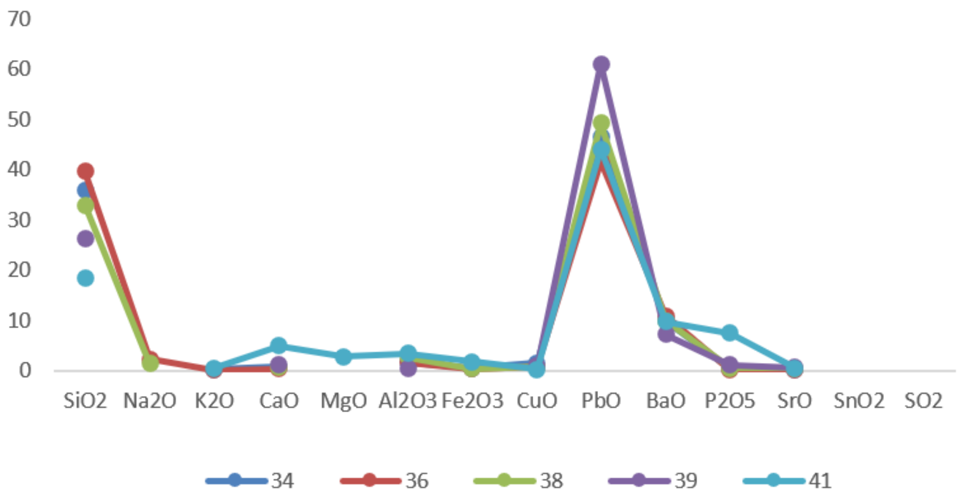

Figure 4. The chemical composition content of the ornamented C weathered lead barium glass is light green or dark green

According to Figure 5 and Figure 6, if the content of silica is more than 50% and the content of lead oxide is less than 30%, the glass product may not contain weathering chemical composition; if the content of silica is less than 40% and the content of lead oxide is more than 30%, the glass product may contain weathering chemical composition.

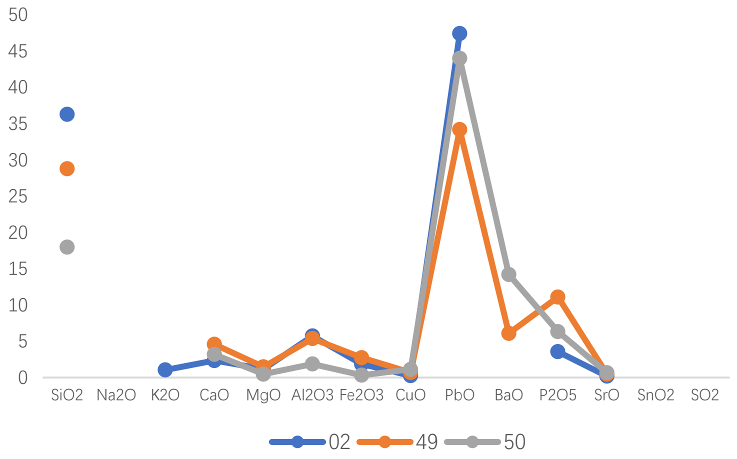

Figure 5. Chemical composition content of grain A weathered lead barium glass

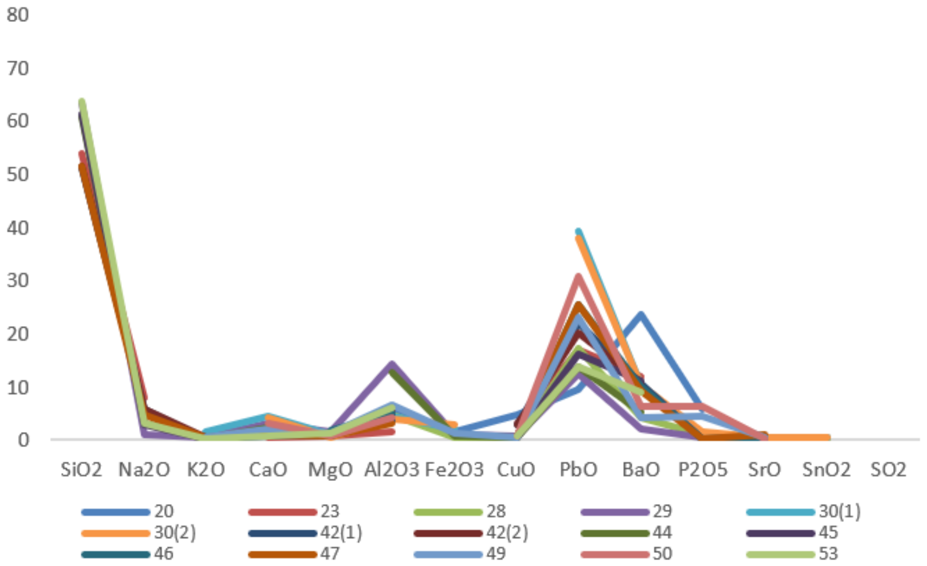

Figure 6. Texture A no weathering point chemical composition content of lead barium glass

3.2. Prediction of the chemical composition content of ancient glass before weathering based on the weighted mean solution

From the data analysis, it can be obtained that because there are more "0" values in the data, in order to ensure the accuracy of the prediction results and weaken the influence of 0 values on the prediction, this paper intends to use the weighted average method for prediction, using the standard normal distribution function for the calculation of weights [7-9]. The calculation process is as follows: Fourteen chemical contents (silicon dioxide (SiO_2), sodium oxide (Na_2O), potassium oxide (K_2O), calcium oxide (CaO), magnesium oxide (MgO), aluminum oxide (Al_2O_3), iron oxide (Fe_2O_3), copper oxide (CuO), lead oxide (PbO), barium oxide (BaO), phosphorus pentoxide (P2O5), strontium oxide (SrO), sulfur dioxide (SO_2), tin oxide (SnO_2)) before weathering are $x_1,x_2,\cdots,x_{14}$, respectively, and their weights are $w_1,w_2,\cdots,w_{14}$, respectively, and after weathering are x_i' and their weights are w_i', and the relationship can be obtained by the standard normal distribution function with the following Function (4) [10]:

$$w_i = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} \tag{4}$$

Its weighted average Function (5) is

$$\begin{cases} \overline{x} = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_{14} w_{14}}{w_1 + w_2 + \cdots + w_{14}} \\ \overline{x}' = \frac{x_1' w_1' + x_2' w_2' + \cdots + x_{14}' w_{14}'}{w_1' + w_2' + \cdots + w_{14}'} \end{cases} \tag{5}$$

Multiplying the 14 chemical contents after weathering with the ratio of $\bar{x}$ to $\bar{x}'$, the result is the corresponding chemical content before weathering, as shown in Function (6):

$$x_i = x_i' \frac{\overline{x_i}}{\overline{x_i}'} \tag{6}$$

The predicted contents of some chemical components of each sample before weathering are shown in Table.9 below:

Table 9. Predicted chemical content of the pre-weathering fraction of each sample

Number	Type	x_1	……	x_9	x_{10}	……	x_{14}
51	Lead Barium	46.91	……	3.19	1.77	……	0.17
52	Lead Barium	52.54		6.17	1.20		0.00
54	Lead Barium	42.20		2.51	1.93		0.00
56	Lead Barium	59.50		11.04	0.53		0.00
57	Lead Barium	51.89		12.36	0.00		0.00
58	Lead Barium	62.03		5.47	1.89		0.00
07	High Potassium	69.72	……	0.00	0.00	……	0.00
09	High Potassium	71.52		0.00	0.00		0.00
10	High Potassium	72.83		0.00	0.00		0.00
12	High Potassium	70.97		0.00	0.00		0.00
22	High Potassium	69.51		0.00	0.00		0.00
27	High Potassium	69.78		0.00	0.00		0.00

The results are plotted as shown in Figure 7:

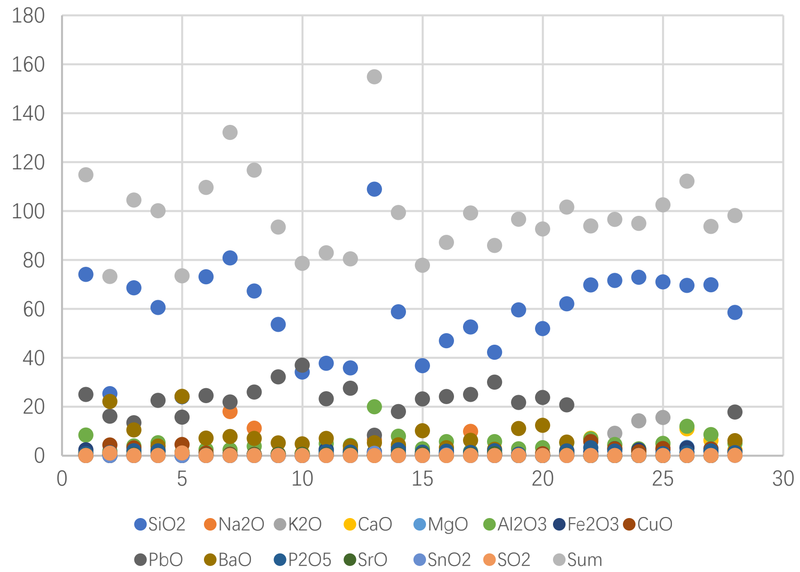

Figure 7. Predicted chemical composition content of each sample before weathering

From Figure 7, it is known that most of the total chemical composition of each sample fluctuates above and below 100%, and the mean value of 98.05% was obtained for all samples, which meets the requirement of valid data.

4. Conclusion

In this paper, in order to investigate the statistical patterns of chemical composition content of ancient glass products with and without weathering. Multiple linear regression models, heteroskedasticity White's test, multiple covariance test and weighted mean method were used to analyze 67 samples and predict the chemical composition content before weathering for 27 samples. Among them, a strong correlation between glass weathering and glass type and grain decoration was obtained by models such as multiple linear regression, but a weak correlation with glass color. The pre-weathering chemical composition of 27 samples was predicted by the weighted average method, and the sum of chemical composition of each predicted sample was obtained to be mostly above and below 100%, and the mean value of 98.05% was obtained for all samples, which meets the requirement of valid data.

References

[1] Sichin Bilig, Li Qinghui, Gan Fuxi. Analysis of ancient Chinese potassium glass fractions by laser exfoliation-inductively coupled plasma-atomic emission spectrometry/mass spectrometry [J]. Analytical Chemistry, 2013, 41 (09): 1328 - 1333.

[2] Li Jiangfeng, Zhu Xiaoyan, Xu Jian. Empirical analysis of factors influencing digital economy based on multiple linear regression model [J]. Science and Technology for Development, 2022, 18 (08): 985 - 992.

[3] Wang Jiaming, Yang Haibin, Zhao Tianyi. Research on online predictive control method of air conditioning and refrigeration station based on temperature multiple linear regression model [J]. HVAC, 2023, 53 (02): 140 - 147. DOI: 10.19991/j. hvac1971.2023.02.22.

[4] Qin Qiusheng. Construction of a prediction model for the number of college students in universities based on multiple linear regression analysis [J]. Science and Technology Wind, 2022, No.498 (22): 152 - 154. DOI: 10.19392/j.cnki.1671 - 7341.202222049.

[5] Zhu Yu, Zheng Yiran, Yin Mo. The method of testing for multicollinearity under statistical significance [J]. Statistics and Decision Making, 2020, 36 (07): 34 - 36. DOI: 10.13546/j.cnki.tjyjc.2020.07.007.

[6] He Jing, Li Zhanjiang, Su Jinmei. Construction of green economy evaluation index system based on coefficient of variation analysis-multicollinearity test [J]. Inner Mongolia Statistics, 2017 (01): 22 - 26. DOI: 10.19454/j.cnki.cn15 - 1170/c.2017.01.008.

[7] Chen Qian. Research on the evaluation of college students' employment quality based on gray correlation analysis [J]. Microcomputer Applications, 2022, 38 (10): 71 - 74.

[8] Yin Yulong. Composition analysis of ancient glass products by correlation prediction [J]. Contemporary Chemical Research, 2023, No.126 (01): 122 - 126.

[9] Wang Zhihao, Zhao Xiangwei, Li Zhiqun. A machine learning-based method for prediction and subclassification of original composition of weathered silicate glass [J]. Journal of Silicates, 2023, 51 (02): 416 - 426. DOI: 10.14062/j.issn.0454 - 5648.20220985.

[10] Liu Zhaoyu, Sheng Hu, Dong Yingying. Research on time-varying Hurst index estimation method based on weighted average [J]. Information Technology, 2019, 43 (11): 1 - 4+9. DOI: 10.13274/j.cnki.hdzj.2019.11.001.