Analysis and Identification of the Composition of Ancient Glass Articles

Xu An [*, #], Xiaochun Zhang [#], Chenyang Wang [#]

An De College, Xi'an University of Architecture and Technology, Xi'an, China, 710311

* Corresponding Author Email: 13072959979@163.com

[#]These authors contributed equally.

Abstract. The composition analysis and identification of ancient glass products are of great significance to archaeology and cultural relics protection. Because ancient glass is easily affected by the environment in the burial process of weathering, its chemical composition will change, thus affecting the accurate judgment of its type. Therefore, this paper proposes a method of analyzing and identifying the composition of ancient glass products. According to the network open data set, this paper first use the chi-square test to study the relationship between the surface weathering of glass relics and its type, pattern and color, and then use the variance analysis to study the change of chemical composition before and after weathering, finally using the distribution matching combined with Bootstrap interval prediction method, interval prediction according to the weathering point data, the mean and 90% confidence interval of 10000 resampling prediction results as reference results. Finally, this study revealed that the surface weathering of cultural relics is significantly related to the type, and the composition change law of lead-barium and high-potassium glass after weathering, and proposed an effective composition prediction method, which provides a scientific basis for the restoration and identification of cultural relics.

Keywords: Ancient glass, Chi-square test, Analysis of variance, Distribution matching Bootstrap method.

1. Introduction

After absorbing foreign technology, ancient Chinese glass was made combined with local materials. Therefore, although its appearance is similar to that of foreign glass products, there are significant differences in chemical composition. In addition, these glass products are highly vulnerable to the buried environment and weathering. In the process of weathering, the elements inside the glass and the elements in the surrounding environment exchange a lot, resulting in the change of the proportion of its components, and then affect the accurate judgment of its category.

Yu Fei and Zhao Yanhong [1] employed the chi-square test method to analyze the influencing factors of rural e-commerce development. Li Qun et al. [2] combined the Pearson chi-square test with machine learning algorithms to evaluate debris flow susceptibility. Tu Peng et al. [3] proposed a chi-square test-based outlier suppression filtering algorithm to correct direction-finding errors. Ma Li et al. [4] used univariate analysis of variance to study plant trait variation. Zhang Yue and Wang Zhaolin[5] applied variance analysis to reveal spatio-temporal differences in tourism demand. Li Wenqing and Shi Yumei [6] adopted analysis of variance to assess air pollution characteristics. Li Qi and Xu Su'an[7] introduced a power load interval prediction method combining GSABO-BP and Bootstrap. Chen Meiling and Yu Hanjun[8] utilized the CLR transformation to address symbolic data modeling issues. Peng Lihong et al. [9] and Han Li et al. [10] integrated the chi-square test with machine learning algorithms for the classification of glass artifacts. Ye Rendao and Yang Jinan [11] applied the biased normal Bootstrap method to evaluate exposure levels. These studies have proposed effective solutions, but they are generally constrained by insufficient sample sizes, missing data, sensitive model assumptions, or high computational complexity, which may affect the robustness and practical applicability of the findings.

In view of the deficiencies of the above documents, this paper comprehensively analyzed the multi-dimensional chi-square test (type, pattern, color), and limited the single factor; adopted the Bootstrap resampling (10000 times) to improve the robustness of the prediction under small samples and the closing effect of component data; avoided the limitation of zero-value processing and data; through the normal distribution assumption and linear conversion simplified model, while retaining the reliability of interval prediction. Finally, the method proposed in this paper is both scientific and practical in the prediction of cultural relic composition.

Therefore, this paper aims to analyze the relationship between the weathering of the surface of glass relics and its glass type, pattern and color, and explore the statistical law of the chemical composition of the surface of cultural relics. Finally, the composition content of cultural relics before the weathering is predicted by the detection data of weathering points, which provides data reference for researchers in the same field to analyze and identify the types of glass cultural relics, and also provides scientific basis for the protection and restoration of cultural relics.

2. The main method

2.1. Chi-square test

The chi-square test is a statistical method used to test whether a significant difference exists between two categorical variables. It is suitable for the analysis of the relationship between definite class variables and definite class variables. The basic idea of the chi-square test is to compare the difference between the observed and expected frequencies, and to determine whether this difference is significant by constructing the chi-square statistic. The formula for calculating the chi-square statistics is:

$$x = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{1}$$

where O_{ij} represents the number of observed frequencies, E_{ij} represents the number of expected frequencies, and r and c are the number of rows and columns, respectively.

First, the null hypothesis is that H_0 is that there is no significant difference between two variables, and the variables are independent. The alternative hypothesis H_1 is a significant difference between two variables, and the variables are not independent. Then, the critical value is found from the chi-square table according to the degree of freedom and significance level. If $x^2 >$ ritical, the null hypothesis is rejected as significant difference between variables; if $x^2 \leq$ critical, the null hypothesis is accepted as no significant difference between variables.

In this study, we used the chi-square test to analyze the relationship between the degree of surface weathering of glass relics and their type, ornamentation and color.

2.2. ANOVA analysis

Analysis of variance (ANOVA) was used to test if multiple population means are equal. The principle is to determine whether there are significant differences between groups and within variance in different groups. When the variance between groups is significantly greater than the variance within groups, the mean value of different groups is significantly different; otherwise, the mean difference of different groups is considered not significant. The variance formula is calculated as follows:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \tag{2}$$

where σ^2 represents the overall variance, N represents the overall data amount, x_i represents the i-th data point, and μ represents the population mean.

First, the original hypothesis is that H_0 is equal to multiple population means, that is, there is no significant difference between different groups; the alternative hypothesis H_1 is that there are at least two means of different means, that is, significant differences between different groups of data. Then the inter-group variance and within-group variance were calculated, and then the F statistic was obtained using the quotient between intergroup variance and intragroup variance. The critical value F_α was found from the F distribution table according to the given significance level α and degree of freedom. If $F > F_\alpha$, the null hypothesis is rejected as a significant difference between different groups; if $F < F_\alpha$, accepts the null hypothesis as no significant difference between different groups.

In this study, the use of variance analysis to identify different types of cultural relics samples before weathering and weathering chemical composition content have obvious change and change rule, made clear the different types of chemical content of weathering, for the analysis and identification of glass cultural relics provides the key chemical composition data reference.

2.3. Distribution matching combined with the Bootstrap interval prediction method

Due to the lack of data before and after the weathering measurement of the same cultural relics in this study, it can only be considered as the prediction problem of unpaired samples. We need to predict the pre-weathering chemical content according to the weathering point data, and predict the pre-weathering chemical content data of glass cultural relics through distribution matching combined with Bootstrap method. We assumed that the data before and after weathering followed the normal distribution, used the normal distribution characteristics to establish the linear conversion relationship of the chemical content before and after weathering, and improved the robustness of the prediction model with Bootstrap resampling technology.

Distribution matching is used to infer the distribution or data feature of another population on a specific relationship when the distribution feature of one population is known. If the two-population data follow normal distribution, set population $A \sim N(\mu_A, {\sigma_A}^2)$,and population $B \sim N(\mu_B, {\sigma_B}^2)$, have the following linear relationship:

$$B = \mu_B + \frac{\sigma_B}{\sigma_A}(A - \mu_A) \tag{3}$$

Bootstrap Method is a non-parametric statistical method, by constructing a large number of Bootstrap samples from the original samples, so as to estimate the confidence interval of the overall parameters, relative point prediction, and the interval prediction provided by Bootstrap makes the model more accurate and robust. Assuming that the original sample is c_1, c_2 to c_m, where m is the number of samples, put back to extract m, get a sample set c_1, repeat K times, get K Bootstrap sample sets, CK=(C1, C2 to Ck), k = 1, 2, …, K, calculate the required statistics for each sample set, and then arrange all statistics in ascending order, determine the upper and lower limits of the confidence interval according to the significance level α . The Bootstrap calculation formula for the confidence interval is:

$$CI = (\bar{\theta}_{(\partial/2),} \bar{\theta}_{(1-\partial/2)}) \tag{4}$$

where $\bar{\theta}$ is the parameter estimate, ∂ is the significance level, and the two-term expression in parentheses is the quantiles of the Bootstrap sample.

3. Analog simulation

3.1. Data preprocessing

The data set of this study comes from the China Archaeology Network, a total of 58 valid glass cultural relics data, which glass relics are divided into high potassium glass and lead barium glass, each type has weathered and unweathered data. There are 6 high-potassium weathered glasses, 12 high-potassium unweathered glasses, 21 lead-barium weathered glasses, and 19 lead-barium unweathered glasses. The chemical composition of each glass data is as follows: Silica, sodium oxide, potassium oxide, calcium oxide, magnesium oxide, alumina, iron oxide, copper oxide, lead oxide, barium oxide, phosphorus pentoxide, strontium oxide, tin oxide, sulfur dioxide a total of 14 chemical components, and a variety of cultural relics decoration, color information for reference analysis.

Since the sum of the chemical composition data of cultural relics is about 100%, which has a "closed effect", direct modeling will lead to statistical bias. In order to avoid the pseudo-correlation of the composition data, we do the central log-ratio transformation (CLR) for each data, and the transformation formula is shown as follows:

$$y_i = \ln(\frac{x_i}{g(x)}) = \ln(x_i) - \frac{1}{D}\sum_{j=1}^{D}\ln(x_j) \tag{5}$$

where y_i is the value of the i-th component after the CLR transformation, $g(x)$ is the geometric mean of the original component vector x, $\ln(x_i)$ is the natural logarithm of the i-th component, and D is the total number of components.

3.2. Analyze the relationship between the surface weathering degree of glass cultural relics and its type, pattern and color

In this study, chi-square test was used to analyze the relationship between the weathering degree of the surface of glass relics and their type, pattern and color, so as to reveal the intuitive characteristics before and after the weathering of glass relics and provide external reference for the analysis and identification of glass relics.

After we import the data into the SPSSPRO analysis tool, the results are presented in Table 1:

Table.1. Results of the Chi-square test

Analytic target	Chi-square value	P value	Significant difference
Type	5.400	0.020	exist
Color	6.287	0.507	non-existent
Pattern	5.747	0.056	non-existent

As can be seen from the above table, only the surface weathering degree of glass relics and its type significance level are <0.05, indicating that there is a significant difference between surface weathering and type data, but no significant difference with color and ornamentation.

Thereafter, draw heat maps for surface weathering and type data as shown in Figure 1:

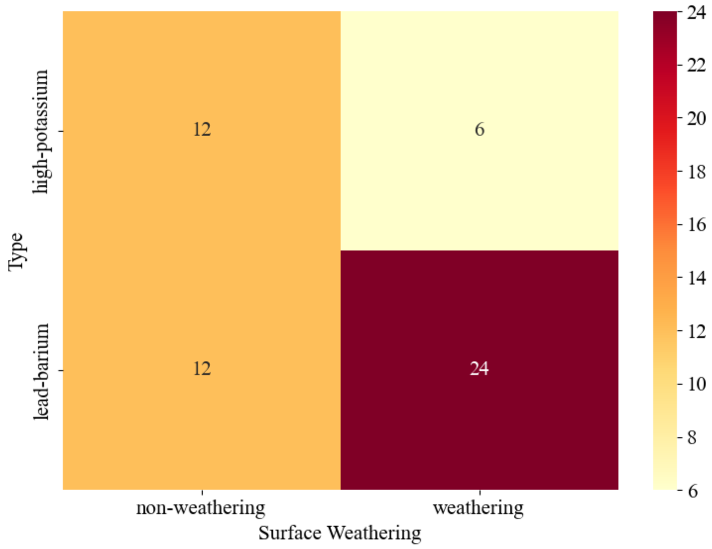

Figure 1. Surface weathering and type heat map

It can be seen from the heat map that the correlation of lead barium type and the weathering of cultural relics is strong, that is, lead barium glass is easier to weathering, and the weathered glass is mostly lead barium glass.

3.3. Analyze the statistical law of weathering chemical composition content on the surface of cultural relic samples

In view of the chemical composition content of different types of cultural relics samples, as well as the change rules, variance analysis was used in this study to provide the reference of the chemical composition data for the analysis and identification of glass cultural relics.

After importing the data into the SPSSPRO analysis tool, the results are shown in Table 2:

Table.2. Results of ANOVA for high potassium-type glass

Analytic target	P value	Significant difference
Silica	0.000	exist
Sodium oxide	0.172	non-existent
Burnt potash	0.052	non-existent
Calcium oxid	0.232	non-existent
Magnesium oxide	0.127	non-existent
Alumina	0.534	non-existent
Ferric oxide	0.697	non-existent
Cupric oxide	0.066	non-existent
Yellow lead	0.137	non-existent
Baryta	0.767	non-existent
Phosphoric anhydride	0.981	non-existent
Strontia	0.155	non-existent
Tin anhydride	0.423	non-existent
Sulfur dioxide	0.727	non-existent

Later, for the silica with significant difference, continue to explore the law of its change, and establish the boxplot as shown in Figure 2:

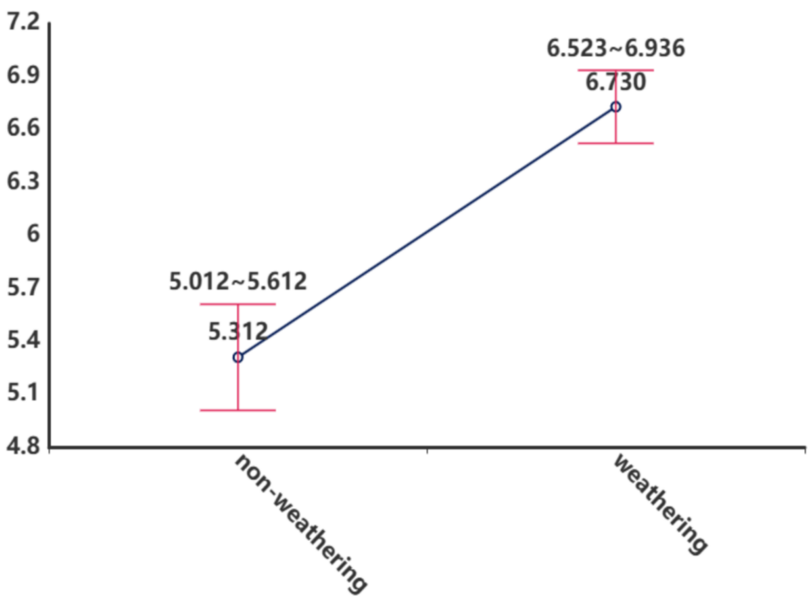

Figure 2. Box plot of surface weathering and silica

As can be seen from the Figure 2, the content of silica in high potassium glass increased from no weathering to weathering.

For lead and barium cultural relics, the results are shown in Table 3:

Table.3. Results of ANOVA for lead-barium glass

Analytic target	P value	Significant difference
Silica	0.000	exist
Sodium oxide	0.017	exist
Burnt potash	0.041	exist
Calcium oxide	0.004	exist
Magnesium oxide	0.751	non-existent
Alumina	0.004	exist
Ferric oxide	0.501	non-existent
Cupric oxide	0.872	non-existent
Yellow lead	0.011	exist
Baryta	0.648	non-existent
Phosphoric anhydride	0.001	exist
Strontia	0.142	non-existent
Tin anhydride	0.736	non-existent
Sulfur dioxide	0.952	non-existent

Later, continue to explore the variation rules of silica, sodium oxide, potassium oxide, calcium oxide, alumina, lead oxide and phosphorus pentoxide, with significant differences, and the straight-line diagram is established in Figure 3 as follows:

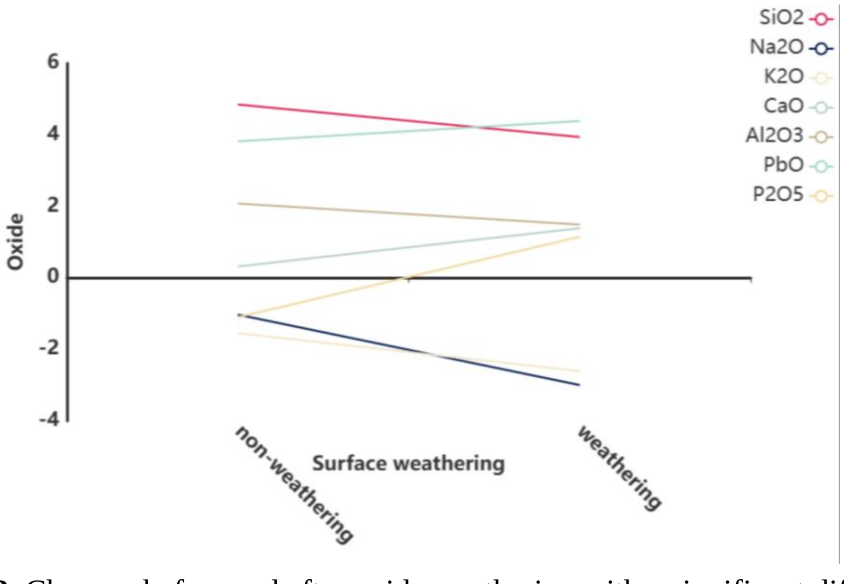

Figure 3. Changes before and after oxide weathering with a significant difference

It can be seen from the figure, the content of calcium oxide, lead oxide, phosphorus pentoxide increased, and the content of silicon dioxide, potassium oxide, potassium oxide and alumina decreased in the process of lead and barium glass from no weathering to weathering.

3.4. Predict the chemical composition content before weathering according to the weathering point detection data

For the data that have undergone the central log ratio transformation, it is assumed that the transformed weathered and unweathered data are normally distributed respectively:

$$\begin{cases} CLR(X_{weathering}) \sim N(\mu_x, \sigma_x^2) \\ CLR(Y_{non-weathering}) \sim N(\mu_y, \sigma_y^2) \end{cases} \tag{6}$$

And they meet the linear relationship:

$$CLR(Y) = \mu_y + \frac{\sigma_y}{\sigma_x}(CLR(X) - \mu_x) \tag{7}$$

Based on this formula, unweathered data can be calculated from weathered data. However, due to the small number of data sets used in this study, in order to avoid the interference of individual extreme data, Bootstrap resampling method is introduced for interval prediction to improve the robustness of the model.

For this study, taking the average value of each content of high K and lead-barium glass as an example, predict the data before weathering respectively. The steps are as follows:

Step1: Calculate the average value of high potassium and lead-barium glass;

Step2: calculate 18 high potassium glasses, resample 18 times, get 18 new high potassium glass data sets, calculate the average and variance of the new data set; 40 lead barium glass, resample 40 times, get 40 new high potassium glass data sets, calculate the average and variance of the new data set;

Step3: Using the new mean and variance, substitute the formula to calculate the predicted value and save it;

Step4: Repeat Step2 and Step310000 times, obtaining 10000 predicted values;

Step5: The predicted value is reduced to the original component percentage through the inverse CLR transformation;

Step6: Take the mean of 10000 predicted values and the 500-9500 interval (90% confidence level) as the result reference.

Follow the above steps through the Python programming solution, and the results are shown in Table 4 and Table 5:

Table.4. Prediction results of high-potassium-type glass

Chemical composition	Weighing point data	Predict the mean value of the unweathered results	Prediction unweathered result interval
Silica	93.96	82.6207	(75.4765, 88.8204)
Sodium oxide	0.01	0.1827	(0.0308, 0.5218)
Burnt potash	0.55	7.1768	(1.7795, 12.9377)
Calcium oxide	0.87	1.2762	(0.1296, 3.8049)
Magnesium oxide	0.20	0.8021	(0.2131, 1.6489)
Alumina	1.93	4.6327	(1.3661, 8.2983)
Ferric oxide	0.27	0.8830	(0.2833, 1.7773)
Cupric oxide	1.56	1.5259	(0.5989, 2.6257)
Yellow lead	0.01	0.1074	(0.0362, 0.2312)
Baryta	0.01	0.0336	(0.0126, 0.0765)
Phosphoric anhydride	0.28	0.6797	(0.2221, 1.2829)
Strontia	0.01	0.0235	(0.0154, 0.0346)
Tin anhydride	0.01	0.0220	(0.0112, 0.0478)
Sulfur dioxide	0.01	0.0339	(0.0158, 0.0652)

Table.5. Prediction results of lead-barium glass

Chemical composition	Weighing point data	Predict the mean value of the unweathered results	Prediction unweathered result interval
Silica	27.58	62.5104	(56.7309, 67.9715)
Sodium oxide	0.28	0.6188	(0.0457, 1.1360)
Burnt potash	0.15	0.1192	(0.0439, 0.2433)
Calcium oxide	2.49	0.7299	(0.2523, 1.3059)
Magnesium oxide	0.64	0.2088	(0.0439, 0.5310)
Alumina	3.11	3.9415	(2.7400, 5.5054)
Ferric oxide	0.64	0.1321	(0.0257, 0.3577)
Cupric oxide	2.33	0.8126	(0.2877, 1.5784)
Yellow lead	43.30	22.2595	(18.1133, 26.8319)
Baryta	11.32	8.2763	(4.8788, 12.0216)
Phosphoric anhydride	4.19	0.1846	(0.0402, 0.4351)
Strontia	0.39	0.1606	(0.0581, 0.3109)
Tin anhydride	0.09	0.0199	(0.0105, 0.0411)
Sulfur dioxide	0.23	0.0258	(0.0098, 0.0665)

4. Conclusions

This paper presents a comprehensive statistical analysis method for the composition change and identification difficulty of ancient glass relics due to weathering. In addition, whether weathering and type, pattern and color are significantly related, the variance analysis is determined whether the chemical composition of high potassium glass and lead-barium glass is predicted based on the data of weathering point, and provide 90% confidence interval prediction and mean prediction, which provides a scientific basis for cultural relics restoration and type identification.

Through the above methods, this study through the square test and variance analysis reveals the ancient glass weathering and type, and color and grain no significant relationship, shows that lead barium glass is easier to weathering, while the high potassium glass weathering after silica content is increased significantly, and lead barium glass in weathering regular changes of 7 kinds of oxides: calcium oxide, lead oxide, phosphorus pent, silicon dioxide, nano oxide, potassium oxide, alumina content decreased, not mentioned chemical composition content is not significant. Finally, the prediction model based on distribution matching and Bootstrap resampling is proposed, which successfully realizes the reverse inference of pre-weathering components under small samples, and provides the composition prediction results and mean value of 90% confidence interval for high potassium and lead-barium glass, respectively, and provides a quantitative basis for cultural relic restoration.

Compared with the existing research, this paper innovatively combines the central log-ratio transformation and Bootstrap resampling prediction method, effectively solves the closure effect and small sample deviation of cultural relics component data, and realizes the pre-weathering data based on weathering data. Through the multi-dimensional statistical verification (type, pattern, color) and linear distribution matching, it improves the scientific and practicability of the model, especially suitable for the scenario of data scarcity in archaeological practice, and provides a theoretical rigor and operational feasible solution for the identification and protection of cultural relics.

Although this study provides an effective statistical method for the composition analysis and weathering prediction of ancient glass relics, there are still some limitations, such as small sample size, simplified normal distribution hypothesis, and failure to consider multi-component interactions. Future work will focus on expanding the sample size, optimizing distribution fitting and prediction accuracy in combination with nonparametric methods and machine learning techniques, and exploring multivariate statistical models to reveal cooperative changes between components. In addition, we should continue to verify the universality of this method in different regions and periods of glass cultural relics and constantly optimize and improve, and finally promote the transformation of research results into cultural relics protection practice, and establish a more perfect "detection-analysis-restoration" technical system.

References

[1] YU Fei, ZHAO Yanhong. Analysis of Influencing Factors of Rural E-Commerce Development in Hebei Province Based on Chi-Square Test Method[J]. Journal of Hebei Software Institute, 2021, 23(02): 17-20.

[2] LI Qun, XU Hongjian, YANG Jin, et al. Evaluation Index Optimization Based on Pearson Chi-Square Test Algorithm for Debris Flow Susceptibility Assessment in the Bomi-Motuo Area[J]. Geological Science and Technology Bulletin, 2025: 1-16.

[3] TU Peng, XIONG Yiyin, WU Yu, et al. Direction-Finding Outlier Suppression Filtering Algorithm Based on Chi-Square Test[J]. Audio Engineering, 2023, 47(06): 148-152.

[4] MA Li, DONG Yanqiu, LI Jia, et al. Univariate Analysis of Variance on Space-Induced Mutation Impatiens[J]. Gansu Agriculture, 2005(09): 162.

[5] ZHANG Yue, WANG Zhaolin. Empirical Study on Tourist Demand Based on Variance Analysis—A Case Study of Jilin Province[J]. Modern Business, 2025(04): 37-40.

[6] LI Wenqing, SHI Yumei. Application of Variance Analysis in Real-Time Environmental Monitoring Data[J]. Economic Research Guide, 2021(26): 44-46.

[7] LI Qi, XU Su’an. Power Load Interval Prediction Based on GSABO-BP and Bootstrap[J]. Modern Electronics Technique, 2024, 47(10): 28-33.

[8] CHEN Meiling, YU Hanjun. Time Series Modeling of Distributional Symbolic Data Based on Centered Log-Ratio Transformation[J]. Statistics & Information Forum, 2024, 39(06): 3-14.

[9] PENG Lihong, WANG Yingliang, TANG Yue, et al. Statistical Analysis of Identification and Classification of Ancient Glass Artifacts[J]. Heilongjiang Science, 2024, 15(24): 102-105.

[10] HAN Li, WANG Junyin. Composition Analysis and Type Identification of Ancient Glass Artifacts[J]. Journal of Yuxi Normal University, 2024, 40(03): 78-86.

[11] YE Rendao, YANG Jianan. Bootstrap Inference for Exposure Levels Under Skew-Normal One-Way Random-Effects Model[J]. Chinese Journal of Engineering Mathematics, 2024, 41(06): 1144-1154.