

Class classification of glass products based on K-means clustering

Jie Yu ^{1, *, #}, Ruiqi Lv ^{1, #}, Sengying Zhang ^{2, #}

¹ School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, China, 450001

² School of Management, Zhengzhou University, Zhengzhou, China, 450001

* Corresponding author: 13974558428@163.com

#These authors contributed equally.

Abstract. The main raw material of glass is quartz sand, due to the high melting point of pure quartz sand, to reduce the melting temperature, it is necessary to add fluxes during refining. Different fluxes are added, and their main chemical composition is also different. To accurately classify high potassium glass and lead-barium glass, this paper establishes a K-means clustering model to classify glass types. High potassium glasses were classified into three subclasses by the content of iron oxide and barium oxide and lead barium glasses were classified into three subclasses by the content of magnesium oxide and barium oxide. Sensitivity analysis was performed on the model and the results remained significantly correlated.

Keywords: K-means, Clustering, Glass Category, Sensitivity Analysis.

1. Introduction

The Silk Road was the channel for cultural exchange between China and the West in ancient times, of which glass was a valuable physical evidence of early trade exchanges. Early glass was often made into bead-shaped jewelry in West Asia and Egypt and was introduced into China. Our ancient glass absorbed it's technology and then made locally, so the appearance of glass products with foreign similarities, but the chemical composition is not the same. As we all know, the main component of glass is silicon dioxide (SiO_2). However, due to the high melting point of silica, it is often necessary to add flux to lower the melting temperature in the glassmaking process. The fluxes commonly used in ancient times were grass ash, natural alkali, saltpeter and lead ore and limestone was added as a stabilizer, which was converted to calcium oxide (CaO) after calcium. With the addition of different fluxes, the main composition of the glass will be changed. Lead-barium glass with lead ore as flux has lead oxide (PbO) and barium oxide (BaO) as its main components. This type is usually considered to be our own invented glass variety, such as the glass of Chu culture is dominated by lead-barium glass. While potassium glass is refined with high potassium substances such as glass wood ash as flux. And this kind of glass is mainly popular in China's Lingnan as well as Southeast Asia, India and other regions. Archeologists have classified the glass products into two types, high potassium glass and lead-barium glass, based on the chemical composition of the artifact samples and other testing methods through the relevant data of ancient glass products in China. We will analyze the classification rules of high potassium glass and lead-barium glass; For each category, we will select the appropriate chemical composition to classify them into subcategories, give specific division methods and division results, and analyze the rationality and sensitivity of the classification results. The structure of this paper is as follows: The first part is the introduction, the second part is the literature review, the third part is the subclass classification and sensitivity analysis of glasses, and the fourth part discusses the conclusion.

2. Literature Review

To explore the specific classification of glass, many researchers have analyzed the chemical composition of glass by applying improved methods and techniques. Fu Xiufeng and Gan

Fuxi(2006)[1] used multivariate statistical methods to analyze and process the chemical composition data of more than 100 excavated ancient glass samples and the results showed that the ancient glass products from southern and southwestern China mainly consisted of lead-barium silicate glass and potassium silicate glass with Chinese characteristics, which could be subdivided into five major system categories: K_2O - CaO ($\sim 10wt\%$)- SiO_2 system, K_2O -system, PbO manufacture O - SiO_2 system, PbO ($\sim 25wt\%$)- O - SiO_2 system, and CaO - PbO ($\sim 40wt\%$)- SiO_2 system. In addition, Fuxi Gan, Huan-Sheng Cheng, and Qing Li(2007)[2] would determine the chemical composition of the glasses using proton-excited X-ray fluorescent technique, energy representative X-ray fluorescent technique, and inductively coupled plasma emission spectroscopy. The chemical compositions of these ancient glass samples belong to the following glass system: $Na_2O(K_2O)$ - CaO - SiO_2 ($K_2O:Na_2O < 1$), $K_2O(Na_2O)$ - CaO - SiO_2 ($K_2O:Na_2O > 1$), K_2O - SiO_2 , and BaO - PbO - SiO_2 .

Yang Bo, Zhen Xuming, and Liu Jianxin(2022)[3] proposed A pipeline impurity glass detection method based on YOLOv4 algorithms. The Kmeans++ algorithm was used to reset the a priori frame, and the high Fiatel feature map with 4 times downsampling rate was used for feature fusion to improve the detection performance of small targets. The results show that the average means accuracy (mAP) of model detection reaches 97.88%, including 95.47% for the adhesive strip glass; compared with other methods, the amount of model parameters is reduced by 31.11%, the AP for adhesive strip glass is improved by 6.70 percentage points and the detection speed is 42.82 frame/s, which can meet the real-time requirements. Therefore, the detection accuracy of the adjusted network is improved relative to the original network and the model is more lightweight and the proposed detection method is suitable for the vision part of the automated sorting impurity glass task.

3. Subcategories of high potassium glass and lead-barium glass

3.1. Data processing and analysis

According to the available data related to a batch of ancient glass objects in China, archeologists have classified them into two types high potassium glass and lead-barium glass based on the chemical composition of these artifact samples and other testing means. We used the test to study the variability of types of lead oxide and barium oxide and made scatter plots to study the classification pattern of high potassium glass and lead-barium glass. To achieve subclassification of different types of glass, we build a K-means cluster analysis model. To make the clustering results better, we gradually remove the chemical components that have little influence on the clustering and iterate sequentially. Finally, we achieve the determination of the subclassification of different glass types and analyze the characteristics of different subclassifications.

3.2. Introduction to K-means

The k-means clustering algorithm is one of the most widely used divisional clustering algorithms. k-means clustering algorithm uses distance as the evaluation index of similarity, that is, the closer two samples are, the greater their similarity, and takes getting compact and independent clusters as the ultimate goal. K-means clustering algorithm is a kind of clustering algorithm based on the idea of division, which has the advantages of simple thinking, fast clustering, and strong local search ability. The K-means clustering algorithm, a division-based clustering algorithm, has the advantages of being easy to understand, simple to implement, and fast to converge (Jiang Linzen,2022) [4]. Therefore, the K-means clustering algorithm is commonly used when the fast classification of objects is required.

The basic steps of the K-means algorithm are as follows:

Firstly, Randomly select k clustering center of mass points as $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$;

Secondly, Repeat the following procedure until convergence: $c^{(i)} = \arg \min_j ||x^{(i)} - \mu_j||^2$ For each sample I, calculate the class to which it should belong. For each class j, recalculate the center of mass of that class $\mu_j = \frac{\sum_{i=1}^m 1_{\{c^{(i)}=j\}}x^{(i)}}{\sum_{i=1}^m 1_{\{c^{(i)}=j\}}}$.

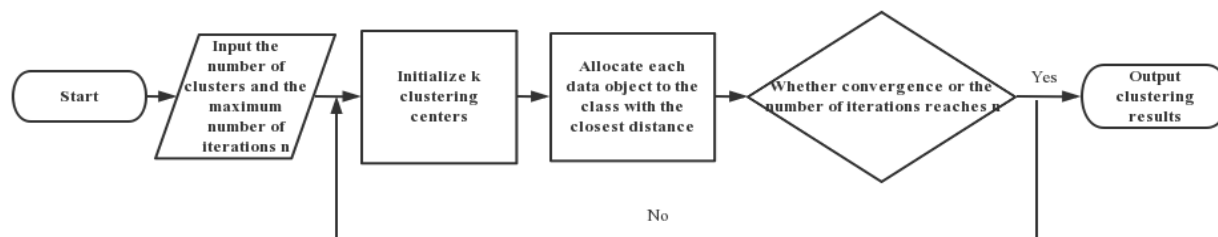


Figure 1. Basic steps of K-means clustering algorithm

3.3. Solution of the model

The t-test (known as independent sample t-test) was used to investigate the difference in type for lead oxide (PbO), and barium oxide (BaO), and the results are shown in Table 1.

Table 1. Results of t-test analysis

	Lead barium (n=49)	High potassium (n=18)	t-value	p-value
PbO	33.35	0.27	15.464	0.000**
BaO	10.49	0.40	8.364	0.000**

*p<0.05 **p<0.01

In conclusion, all samples of different types showed significant differences for PbO and BaO. The types were analyzed separately with PbO, and BaO using scatter plots.

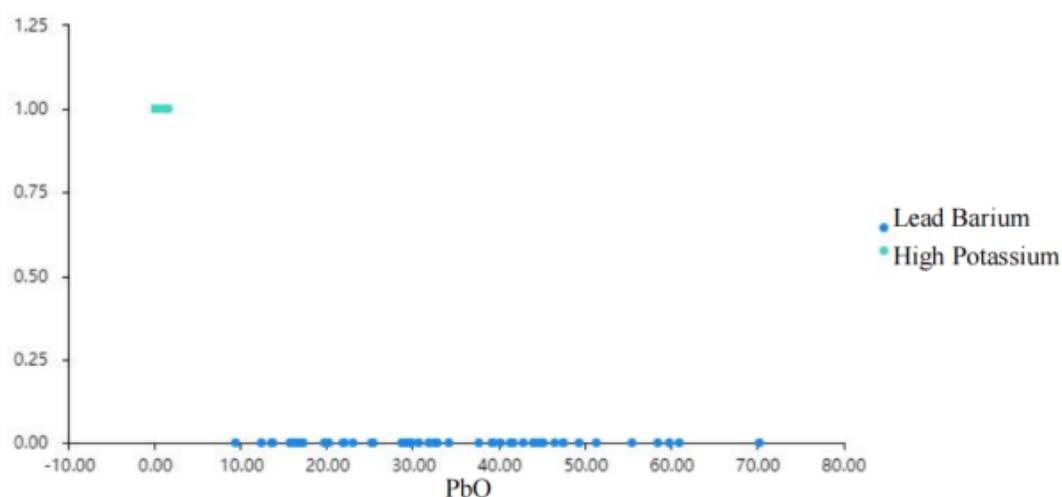


Figure 2. Scatter plot of lead oxide (PbO)

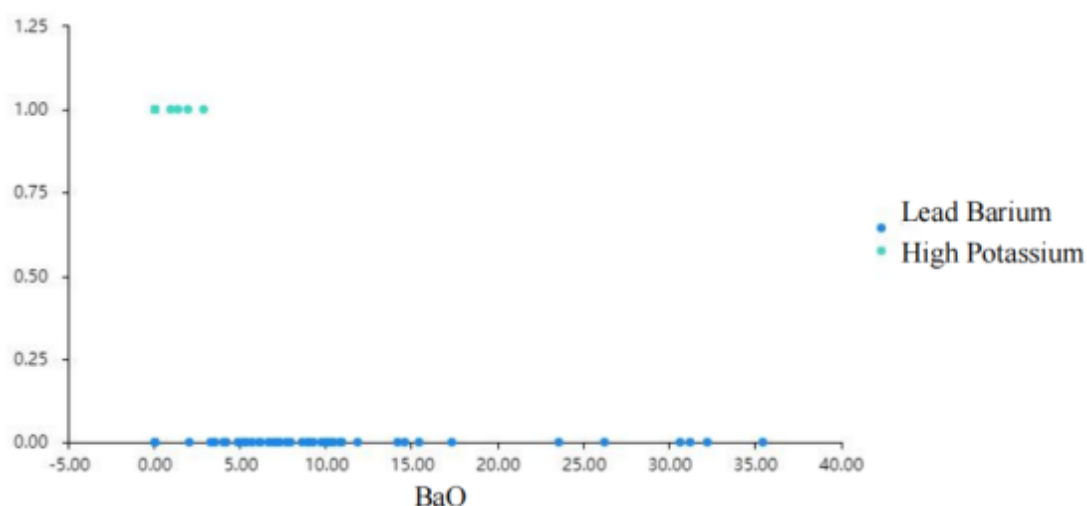


Figure 3. Scatter plot of barium oxide (BaO)

From the scatter diagram, it can be seen that lead barium glass and high potassium glass overlap for barium oxide (BaO) content, which cannot be well visualized for the first-level division. And lead oxide (PbO) two kinds of glass dividing line is obvious and can be good for the preliminary division.

3.4. Sub-categories of glass products

3.4.1. Sub-categories of high potassium glass

Each category was subclassified according to its chemical composition and solved by the K-means clustering algorithm using SPSS software, and the results are shown in Table 2.

Table 2. Results of subclassification of high potassium glass (partial)

Category	Heritage Samplin Sites
3	1
2	3.1
1	3.2
3	4
3	5
1	6.1
1	6.2
2	7
2	9
2	10

Note:3.1 indicates part 1 of artifact #3, 3.2 indicates part 2 of artifact #3, and the rest in the same way.

As can be seen from Table 9, the high potassium glass was divided into three types in the subclass division, and their percentages were 22.22%, 50.00%, and 27.78%, respectively. Overall, the distribution of the three types is more even, which indicates a good clustering effect.

To find the specific characteristics of each category, we used ANOVA to investigate the variability of each category. Through the first cluster analysis, we found that a total of six chemical contents of Na₂O, SnO, CaO, K₂O, SO₂, and PbO would not show significance ($p > 0.05$), indicating that these six chemical contents would not play a role in the cluster analysis.

After removing the above six chemical contents and performing the cluster analysis again, the results of the comparison of the variance analysis of the clustering category groups of high potassium glass obtained are shown in Table 10. As can be seen from Table 10, the clustering category groups showed significant variability for a total of five chemical contents: iron oxide, barium oxide, phosphorus pentoxide, strontium oxide, and magnesium oxide.

Table 3. Comparison results of ANOVA for clustering categories of high potassium glass

	Cluster category ANOVA variance comparison results (mean)			F-value	p-value
	Category1(n=4)	Category2(n=9)	Category3(n=5)		
Fe ₂ O ₃	3.24	0.28	1.86	12.303	0.001**
BaO	1.79	0.00	0.00	37.261	0.000**
P ₂ O ₅	2.62	0.28	1.11	9.035	0.003**
SrO	0.08	0.00	0.03	8.011	0.004**
MgO	1.51	0.26	1.15	11.503	0.001**

* $p < 0.05$ ** $p < 0.01$

From the scatter plot distribution of the clustering results of high potassium glasses, we can see that the subclassification of high potassium glasses is mainly based on the content of iron oxide and barium oxide as the classification index, in which the content of iron oxide and barium oxide is high in Category 1, low iron oxide, and barium oxide is 0 in Category 2, and high iron oxide and barium oxide is 0 in Category 3. We named the subclasses K₂O-BaO, K₂O-Fe₂O₃ (average<1), and K₂O-Fe₂O₃ (average>1) (Fu Xiufeng and Gan Fuxi,2006) [1]

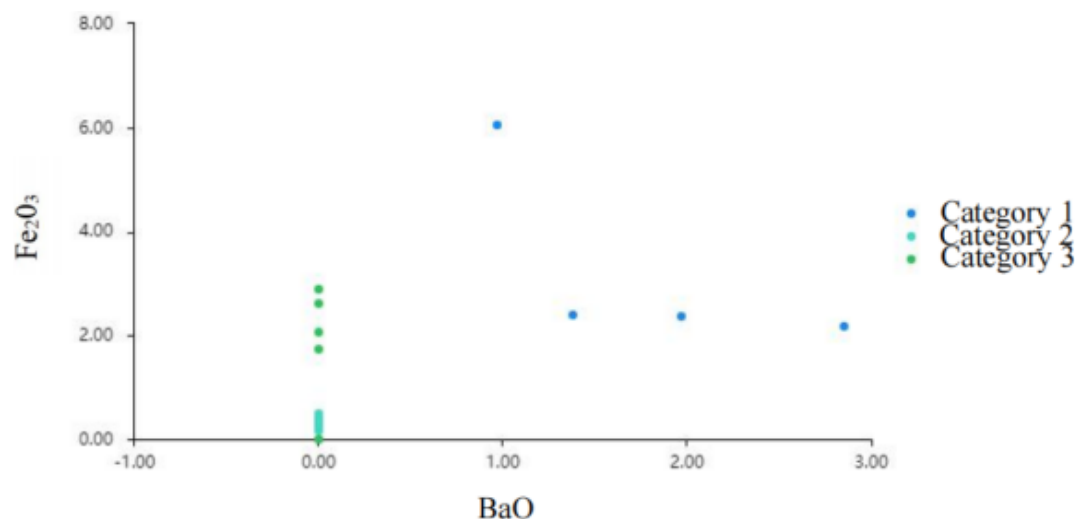


Figure 4. Scatter plot of barium oxide and iron oxide

3.4.2. Sub-categories of lead barium glass

The K-means clustering algorithm was then solved using SPSS software, and the results are shown in Table 6.

Table 4. Lead barium glass sub-classification results (partial)

Category	Heritage Sampling Sites
3	2
2	8.1
2	8.2
1	11
1	19
1	20
1	23
2	24
1	25
2	26.1

Note: 8.1 indicates part 1 of artifact #8, 8.2 indicates part 2 of artifact #8, and the rest in the same way.

As can be seen from Table 7, lead barium glass is divided into three categories in subclassification, and their percentages are 51.02%, 10.20%, and 38.78%, respectively. Overall, Category 1 accounts for more than 50%, which means that this category is to be the majority.

Through the first cluster analysis, we found that a total of four chemical contents, phosphorus pentoxide, strontium oxide, nano oxide, and calcium oxide, did not reflect significance ($p > 0.05$), indicating that the four chemical contents would not play any role in the cluster analysis process.

After removing the four chemical contents mentioned above, we performed a second cluster analysis and found that tin oxide did not reflect significance ($p > 0.05$), indicating that tin oxide content does not play a role in cluster analysis during the cluster analysis.

After removing tin oxide, we performed a third cluster analysis and found that potassium oxide did not reflect significance ($p > 0.05$), indicating that the amount of potassium oxide does not play a role in the cluster analysis process.

After removing potassium oxide, we performed the fourth cluster analysis and obtained the results of the comparison of the ANOVA differences for the clustered categories of lead and barium glass as shown in Table 7. As can be seen from Table 7 above, the clustering category groups showed significance ($p < 0.05$) for barium oxide, sulfur dioxide, magnesium oxide, and copper oxide.

Table 5. Comparison results of ANOVA for clustering categories of lead barium glass

	Comparison results of ANOVA for clustering categories (mean)			F-value	p-value
	Category1(n=25)	Category2(n=5)	Category3(n=19)		
BaO	9.92	31.16	5.81	76.337	0.000**
SO ₂	0.15	7.10	0.00	20.274	0.000**
MgO	0.28	0.00	1.29	52.485	0.000**
CuO	1.59	7.24	0.85	29.986	0.000**

From the scatter diagram of lead barium glass clusters, it can be seen that the subclassification of lead barium glass is mainly based on the content of magnesium oxide and barium oxide as classification indexes. Category 1 class has lower magnesium oxide content and barium oxide content, Category 2 has zero magnesium oxide content and higher barium oxide content, and Category 3 class has higher magnesium oxide content and lower barium oxide content. We named the subclasses PbO-BaO-MgO (average<1), PbO-BaO (average>20), and PbO-BaO-MgO (average>1) (Fu Xiufeng and Gan Fuxi,2006) [1].

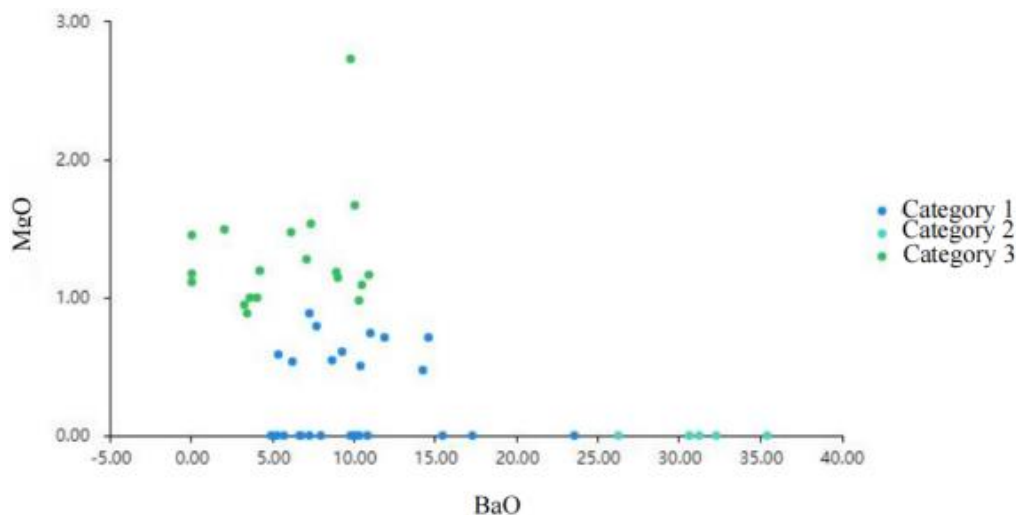


Figure 5. Scatter plot of barium oxide and magnesium oxide

3.5. Sensitivity Analysis

We conclude with a sensitivity analysis. In the context of sensitivity analysis, regression analysis involves fitting a linear regression to the model response and using the standardized regression coefficients as a direct measure of sensitivity. Therefore, this method is most appropriate when the model response is linear.

Regression analysis is used to study the influence relationship of X (quantitative or fixed class) on Y (quantitative), whether there is an influence relationship, the direction of influence, and the degree of influence situation. Iron oxide (Fe₂O₃), barium oxide (BaO), phosphorus pentoxide (P₂O₅), strontium oxide (SrO), and magnesium oxide (MgO) of high potassium glass were used as independent variables, while the subclassification type of high potassium glass was used as the dependent variable for linear regression analysis, and the model equation was obtained from SPSS analysis as

$$y_1 = 2.141 + 0.136x_1 - 0.710x_2 - 0.313x_3 - 0.437x_4 + 0.439x_5 \quad (1)$$

Where y_1 denotes high potassium glass subcategory type, x_1 denotes iron oxide (Fe₂O₃), x_2 denotes barium oxide (BaO), x_3 denotes phosphorus pentoxide (P₂O₅), x_4 denotes strontium oxide (SrO), and x_5 denotes magnesium oxide (MgO).

The model R-squared value of 0.691 implies that iron oxide (Fe_2O_3), barium oxide (BaO), phosphorus pentoxide (P_2O_5), strontium oxide (SrO), and magnesium oxide (MgO) can explain 69.1% of the variation in subclass classification.

The multiple cointegrations of the model were tested and found that all the VIF values in the model were less than 5, implying that there was no cointegration problem; and the D-W values were around the number 2, thus indicating that there was no autocorrelation in the model, and there was no correlation between the sample data, and the model was good.

Similarly, barium oxide (BaO), sulfur dioxide (SO_2), magnesium oxide (MgO), and copper oxide (CuO) were used as independent variables, while the lead-barium glass subcategory type was used as the dependent variable for linear regression analysis. The model equation was.

$$y_2 = 1.032 - 0.010x_2 + 0.074x_6 + 1.222x_5 + 0.056x_7 \quad (2)$$

where y_1 denotes high potassium glass subcategory type, x_2 denotes barium oxide (BaO), x_6 denotes sulfur dioxide (SO_2), x_5 denotes magnesium oxide (MgO), and x_7 denotes copper oxide (CuO).

The model R-squared value of 0.605 implies that barium oxide (BaO), sulfur dioxide (SO_2), magnesium oxide (MgO), and copper oxide (CuO) can explain 60.5% of the variation.

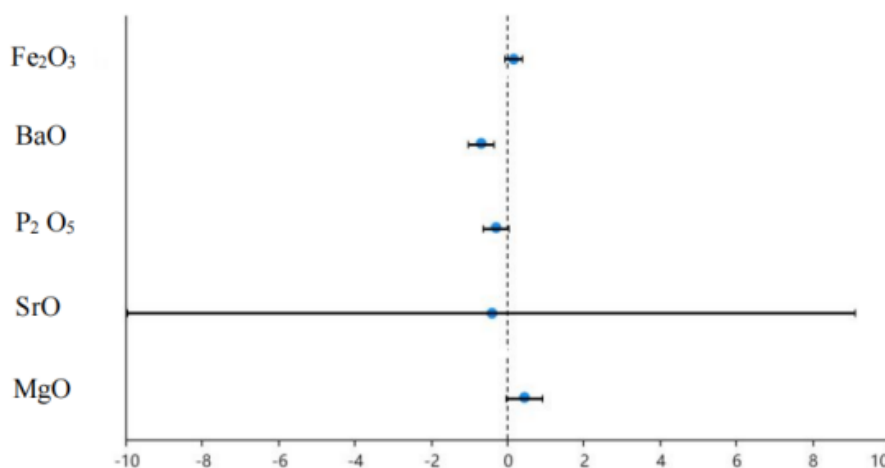


Figure 6. Confidence interval of regression coefficient of high potassium glass

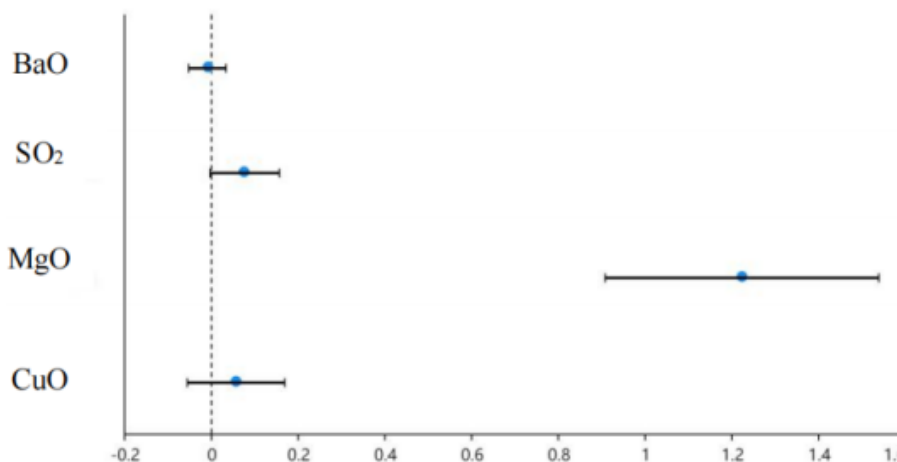


Figure 7. Confidence interval of regression coefficient of lead barium glass

4. Conclusion

In this paper, the t-test found that the content of both lead oxide and barium oxide were significantly different for glass types. It was found by scatter plot that the two type divisions did not

overlap when lead oxide was used as the classification criterion. A K-means cluster analysis model was established, and the chemical components that had less influence on the clustering were gradually removed iteratively, and the high-potassium glasses were divided into three subclasses of K_2O -BaO, K_2O -Fe $_2$ O $_3$ (average<1), and K_2O -Fe $_2$ O $_3$ (average>1) by the content of iron oxide and barium oxide, and the lead-barium glasses were divided into three subclasses by the content of magnesium oxide and barium oxide PbO-BaO-MgO (average<1), PbO-BaO (average>20), and PbO-BaO-MgO (average>1) subclasses. In the context of sensitivity analysis, iron oxide (Fe $_2$ O $_3$), barium oxide (BaO), phosphorus pentoxide (P $_2$ O $_5$), strontium oxide (SrO), and magnesium oxide (MgO) can explain 69.1% of the variation in the division of high potassium glass subclasses; barium oxide (BaO), sulfur dioxide (SO $_2$), magnesium oxide (MgO), and copper oxide (CuO) can explain 60.5% of the various reasons.

References

- [1] Fu XiuFeng, Gan FuXi. A study on the composition of a group of ancient glass from southern and southwestern China based on multivariate statistical analysis[J]. Conservation and Archaeological Science, 2006, (04): 6 - 13.
- [2] Gan FuXi, Cheng HuanSheng, Li QingHui. The Origin of Ancient Chinese Glass - A Study of the Earliest Ancient Glass in China [J]. Chinese Science (Series E: Technical Sciences), 2007 (03): 382 - 391.
- [3] Yang Bo, Xu ZM, Liu JX. Research on YOLOv4-based impurity glass sorting detection [J]. Advances in Laser and Optoelectronics, 2022, 59 (18): 322 - 328.
- [4] Jiang Linzen, Fan Xiaowei, Liu Xiangdong. A study on the initial values of K-means clustering algorithm [J]. Computer Knowledge and Technology, 2022, 18 (11): 95 - 97. DOI:10.14004/j.cnki.ckt. 2022.0698.
- [5] Chen Songlin. Exploring the research and application of standard substances for glass composition analysis [J]. China Petroleum and Chemical Standard and Quality, 2022, 42 (23): 135 - 137.
- [6] Zhang Liyan, Li Hong, Chen Shubin, Li Zhongyan, Ruan Tiachi, Xue Tianfeng, Qian Min, Fan Sijun. Simulation methods for the composition and properties of glass [J]. Journal of Silicates, 2022, 50 (08): 2338 - 2350. DOI: 10.14062/j.issn.0454 - 5648.20220115.
- [7] X. P. Dai, J.Q. Zhu, S.L. Huang, Y.B. Chen, C.C. Tang. Quantitative analysis of complex glass compositions by X-ray fluorescence spectrometry [J]. Glass Fiber, 2021 (03): 11 - 14. DOI: 10.13354/j.cnki.cn32 - 1129/tq.2021.03.003.
- [8] Duan Y. Shuai. Construction of vehicle driving conditions based on principal component analysis and K-means clustering [J]. Software Guide, 2022, 21 (05): 175 - 180.
- [9] Chen Zhuang, Jia Chenghe, Jiang Hong. Classification of convenient sticker microscopic confocal laser Raman spectra based on principal component analysis and K-means [J]. Journal of the Chinese People's Public Security University (Natural Science Edition), 2022, 28 (03): 9 - 14.
- [10] Liu Yanpeng, Zhu Lixin, Ma Shengming, Duan Jilin, Gong Qiuli. Constituent data theory and unsupervised clustering K-means method for extracting background and anomaly information--an example of Zhaogikou Pb-Zn deposit in Anhui Province [J]. Journal of Geology, 2022, 96 (11): 4038 - 4055. DOI: 10. 19762/j.cnki.dizhixuebao.2022088.
- [11] Chen Songlin. Exploring the research and application of standard substances for glass composition analysis [J]. China Petroleum and Chemical Standard and Quality, 2022, 42 (23): 135 - 137.
- [12] Zhang Liyan, Li Hong, Chen Shubin, Li Zhongyan, Ruan Tiachi, Xue Tianfeng, Qian Min, Fan Sijun. Simulation methods for the composition and properties of glass [J]. Journal of Silicates, 2022, 50 (08): 2338 - 2350. DOI: 10.14062/j.issn.0454 - 5648.20220115.
- [13] Ren Yubo, Wen Rui, Xian Yiheng, Zhang Cui Min. Chemical composition and production process of glass ear dangles excavated from Yingchengzi Han tomb [J]. DOI: 10.16334/j.cnki.cn31-1652/k.20210101995.