

Analysis of each components of glass samples based on the Spearman correlation coefficient model

Xin Zheng *, Yusi Feng, Hongkai Chen

Taiyuan University of Technology, Taiyuan, China

* Corresponding Author Email: tyut_xin@163.com

Abstract. In this paper, Spass is used to draw the scatter plot between the chemical component content, and then the Shapiro Verk test (Shapiro-Wilk test) is used to determine whether the sample follows a normal distribution. The Spearman correlation coefficient (Spearman correlation coefficient) model was then established to infer the correlation between the chemical components of the glass sample. At the same time, the chemical composition associations of different categories of glass were analyzed. This paper predicts the content of chemical composition of different types of glass based on rich data, refine, classification and use gradual regression model, establish logistic regression model and classify unknown samples; establish Spearman correlation coefficient model to obtain the correlation of different types of glass, and then analyze the difference of the correlation relations, making the results more significant and reliable.

Keywords: Glass, Shapiro-Wilk Test, Spearman Correlation Coefficient.

1. Introduction

The Silk Road is an important channel of trade between China and the West through the ages, and glass was also introduced to China through it. Ancient Chinese glass is different from the western glass system, with different chemical composition. The main raw material of glass is quartz sand, and the main component is silica (SiO_2). To reduce the melting point of quartz sand, a flux needs to be added during refining. In ancient times, the common fluxes were plant ash, natural bubble ash, saltpeter and lead ore, and limestone was added as a stabilizer [1]. Adding different fluxes, the main chemical composition of the fired glass is also different. For example, lead barium glass with lead ore as a flux has a high content of lead oxide (PbO) and barium oxide (BaO), which is usually considered to be the glass variety invented by China itself [2-3]. The glass of Chu culture is mainly lead barium glass. Potassium glass, mainly popular in Lingnan and Southeast Asia and India, is made with high potassium content, such as plant ash, as a flux [4-5].

The main raw material of glass is quartz sand, and the main component is silica (SiO_2). To reduce the melting point of quartz sand, a flux needs to be added during refining. In ancient times, the commonly used fluxes were plant ash, natural bubble ash, saltpeter and lead ore, and limestone was added as a stabilizer. Adding different fluxes, the main chemical composition of the fired glass is also different. For example, lead barium glass with lead ore as a flux has a high content of lead oxide (PbO) and barium oxide (BaO), which is usually considered to be the glass variety invented by China itself. The glass of Chu culture is mainly lead barium glass. Potassium glass, mainly popular in Lingnan of China and Southeast Asia and India, is fired with high potassium content, such as plant ash, as a flux [6-7].

We were asked to analyze the relationship between the chemical composition of the different categories of glass samples, and to compare the variability of the chemical composition associations between the different categories [8]. This idea for this question is as follows: first, using Spass to draw the scatter plot between the chemical components, preliminary to determine whether there is a correlation between the chemical composition, and then use the Shapiro Wilk test (Shapiro-Wilk test) to determine whether the sample is normally distributed, establish the Spearman correlation coefficient (Spearman correlation coefficient) model to infer the correlation between the chemical components of the glass sample. At the same time, the difference of the chemical composition association of different categories of glass was analyzed [9-10].

2. Model building and solution

2.1. Scatter plot analysis

The proportion of 14 chemical components of each sampling point was accumulated and screened, and the data that did not meet the requirements of the title was excluded, and the content statistical table of 14 chemical components of each sampling point was obtained. The data were imported into Spass to describe the statistical results, and the basic statistical characteristics of the contents of the 14 chemical components at each sampling point are shown in Figure 1.

Descriptive Statistics										
	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Variance Statistic	Skewness		Kurtosis	
							Statistic	Std. Error	Statistic	Std. Error
SiO ₂	18	59.01	96.77	76.6439	14.46673	209.286	0.205	.536	-1.816	1.038
Na ₂ O	18	0.00	3.38	0.4633	1.08871	1.185	2.127	.536	3.143	1.038
K ₂ O	18	0.00	14.52	6.4017	5.30775	28.172	-0.045	.536	-1.715	1.038
CaO	18	0.00	8.70	3.8450	3.30813	10.944	0.189	.536	-1.782	1.038
MgO	18	0.00	1.98	0.7850	0.71180	0.507	0.287	.536	-1.398	1.038
Al ₂ O ₃	18	0.81	11.15	5.0567	3.07666	9.466	0.477	.536	-0.627	1.038
Fe ₂ O ₃	18	0.00	6.04	1.3761	1.56603	2.452	1.630	.536	3.373	1.038
CuO	18	0.00	5.09	2.1556	1.49224	2.227	0.514	.536	-0.639	1.038
PbO	18	0.00	1.62	0.2744	0.51414	0.264	1.972	.536	2.723	1.038
BaO	18	0.00	2.86	0.3989	0.84163	.708	2.112	.536	3.766	1.038
P ₂ O ₅	18	0.00	4.50	1.0283	1.28060	1.640	2.154	.536	4.125	1.038
SrO	18	0.00	0.12	0.0278	0.04387	0.002	1.231	.536	-0.066	1.038
SnO ₂	18	0.00	2.36	.1311	0.55626	0.309	4.243	.536	18.000	1.038
SO ₂	18	0.00	0.47	0.0678	0.15716	0.025	2.025	.536	2.510	1.038
Valid N	18									

Figure 1. Basic statistical characteristics of each chemical composition

In order to preliminarily determine whether there is a correlation between each chemical components, Spass was used to draw out the scatter plot between each chemical components in Figure 2, so as to realize the correlation visualization.

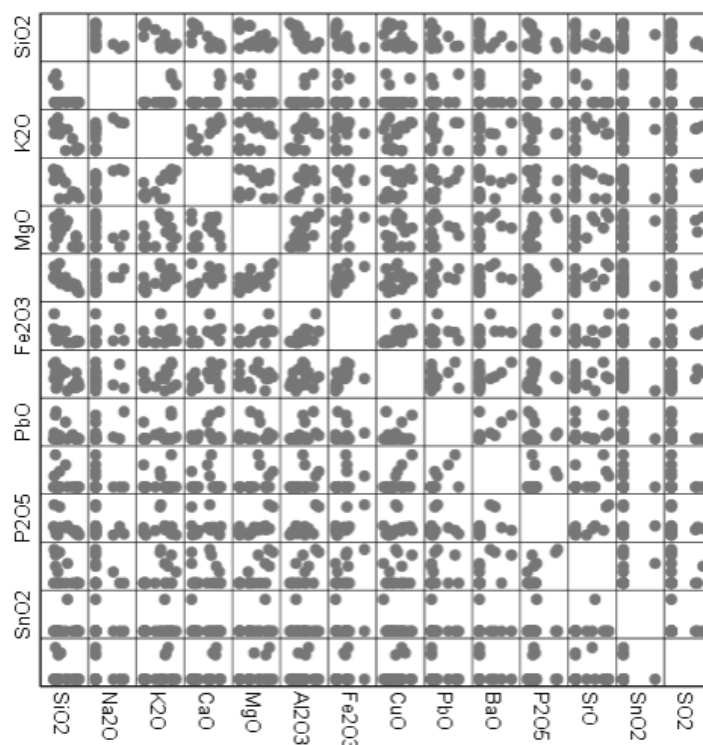


Figure 2. Scatter plot analysis of each chemical composition

It is not difficult to see from the scatter plot that most variables have no obvious linear relationship between them, so the Pearson correlation coefficient model is no longer applicable. Consider using the Spearman correlation coefficient (Spearman correlation coefficient) to further explore the correlation between the chemical components.

2.2. Correlation test

According to the statistics, the correlation between the two variables is commonly measured by Pearson's correlation coefficient (Pearson correlation coefficient) and Spearman's correlation coefficient (Spearman correlation coefficient). The Pearson's correlation coefficient is the most appropriate one if the data is continuous and is normally distributed, and the Spearman's correlation coefficient if the above conditions are not met.

The number of samples in both categories was less than 50, so the Shapiro Wilk test (Shapiro Wilk test) was used.

The test value w of this variable was calculated using SPSS, and the corresponding p -value was calculated, comparing the p -value to 0.05, and the null hypothesis can be rejected, otherwise the original hypothesis cannot be rejected.

The results show that silicon dioxide (SiO_2), sodium oxide (Na_2O), potassium oxide (K_2O), calcium oxide (CaO), magnesium oxide (MgO), iron oxide (Fe_2O_3), lead oxide (PbO), barium oxide (BaO), phosphorus pentoxide (P_2O_5), strontium oxide (SrO), tin oxide (SnO_2), sulfur dioxide (SO_2) data p value < 0.05 , that is to reject the null hypothesis, so it does not meet the normal distribution, the Pearson correlation coefficient cannot be To accurately reflect its correlation, the Spearman correlation coefficient should be used.

Spearman's correlation coefficient, used to reflect the degree of correlation of two ordered or rank variables. When calculating the Spearman correlation coefficient, it is required to rank the variable values first. After ranking the measurements of the variables X and Y for the two paired measurements, the Spearman correlation coefficient R_s has the same formula as the Pearson's correlation coefficient formula. If there is no knot (ties) after the ranking of the rank variables, the calculation formula of the Spearman correlation coefficient R_s can be simplified as:

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

$$d_i = x_i - y_i (i = 1, 2, \dots, n)$$

Because the imported chemical composition content data are not all normally distributed, the Spearman correlation coefficient is taken and the demand is hypothesis-tested. Matlab was used to obtain different types of Spearman correlation coefficients, and the results are shown in Figure 3 and Figure 4.

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1	-0.475	-0.781	-0.752	-0.546	-0.831	-0.783	-0.43	-0.443	-0.346	-0.53	-0.524	0.07	-0.284
Na2O	-0.475	1	0.606	0.64	-0.223	0.313	0.218	-0.113	0.266	-0.236	-0.221	-0.11	-0.108	-0.198
K2O	-0.781	0.606	1	0.69	0.256	0.506	0.457	0.179	0.241	-0.01	0.156	0.404	0.07	0.307
CaO	-0.752	0.64	0.69	1	0.087	0.521	0.537	0.344	0.256	-0.073	0.023	-0.052	-0.374	0.389
MgO	-0.546	-0.223	0.256	0.087	1	0.735	0.569	0.204	0.295	0.521	0.643	0.666	0.214	0.429
Al2O3	-0.831	0.313	0.506	0.521	0.735	1	0.748	0.246	0.602	0.473	0.518	0.511	-0.164	0.246
Fe2O3	-0.783	0.218	0.457	0.537	0.569	0.748	1	0.643	0.323	0.539	0.517	0.409	-0.374	0.313
CuO	-0.43	-0.113	0.179	0.344	0.204	0.246	0.643	1	0.047	0.483	0.466	0.141	-0.398	0.354
PbO	-0.443	0.266	0.241	0.256	0.295	0.602	0.323	0.047	1	0.682	0.121	0.368	-0.186	-0.341
BaO	-0.346	-0.236	-0.01	-0.073	0.521	0.473	0.539	0.483	0.682	1	0.471	0.565	-0.128	-0.236
P2O5	-0.53	-0.221	0.156	0.023	0.643	0.518	0.517	0.466	0.121	0.471	1	0.5	0.304	0.246
SrO	-0.524	-0.11	0.404	-0.052	0.666	0.511	0.409	0.141	0.368	0.565	0.5	1	0.306	-0.023
SnO2	0.07	-0.108	0.07	-0.374	0.214	-0.164	-0.374	-0.398	-0.186	-0.128	0.304	0.306	1	-0.108
SO2	-0.284	-0.198	0.307	0.389	0.429	0.246	0.313	0.354	-0.341	-0.236	0.246	-0.023	-0.108	1

Figure 3. The R value of the high potassium glass correlation coefficient

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1	0.405	0.308	-0.49	0.123	0.425	0.044	-0.447	-0.757	-0.28	-0.538	-0.552	0.098	-0.313
Na2O	0.405	1	0.067	-0.367	0.023	0.145	-0.204	-0.102	-0.357	0.087	-0.57	-0.149	-0.093	-0.209
K2O	0.308	0.067	1	-0.035	0.362	0.512	0.149	-0.198	-0.336	-0.001	-0.14	-0.167	0.14	-0.022
CaO	-0.49	-0.367	-0.035	1	0.325	0.123	0.411	-0.037	0.351	-0.157	0.502	0.308	0.284	0.078
MgO	0.123	0.023	0.362	0.325	1	0.674	0.258	-0.266	-0.076	-0.435	0.153	0.074	0.259	-0.368
Al2O3	0.425	0.145	0.512	0.123	0.674	1	0.387	-0.325	-0.389	-0.369	-0.019	-0.197	0.333	-0.385
Fe2O3	0.044	-0.204	0.149	0.411	0.258	0.387	1	-0.39	0.081	-0.443	0.225	-0.116	0.363	-0.337
CuO	-0.447	-0.102	-0.198	-0.037	-0.266	-0.325	-0.39	1	0.094	0.495	0.221	0.182	-0.356	0.449
PbO	-0.757	-0.357	-0.336	0.351	-0.076	-0.389	0.081	0.094	1	-0.1	0.361	0.362	-0.092	-0.108
BaO	-0.28	0.087	-0.001	-0.157	-0.435	-0.369	-0.443	0.495	-0.1	1	-0.169	0.182	-0.02	0.467
P2O5	-0.538	-0.57	-0.14	0.502	0.153	-0.019	0.225	0.221	0.361	-0.169	1	0.255	-0.033	0.2
SrO	-0.552	-0.149	-0.167	0.308	0.074	-0.197	-0.116	0.182	0.362	0.182	0.255	1	0.014	0.227
SnO2	0.098	-0.093	0.14	0.284	0.259	0.333	0.363	-0.356	-0.092	-0.02	-0.033	0.014	1	-0.113
SO2	-0.313	-0.209	-0.022	0.078	-0.368	-0.385	-0.337	0.449	-0.108	0.467	0.2	0.227	-0.113	1

Figure 4. R value of lead-barium glass correlation coefficient

For high potassium glass, the number of samples $n=18 < 30$ belongs to small samples. In the case of small samples, directly query the Spear, Mann rank correlation critical value table (see Appendix 11 for details) to complete the hypothesis test of the Spearman correlation coefficient: the sample correlation coefficient r must be greater than or equal to the critical value to reach a significant conclusion.

When the number of samples was 18, the cut-off value was 0.472. To reject the null hypothesis when $r > 0.472$, significant conclusions can be drawn.

For the lead-barium glass, the number of samples is $n=49 > 30$, which belongs to the large sample. In the case of the large sample, the statistic is: $r_s \sqrt{n-1} \sim N(0,1)$.

So we calculated the test value of $r_s \sqrt{n-1}$ and found the corresponding p-value compared with 0.05. A p-value is less than 0.05, rejecting the null hypothesis, significant conclusions can be drawn.

	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	BaO	P2O5	SrO	SnO2	SO2
SiO2	1	0.004	0.031	0	0.4	0.002	0.765	0.001	0	0.052	0	0	0.502	0.028
Na2O	0.004	1	0.649	0.01	0.873	0.32	0.158	0.484	0.012	0.554	0	0.307	0.526	0.15
K2O	0.031	0.649	1	0.812	0.011	0	0.306	0.173	0.018	0.997	0.338	0.252	0.337	0.878
CaO	0	0.01	0.812	1	0.023	0.402	0.003	0.8	0.013	0.282	0	0.031	0.048	0.593
MgO	0.4	0.873	0.011	0.023	1	0	0.074	0.065	0.604	0.002	0.295	0.611	0.072	0.009
Al2O3	0.002	0.32	0	0.402	0	1	0.006	0.023	0.006	0.009	0.895	0.176	0.019	0.006
Fe2O3	0.765	0.158	0.306	0.003	0.074	0.006	1	0.006	0.579	0.001	0.12	0.427	0.010	0.018
CuO	0.001	0.484	0.173	0.8	0.065	0.023	0.006	1	0.523	0.000	0.128	0.210	0.012	0.001
PbO	0	0.012	0.018	0.013	0.604	0.006	0.579	0.523	1	0.496	0.011	0.011	0.530	0.461
BaO	0.052	0.554	0.997	0.282	0.002	0.009	0.001	0.000	0.496	1	0.246	0.212	0.891	0.001
P2O5	0	0	0.338	0	0.295	0.895	0.12	0.128	0.011	0.246	1	0.077	0.823	0.168
SrO	0	0.307	0.252	0.031	0.611	0.176	0.427	0.210	0.011	0.212	0.077	1	0.924	0.117
SnO2	0.502	0.526	0.337	0.048	0.072	0.019	0.010	0.012	0.530	0.891	0.823	0.924	1	0.439
SO2	0.028	0.15	0.878	0.593	0.009	0.006	0.018	0.001	0.461	0.001	0.168	0.117	0.439	1.000

Figure 5. P-value of lead-barium glass correlation coefficient

2.3. Difference analysis of the association relationship among different types of glass

Combined with the Figures. 1,2,3,4 and 5 mentioned above, the correlation relationship of the chemical composition between the different categories of glass is very different. Overall, the chemical composition of high potassium glass is associated with less chemical components, but the correlation coefficient is generally larger and stronger. However, the chemical composition of lead-barium glass

is associated with more chemical components, but the correlation coefficient is small, and the correlation is weak.

3. Conclusion

In a progressive way, the relationship scatter chart is first made to roughly analyze the correlation relationship, and then to use the Spearman correlation coefficient model for further quantitative analysis, so that the problem solution and elaboration are more reasonable and more scientific. For the multiple linear regression model, the prediction results may differ greatly from the actual results because of ignoring the interaction effects and the non-linear causality. Although Spearman's correlation coefficient model has a wide range of applications, the accuracy decreases compared with Pearson's correlation coefficient model, which does not well represent the correlation between the two variables. Through the composition analysis, identification and prediction of glass products, the important influencing factors of ancient glass weathering can be found, which can be used for more targeted protection and restoration of unearthed cultural relics. The obtained prediction model can be used as a basis for the investigation of ancient glass age.

References

- [1] Zhao Lei. Signal Detection algorithm Based on Spearman's Simple Correlation coefficient in Pulse Noise [D]. Guangdong University of Technology, 2022.DOI: 10.27029/d.cnki.ggdgu. 2022.001595.
- [2] Zhang Weifeng. Statistical characterization of Spearman simple correlation and Kininegamma correlation coefficient [D]. Guangdong University of Technology, 2020.DOI: 10.27029/d.cnki.ggdgu. 2020.001333.
- [3] Zhang Weifeng, Xu Weichao. Analysis of the robustness of Spearman's simple correlation coefficient and Kiney Gamma correlation coefficient to pulse noise [J]. The Electronic World, the 2020(10):81-82. DOI: 10.19353/j.cnki.dzsj. 2020.10.042.
- [4] Fang Xiangzhong. The Chi-square distribution was compared with the chi-square test [J]. China Statistics, 2022 (05): 29-31.
- [5] Lai Si-Yin., PCA-based, and an improved SVDD anomaly detection algorithm [J]. Microcomputer applications, 2022,38 (08): 129-132.
- [6] Yan Qiangqiang, Zhang Min, Xun Yaling. Using a density clustering algorithm with effective proximity and adapted density [J]. Computer Technology and Development, 2022,32 (09): 14-22.
- [7] Timoneda J C. Estimating group fixed effects in panel data with a binary dependent variable: How the LPM outperforms logistic regression in rare events data [J]. Social Science Research, 2021,93:102486.
- [8] The Legend of Li, Cui Jiawei, Sun Ce, Duan Mingyin, Ma Mengdie, Yang Xingzi. Global sensitivity analysis of the SWMM model based on the partial rank correlation-stepwise regression method [J]. China Rural Water Resources and Hydropower, 2019 (01): 51-55 + 62.
- [9] Chen Yide, Liu Shujie, Zhou Yingying, Zhang Houquan. Micromicroscopic mechanical properties of CaO-Al₂O₃-SiO₂ glass [J]. Journal of Ceramics, 2022,43(02): 255-262.DOI: 10.13957/j.cnki.tcx. 2022.02.009.
- [10] James, S. N. "GLASS SUBSTRATES WITH MODIFIED SURFACE RESISTANT TO WEATHERING.", EP3874301A1. 2021.