A study of ancient glass classification problem based on multiple logistic regression

Jingping Zuo [1,*,#], Hangxin Gan [2,#], Huixin Li [3,#]

[1] School of Economics, Nankai University, Tianjin, China

[2] School of Mathematical Sciences, Nankai University, Tianjin, China

[3] Business School, Nankai University, Tianjin, China

* Corresponding Author Email: 1286986709@qq.com

[#] These authors contributed equally.

Abstract. The Silk Road was a commercial route linking China and the West in ancient times, and was an important window for China to open up to the outside world during the Han Dynasty, strongly promoting cultural exchange between China and the West. Ancient glass is susceptible to weathering by the burial environment, and in the process the proportions of its various chemical components change. This paper studies the chemical composition content of ancient glass objects, establishes a classification model, analyses the correlation between various parameters of different types of glass objects, gives statistical laws and specific classification methods, and uses glass objects to help the exchange and mutual appreciation between Eastern and Western civilizations and the longevity of Chinese civilization.

Keywords: Multiple logistic regression, Fisher's discriminant analysis, independence test.

1. Introduction

Since the official opening of the Silk Road during the Han Dynasty, glass products have been introduced into China in large quantities [1]. Since the melting point of silica, the main chemical component of glass, is very high, it is necessary to add fluxes in the refining process, and the chemical composition of different kinds of fluxes varies [2]. At the same time, ancient glass is susceptible to weathering by the burial environment, and in this process, the ratio of its various chemical components will also change [3-5]. In this paper, we study the chemical composition content of ancient glass artifacts, establish a classification model, analyze the correlation between various parameters of different types of glass artifacts, give statistical laws and specific classification methods, and use glass artifacts to help the exchange and mutual appreciation of eastern and western civilizations and the perpetual renewal of Chinese civilization [6].

We take a process data as a sample, firstly, we pre-process the data, and then fill in the missing color data by multiple logistic regression model after eliminating invalid data [7]. According to the independence test model, the correlation between the surface weathering and glass type, decoration and color of glass artifacts was analyzed, and it was concluded that there was a correlation between the surface weathering of glass artifacts and their glass type, but not between their decoration and color. Next, frequency histograms of chemical composition before and after weathering were plotted for both types of glass, and statistical patterns of chemical composition content with and without weathering were derived based on t-test models: for high potassium, the content of weathered silica was greater than that without weathering, while copper oxide and strontium oxide were not significantly different, and the rest were smaller than the value without weathering; for lead barium, the content of silica, potassium oxide, iron oxide, iron oxide for lead and barium, the contents of silicon dioxide, potassium oxide, iron oxide, and tin oxide are greater than the values without weathering, and silicon dioxide is less than the values without weathering, and the rest have no significant changes [8]. Finally, the confidence interval of the overall mean difference was found by calculating the mean values of the contents of various chemical components before and after weathering [9]. The various chemical components after weathering were subtracted from the corresponding confidence intervals to obtain the predicted values of the content of various chemical components before weathering of the artifacts.

2. The basic fundamental of Classification

2.1. Multiple logistic regression modeling

The logistic regression model is a modification of the linear regression model, both being linear models derived from the family of exponential distributions, the linear regression assumptions $Y|X$ obeying the Gaussian distribution and the logistic regression assumptions $Y|X$ obeying the Bernoulli distribution, essentially the probability of a sample falling into a certain category [10].

In linear regression the dependent variable y is generally continuous, but in classification problems, for example in the common dichotomous classification $y=0$ or $y=1$ is usually discontinuous, so in order to be able to establish the posterior probability of a classification given the independent variable x and $P(y=0)$ and $P(y=1)$, the linear regression $x^T\beta$ can be varied to establish the logistic regression hypothesis function – the Sigmond function as follows.

$$g(x)=\frac{1}{1+e^{-x^T\beta}} \tag{1}$$

For the binary logistic regression model, the error sum of squares loss in the least squares method cannot be used because the dependent variable of the dichotomous classification is non-continuous. We choose the maximum likelihood method as its parameter estimation method to fit the model and use the gradient ascent method to solve for the optimal parameters. Assuming that the dependent variables for the dichotomous classification are $y=0$ and $y=1$, we have

$$P(y=1|x,\beta)=\frac{1}{1+e^{-x^T\beta}} \tag{2}$$

Their corresponding N-sample log likelihoods are

$$l(\beta)=-\sum_{i=1}^{N}\left[(1-y_i)x_i^T\beta+\log\left(1+e^{-x_i^T\beta}\right)\right] \tag{3}$$

Extending to multicategory logistic regression, in order to use a linear function of the input features X to establish the posterior probabilities of the categories $G=1,2,\cdots,K$, and requiring the posterior probabilities of all categories to be 1 after and within [0,1], we build the following model.

$$\log\frac{P(G=K-1|X=x)}{P(G=K|X=x)}=x^T\beta_{k-1} \tag{4}$$

whose parameters are $\theta=\left(\beta_1^T,\cdots,\beta_{k-1}^T\right)$, according to $\sum_{k=1}^{K}P(G=K|X=x)=1$ which can be calculated as

$$P(G=K|X=x)=\frac{e^{x^T\beta_K}}{1+\sum_{k=1}^{K}e^{x^T\beta_K}},k=1,2,\cdots,K-1 \tag{5}$$

2.2. Modeling of independence tests

In the application process, a random sample is first selected and the observed frequency of each cell in the column table is recorded, and the expected frequency and the value of the test statistic are calculated for each cell. Afterwards, applying the rejection rule, one of the p-value method (reject H_0 if p-value $\leq \alpha$) and the critical value method (reject H_0 if $\chi^2 \leq \chi_\alpha^2$) can be chosen for correlation analysis, with α being the significance level and taking the value of 0.05 to conclude the correlation analysis [11].

3. Results

3.1. The establishment of simulation model

This paper uses a single objective planning model to analyses the calculated data. Due to limitations in testing methods, the sum of the chemical composition of the glass products in the sample data may not be 100%, and data with a sum of chemical composition between 85% and 105% are considered valid according to the conditions given in the question. We therefore removed the invalid data [12]. The 14 chemical compositions, decoration, presence or absence of weathering and type were used as independent variables, and color was used as the dependent variable. After multiple logistic regression, the vacant colors in the data were filled in for subsequent analysis of the relationship between surface weathering and glass type, decoration and cooler of the glass artefacts.

3.2. Analysis of experimental results

Based on the results of the processing of the data, the corresponding observed frequencies of surface weathering and glass type are shown in table 1 below.

Table 1. Table of observed frequencies of surface weathering and glass type

Row labels	High Potassium	Lead Barium	Total
Weathering	6	28	34
No weathering	12	12	24
Total	18	40	58

The corresponding expected frequencies were calculated and are shown in table 2 below.

Table 2. Table of expected frequencies of surface weathering and glass type

Row labels	High Potassium	Lead Barium	Total
Weathering	10.5517	23.4483	34
No weathering	7.4483	16.5517	24
Total	18	40	58

The test statistic can be calculated, with degrees of freedom of 1, whose corresponding p-values are in the range of 0.005-0.01, and since the significance level is 0.05, the original hypothesis is rejected and the conclusion that the column and row variables are not independent of each other is obtained. Therefore, there is a correlation between the surface weathering of glass artifacts and their glass type. Based on the results of the processing of the data in the annex, we can create a 2 × 3 column table to obtain the observed frequencies, as shown in the following table 3.

Table 3. Table of observed frequencies of surface weathering and ornamentation

Row labels	A	B	C	Total
Weathering	11	6	17	34
No weathering	11	0	13	24
Total	22	6	30	58

The corresponding expected frequency is calculated, as shown in the following table 4.

Table 4. Table of expected frequency of surface weathering and ornamentation

Row labels	A	B	C	Total
Weathering	12.8966	3.5172	17.5862	34
No weathering	9.1034	2.4828	12.4138	24
Total	22	6	30	58

The test statistic can be calculated, the degree of freedom is 2, whose corresponding p-values are between 0.05 and 0.10, and the original hypothesis cannot be rejected due to the significance level of 0.05, and the conclusion of independence between the column and row variables is obtained. Therefore, there is no correlation between the surface weathering of glass artifacts and their ornamentation.

Based on the results of the processing of the data in the annex, we can create a 2 × 8 column table to obtain the observed frequencies, as shown in the following table 5.

Table 5. Frequency table of observation of surface weathering and color

Row labels	Black	Blue-Green	Green	Pale Blue	light green	Deep Blue	Dark Green	Purple	Total
Weathering	2	9	0	14	1	1	5	2	34
No weathering	0	6	1	8	2	2	3	2	24
Total	2	15	1	22	3	3	8	4	58

The corresponding expected frequency is calculated, as shown in the following table 6.

Table 6. Table of expected frequency of surface weathering and color

Row labels	Black	Blue-Green	Green	Pale Blue	light green	Deep Blue	Dark Green	Purple	Total
Weathering	1.1724	8.7931	0.5862	12.8966	1.7586	1.7586	4.6897	2.3448	34
No weathering	0.8276	6.2069	0.4138	9.1034	1.2414	1.2414	3.3103	1.6552	24
Total	2	15	1	22	3	3	8	4	58

The test statistic can be calculated, the degree of freedom is 7, whose corresponding p-value is between 0.025-0.05, and since the significance level is 0.05, the original hypothesis cannot be rejected and the conclusion that the column and row variables are not independent of each other is not obtained. Therefore, there is no correlation between the surface weathering of glass artifacts and their color.

4. Conclusions

This paper describes the process of examining glass artefacts with a degree of accuracy and subtlety. A t-test is used to clearly demonstrate the relationship between surface weathering and the type, decoration and colour of the glass, in response to the small amount of data available on glass artefacts. For the same type of glass, we re-classified the artefacts by the degree of weathering at the sampling point and calculated the mean of the various chemical contents before and after weathering, solving for confidence intervals for the overall mean. For a particular type of weathered artefact, the corresponding confidence interval was subtracted from the various chemical constituent contents after weathering to obtain a range of proportions of the various chemical constituent contents before weathering for that artefact, limiting the cumulative sum of the proportions of each constituent to revert to 85%-105% to obtain a predicted value of the various chemical constituent contents before weathering for that artefact.

References

[1] Li Caiyun, Yuan Mei, Zhu Jiajia, Wang Fen, Zhong Yan, Yu Tongfu. The value of DECT in identifying mixed ground glass nodular lung tumor subtypes[J]. Diagnostic Imaging and Interventional Radiology,2022,31(04):252-258.

[2] Li M, Han LZ, Ma Juxiang, Li Q, Wang H, Ye ZX. Differential value of double-layer detector spectral CT for differentiating the invasive degree of pure ground glass nodular lung adenocarcinoma[J]. Chinese Cancer Clinics,2022,49(14):727-733.

[3] Li HY, Han Kun, Jia ZR. MSCT signs of pure ground glass density nodular infiltrative lung adenocarcinoma and differential diagnosis with pre-infiltrative lesions[J]. Chinese Journal of CT and MRI,2022,20(07):40-42.

[4] Tang Junbao, Shen Junjie. MSCT image features of mixed ground glass nodule-like lung adenocarcinoma and its differential diagnostic value[J]. Chinese Journal of CT and MRI,2022,20(07):43-45.

[5] Wang R, Dai XL, Wu YP, Qiu Tie-Feng. CT imaging performance and differential diagnosis of ground glass density in patients with pulmonary nodules[J]. Zhejiang Trauma Surgery,2022,27(03):592-593.

[6] Niu Shengbao, Zhang Hong, Li Xuesong. Analysis of the differentiation value of CT imaging technique in pulmonary microinvasive adenocarcinoma with ground glass-like nodules and pulmonary infiltrative adenocarcinoma[J]. China Medical Equipment,2022,19(06):59-63.

[7] FANG Jian, SUN Yanbing, TAO Guangyu. Effect of CT spatial resolution on the efficacy of imaging histological models to identify infiltrative pulmonary ground glass nodules[J]. Journal of Clinical and Pathological Sciences,2022,42(05):1166-1172.

[8] Zhu Weibin. Differential diagnostic value of CT imaging for the differentiation of benign and malignant pulmonary ground glass nodule-like nodules and analysis of influencing factors[J]. Imaging Technology,2022,34(02):77-80.

[9] Zhao YZ. Discussion of the accuracy of high-resolution CT signs of pulmonary ground glass nodules in the differential diagnosis of benign and malignant nodules[J]. Imaging Research and Medical Applications,2022,6(08):182-184.

[10] Zhang Yadong,Wang Zheng,Li J. Differential diagnostic value of multilayer spiral CT for infiltrating/non-infiltrating lesions in pulmonary adenocarcinoma with ground glass nodules[J]. Primary Medical Forum,2021,25(10):1431-1433.

[11] Lin Shengcheng, Liu Haihua, Huang Yonglin, Chen Lin, Gong Xiang. An analysis of the identification method of sapphire glass for watches [J]. Science and technology innovation and application,2017, (02):88.

[12] Chen Zheng,Li Zhigang,Cao Shumin. Identification of natural glass and glass [J]. Journal of Gemstone and Gemmology,2007, (01):22+42.