

基于统计分析的古代玻璃制品的化学成分含量模型研究

兰奕珩

成都锦城学院建筑学院 四川成都

【摘要】古代玻璃制品的表面特征及化学成分含量进行玻璃制品的分析与鉴别,对所给数据进行预处理,基于多元线性回归建立相关统计模型,进行统计分析古代玻璃表面有无风化化学成分规律。使用处理完成后的数据,完成古代玻璃表面特征与其是否分化的 Spearman 秩相关关系,得到相关系数,其中玻璃类型的相关性系数最高,再运用卡方独立性检验验证其结论正确,并得到玻璃表面特征与是否风化之间都具有中等程度差异性。使用独立样本 T 校验得到氧化钠,氧化钾,氧化铜,氧化铅,二氧化硫的含量是有无风化化学成分分析统计的关键。

【关键词】Spearman 秩相关; 线性回归; 古代玻璃制品

【收稿日期】2022 年 11 月 25 日 **【出刊日期】**2022 年 12 月 28 日 **【DOI】**10.12208/j.jccr.20220018

Research on component content model of ancient glass products based on statistical analysis

Yiheng Lan

School of Architecture, Chengdu Jincheng College, Sichuan Chengdu

【Abstract】The surface characteristics and chemical composition content of ancient glass products are analyzed and identified. The given data are preprocessed, and a relevant statistical model is established based on multiple linear regression to statistically analyze whether there are weathering chemical composition rules on the surface of ancient glass. Using the processed data, complete the Spearman rank correlation between the surface characteristics of ancient glass and whether it is differentiated, and obtain the correlation coefficient, among which the correlation coefficient of glass type is the highest, and then use the chi-square independence test to verify that the conclusion is correct, and obtain the glass surface There is a moderate degree of variability in character and whether or not it is weathered. Using the independent sample T calibration to obtain the contents of sodium oxide, potassium oxide, copper oxide, lead oxide and sulfur dioxide is the key to the analysis and statistics of chemical composition with or without weathering.

【Keywords】Spearman rank correlation, linear regression, ancient glassware

1 引言

玻璃和大气作用发生侵蚀称为风化,玻璃埋在土中(如古墓内的葬品)受到侵蚀,也叫做风化,玻璃是早期贸易往来的宝贵物证,我国古代玻璃制作工艺就地取材制作,化学成分与其他国家不同,具有重要研究意义。从化学作用方向研究玻璃风化是传统研究课题,通过数学建模及计算机模拟风化过程,从理论上彻底弄清玻璃文物风化过程及风化产物,帮助学者更快得出结论。探究玻璃文物的表面风化与其类型、纹饰和颜色的关系及其相关性的结果,在玻璃类型不同的情况下,对表面有无风化化学成分含量的统计规

律进行分析,结合回归方程等方法预测风化前的化学成分含量。研究高钾玻璃、铅钡玻璃的分类规律,再用一定的划分方法对其进行亚类划分,并得出结果,再分析分类结果的合理性和敏感性。分析未知类别玻璃文物的化学成分,根据建立的分类模型对其进行分类,并对分类结果进行敏感性分析。讨论不同类别的玻璃文物样品其化学成分之间的关联关系,比较化学成分关联关系的差异性。

根据数据整理得出数据为分类变量,所有使用 Spearman 相关系数进行相关性分析,分别得出玻璃文物的表面风化与其类型、纹饰和颜色的关系模型,

并运用卡方检验进行差异性分析,研究不同分类变量之间的关系最后得出结论。指出影响统计表面有无风化化学成分含量的规律,并对其进行分析得出风化前后化学成分变化规律,得出合适的回归模型,推测风化前化学成分含量的具体值。依据附件数据来构建机器学习分类模型,运用神经网络分类预测及可视化较好的决策树分类模型,利用化学成分含量不同分析高钾玻璃、铅钡玻璃这两种类型的分类规律。选择合适的化学成分对每个类别进行亚分类,先对化学成分进行聚类分析,得出聚类模型,解释模型的合理性及分类过程中调整模型参数,运用方法分析模型敏感性产生的变化。对未知类型古代文物数据进行整理,带入建立的分类模型进行分类,得出八个样本的分类结果,运用均方根对分类结果的敏感性进行分析。对化学成分运用灰色关联分析和相关性分析,对不同类别的玻璃文物样品其化学成分之间的灰色关联度相关系数进行分别计算,得出相关热力图直观进行对比得出不同类别之间的化学成分关联关系的差异性。

2 数据预处理

将玻璃文物的表面风化与其玻璃类型、纹饰和颜色的数据进行编码量化,用数字 0123456 分别代替是否风化,玻璃的不同类型、纹饰和颜色,删除缺失值。保留成分比例累加和介于 85%-105%之间的有效数据,缺失值设为固定值 0。玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系,看其是否具有相关性。

$$p = \frac{\sum_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i(x_i - \bar{x})^2 \sum_i(y_i - \bar{y})^2}} \quad (1)$$

其中,其中 x_i, y_i 为原始变量 X_i, Y_i , 经过等级映射得到,具体映射方式可以取:每个原始数据的降序位置的平均,记 $d_i = x_i - y_i$, 其简化的计算公式为:

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

计算出相关系数,由于量化后的数据非正态分布,表面风化与玻璃类型之间 Spearman 相关系数为 0.316, p 为 0.02 具有显著性,两变量之间具有相关性,表面风化与玻璃纹饰之间 Spearman 相关系数为 0.128, p 为 0.358 不具有显著性,两变量之间不具有相关性。表面风化与玻璃颜色之间 Spearman 相关系数为 -0.112, p 为 0.421 不具有显著性,两变量之间不具有相关性。对表面风化与其玻璃类型、纹饰和颜色的差异性进行分析,运用卡方独立性检验。

$$\chi^2 = \sum \frac{(A-T)^2}{T} \quad (3)$$

(3) 式中: A 是实际值, T 为理论值, χ^2 值的意义:衡量理论与实际的差异程度,根据卡方值来判断是否有差异,查询卡纸分布的临界值,将计算的值与临界值比较,卡方检验分析的结果显示,对于表面风化,显著性 P 值为 0.020**, 水平上呈现显著性,拒绝原假设,因此对于表面风化和类型数据存在显著性差异,因此存在相关性。对于表面风化,显著性 P 值为 0.507, 水平上不呈现显著性,接受原假设,因此对于表面风化和颜色数据不存在显著性差异,两者不相关。对于表面风化,显著性 P 值为 0.056*, 水平上不呈现显著性,接受原假设,因此对于表面风化和纹饰数据不存在显著性差异,两者不相关,且结果与相关性分析结果相同。

效应量化分析的结果显示,分析项:纹饰 Cramer's V 值为 0.326, 因此纹饰和表面风化的差异程度为中等程度差异,类型 Cramer's V 值为 0.316, 因此类型和表面风化的差异程度为中等程度差异,颜色 Cramer's V 值为 0.341, 因此颜色和表面风化的差异程度为中等程度差异。

3 回归模型的建立

由于玻璃类型与风化具有相关性,根据定类变量有无风化对定量字段化学成分进行分组,分别检验其正态性检验,查看数据的总体分布是否呈现正态性分布,再进行方差齐性检验,查看 P 值是否小于 0.05,使用方差分析,查看 P 值是否呈显著性,若呈显著性,可以根据均值 \pm 标准差的方式对差异进行分析,反之则表明不呈现差异性,若独立样本 T 检验呈显著性,借助效应量化分析对差异性进行量化分析。对各指标成分进行正态性分布检验,得出各化学成分的中位数,平均值,标准差,偏度,峰度;数据都不满足正态分布。二氧化硅,氧化钙,氧化镁,氧化铝,氧化铁,氧化钡,五氧化二磷,氧化锶,氧化锡,不存在显著差异,氧化钠,氧化钾,氧化铜,氧化铅,二氧化硫,存在显著差异。这是对玻璃类型整体起作用的规律,表现有无风化化学成分含量整体的本质和必然的联系。

多元线性回归模型建立,使用最小二乘法建立函数关系,寻找最佳参数,计算模型预测值与真实值之间的差异程度,对风化前后的数据进行最佳拟合。将特征影响因素是否风化与预测值,下式描述变量之间的关系:

$$y_i = b + \sum_{d=1}^p \omega_d x_i, d \quad (4)$$

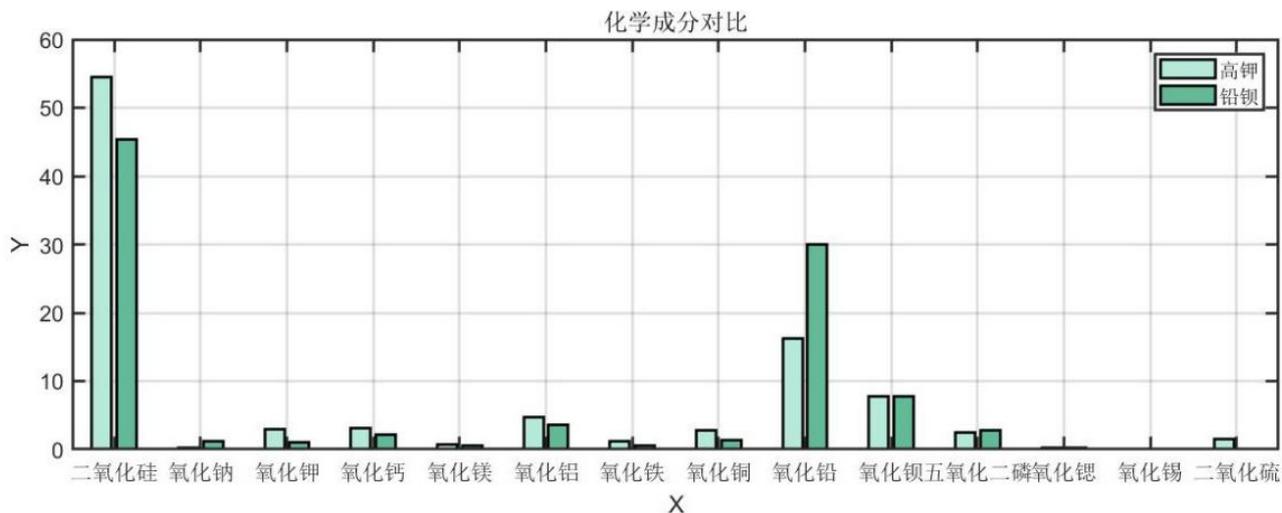


图 1 各化学成分对比图

4 结果

运用 PyCharm 软件进行计算得出风化前后各元素的线性回归模型，得出二氧化硫，氧化钠，五氧化二磷，氧化铜，氧化铁这些化学元素因为不能拒绝原假设，所有回归模型无效，则取其含量的平均值，分别为 0.603，0.786，2.684，1.954，0.849。

参考文献

[1] 王承遇,陶瑛.硅酸盐玻璃的风化[J].硅酸盐学报,2003(01):78-85.
 [2] 许凤雯,狄靖月,李艳,阮燕云,包红军.长江上游流域面雨量与主要水电站发电能力关系分析[J].水电能源科学,2021,39(10):22-26.

[3] 王重阳,马菁华,吴睿钰,严泽凡.基于 BP 神经网络的水华预测模型及其敏感性分析
 [4] 吴云,雷建文,鲍丽山,李春哲.基于改进灰色关联分析与蝙蝠优化神经网络的短期负荷预测[J].电力系统自动化,2018,42(20):67-72.
 [5] 贺超峰,华心祝,马菁花,曹永模,陈康.基于 BP 神经网络的回采巷道围岩稳定性分类[J].矿业工程研究,2012,27(03):6-9.

版权声明：©2022 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>

