
古代玻璃制品的成分分析与鉴别

摘要

本文主要研究古代玻璃制品成分受风化影响导致化学成分变化的现象,根据不同文物表面风化程度与文物特征及相关数据之间的关系,量化附件数据后,利用统计学,相关性分析,独立样本 T 检验、随机森林分类算法等方法,分析了变量之间的相关程度,预测化学成分含量及分布规律。

针对问题一,根据统计学、相关性分析得出表面风化与纹饰、颜色、玻璃类型的关系。在运用相关性分析时需选择合适的相关系数,由于数据是离散的且是定类数据,不满足运用 person 相关系数的条件,故选择 spearman 相关系数对数据进行相关性分析。再次利用相关性分析确定与文物表面风化有关的化学成分,根据独立样本 T 检验获得化学成分分别在玻璃文物类型及文物表面有无风化的含量占比,并利用文物风化前后化学成分比值,建立预测风化前化学成分含量的模型。

针对问题二,根据随机森林分类算法,得出高钾玻璃以及铅钡玻璃分类中占比含量高的化学成分,得出分类规律。对表格二的数据进行汇总,分别分离出高钾玻璃与铅钡玻璃的数据,使用与文物风化相关度高的成分进行不同类别文物的聚类分析,得出划分方法与结果并对方法进行合理性与敏感性验证。

针对问题三,根据问题二得出的类型分类规律,将表格三的有效数据带入,预测出玻璃文物的类型,然后使用 Lasso 回归建立数学预测模型,将有效数据导入,验证结果的准确性。

最后针对问题四,由于两类玻璃文物的化学成分含量符合连续变化的要求且在自然环境中符合正态分布,使用相关性分析的 person 相关系数算法,得出各自关联性,对结果进行统计分析,得出两者的差异性。

关键词: 玻璃风化 化学成分 统计学 相关系数 Lasso 回归模型 分类算法

1、问题重述

1.1 问题背景

我国古代玻璃制品埋藏在地下时，极易受周边环境影响发生风化，导致内部化学成分与外部环境发生交换，掺杂一些无关元素进入并损失内部元素，从而影响或误导人们对其类别及成分的分析。因此需要借助相关数据处理手段，分析玻璃制品表面特征之间的关系，根据已有数据预测其丢失的数据，给出风化后文物类别确定的依据，从而达到文物化学成分数据预测以及关系分类验证的目的。

附件中的数据给出了古代玻璃制品的相关处理数据：附件 1 是将玻璃制品表面特征进行量化处理，并将无效数据进行删除；附件 2 是不同玻璃制品缺省数据处理（空缺值为 0）。

1.2 问题提出

根据以上背景，以及本题给出的附件的三个表单，需要解决以下问题：

- (1) 对这些玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系进行分析。
- (2) 结合玻璃的类型，分析文物样品表面有无风化化学成分含量的统计规律。
- (3) 根据风化点检测数据，预测其风化前的化学成分含量。
- (4) 依据附件数据分析高钾玻璃、铅钡玻璃的分类规律。
- (5) 对于每个类别选择合适的化学成分对其进行亚类划分，给出具体的划分方法及划分结果，并对分类结果的敏感性进行分析。
- (6) 针对不同类别的玻璃文物样品，分析其化学成分之间的关联关系，并比较不同类别之间的化学成分关联关系的差异性。

2、问题分析

2.1 对问题一的分析

问题一分为三部分，第一部分要求分析玻璃文物的表面风化与其玻璃类型、纹饰和颜色的关系，由于附件 1 的数据是不连续的，属于定类变量，所以选择 spearman 相关系数进行数据间关系分析。第二部分要求结合类型统计有无表面风化的化学成分规律，先使用相关性分析确定有关的化学成分，再使用独立样本 T 检验获取风化前后化学成分变化，再进行统计学进行规律统计。第三部分根据风化前后的化学成分比值获得风化前的成分含量分别得到各个成分的模型公式，将公式累加即得总化学成分含量。

2.2 对问题二的分析

问题二分为两部分，第一部分要求分析不同类别玻璃制品的分类规律。也就是说要求分析出将玻璃制品分为这两类的依据，我们初步处理附件数据，将缺省的数据处理为 0，再通过随机森林算法，计算出不同类别玻璃制品中化学成分占比最大的化学成分，并根据该化学成分取值区间的不同得出分类规律。对于问题二，由于需要将每个类别的化学成分进行分类，需要将不同类别的化学成分单独处理，根据聚类算法，将每个类别的化学成分分为两类，得到两类化学成分取值区间，以此区间依据进行分类，观察分析是否存在区间重叠来分析结果的合理性与敏感性。

2.3 对问题三的分析

问题三要求依据化学成分对未知类型的玻璃制品进行推测，我们先依据问题二的分类规律，推测出该玻璃制品的类型，再根据 Lasson 回归，将类别量化作为因变量，以化学成分作为自变量，带入化学成分数据，求出量化的类别，再将类别转换，分析推测是否合理，并根据化学成分区间是否重叠来分析其敏感性。

2.4 对问题四的分析

问题四求两种类型的化学成分的关联性，初步采用灰度关联分析发现效果不符合预期，考虑到数据属于连续性，在自然中符合正态分布，考虑使用相关性的 person 相关系数检验，结合两种玻璃的组成和风化化学反应的不同对差异性进行分析。

3、模型假设

- 1、相关性分析模型假设：假设附件中化学成分数据变化大致成线性关系。
- 2、独立样本 T 检验模型假设：每个数据样本是独立分布的、数据整体分布呈现正态分布。
- 3、数据样本的分布呈随机性。

4、模型的建立与求解

4.1 问题一

4.1.1 玻璃文物表面风化与其玻璃类型、纹饰、颜色的关系分析：

为了探究文物表面风化与文物其他三个特征之间的关系，在其之间建立相关性分析，注意到附件中表格的数据是定类并且离散的，我们先使用 SPSS 软件对数据进行量化处理并剔除无效的数据，在分别对文物有无风化、玻璃类型、纹饰、颜色进行排序，分别求得秩次。再使用斯皮尔曼相关系数，根据以下公式分别计算表面是否风化与其它三个特征的相关系数（ d_i 代表两两秩次之差）。

相关性分析（斯皮尔曼算法）^[1]：

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

表 1 表面风化与颜色、类型、纹饰之间的相关系数

	表面风化	颜色	类型	纹饰
表面风化	1.000(0.000***)	-0.017(0.891)	0.385(0.002***)	0.035(0.780)

表 1.1 不同颜色的秩次

	蓝绿	浅蓝	紫	深绿	深蓝	浅绿	黑	绿
秩次	9	28.5	42.5	49	54	57	60.5	63

表 1.2 不同类别的秩次

	高钾玻璃	铅钡玻璃
秩次	9.5	40.8

表 1.3 不同纹饰的秩次

	A	B	C
秩次	44.5	60.5	16

表 1.4 有无表面风化的秩次

	表面风化	表面无风化
秩次	44.5	13

表面风化与颜色、类型、纹饰之间的相关系数分别为-0.017、0.385、0.035(数值越大，相关程度越高)，根据数值可知，表面风化与文物类型具有强相关性，表面风化与颜色、纹饰相关性并不显著。再将 excel 表格中四个数据进行柱状图统计分别为图 4-1、图 4-2 以及图 4-3。

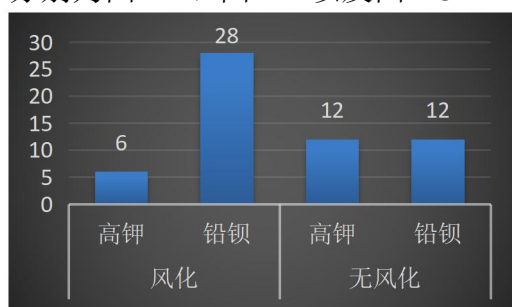


图 4-1

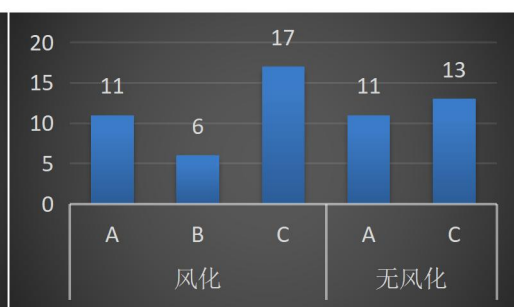


图 4-2

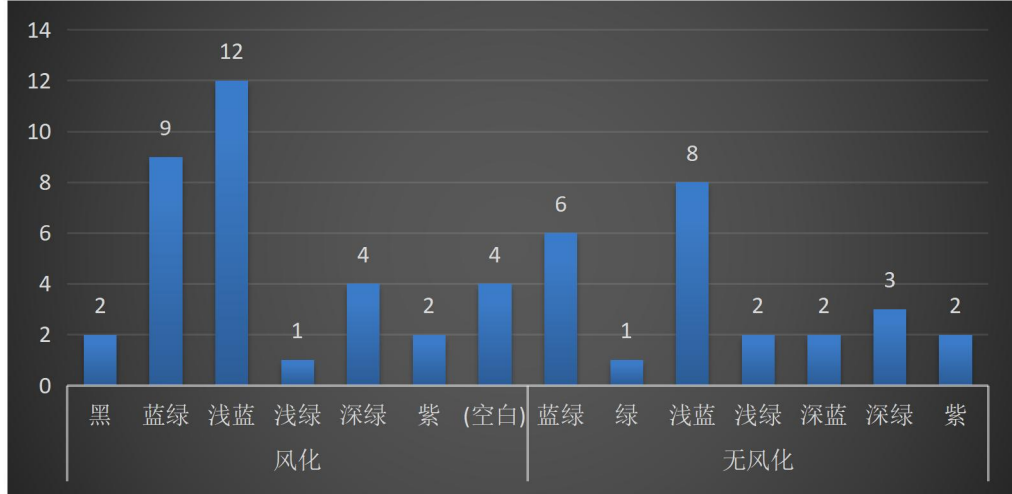


图 4-3

根据图表统计可以明显看出，文物类型中铅钡发生表面风化占比大，纹饰 B、纹饰 C 易发生风化，风化前后主要以浅蓝、蓝绿色为主。

4.1.2 结合类型，文物样品表面有无风化化学成分含量规律：

在分析化学成分含规律之前，先使用相关性分析（斯皮尔曼相关系数）得出与文物表面有无风化相关程度较强的化学成分，剔除其他无关成分的影响。

由相关化学成分的数据（见附录 1）可知，表面有无风化与主要与氧化钾（K₂O）、氧化铝（Al₂O₃）、氧化铁（Fe₂O₃）、氧化铅（PbO）、氧化钡（BaO）、氧化锶（SrO）、二氧化硅（SiO₂）这些化学成分关联程度大，后文将以这些成分作为文物化学成分进行分析。为了分析不同类别，有无风化化学成分含量的统计，分别对不同类型以及有无风化使用独立样本 T 检验^[2]，得出化学成分含量的规律：

独立样本 T 检验：

$$t = \frac{m_A - m_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \quad (2)$$

$$S^2 = \frac{\sum (x - m_A)^2 + \sum (x - m_B)^2}{n_A + n_B - 2} \quad (3)$$

$$df = \left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B} \right) / \left(\frac{S_A^4}{n_A^2(n_B - 1)} + \frac{S_B^4}{n_B^2(n_B - 1)} \right) \quad (4)$$

根据类别分类得出以下统计：

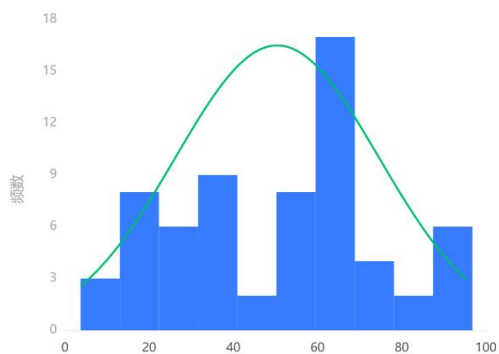


图 4-4 二氧化硅(SiO₂)

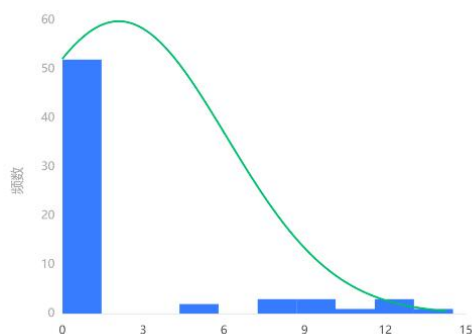


图 4-5 氧化钾(K₂O)

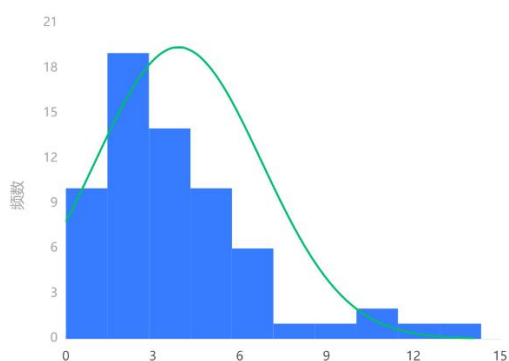


图 4-6 氧化铝(Al₂O₃)

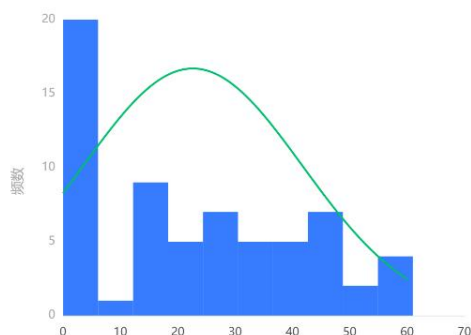


图 4-7 氧化铅(PbO)

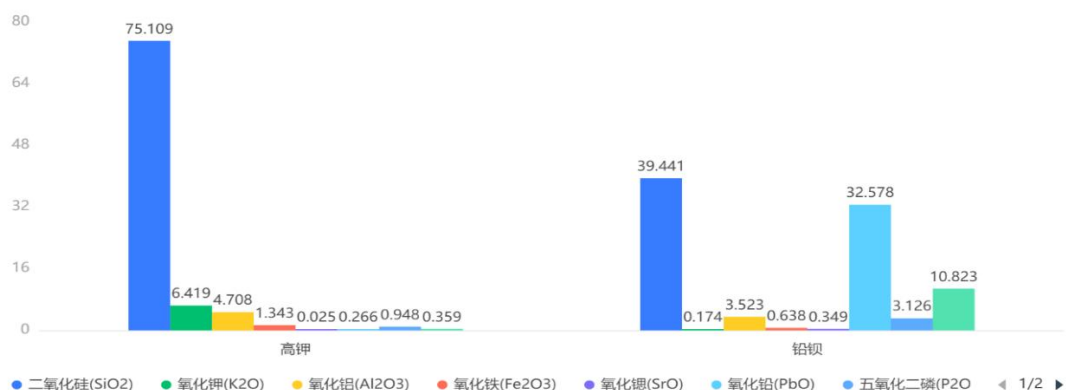


图 4-8 风化前后高钾类型文物所含化学成分占比（左图为风化前，右图为风化后）

可以根据以上图表统计可得规律：高钾类型的文物玻璃制品中，二氧化硅、氧化钾、氧化铝化学成分占比大，而铅钡中二氧化硅、氧化铅、氧化钡、化学成分占比大，详细数值参见附录 2。

根据文物表面有无风化得出以下统计：

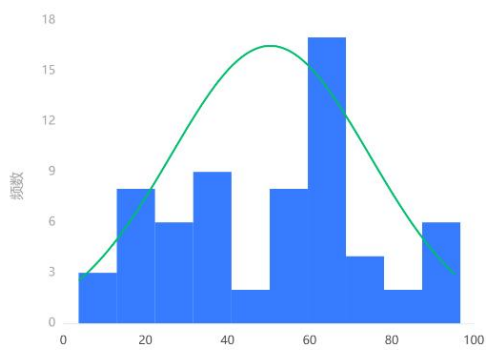


图 4-9 二氧化硅(SiO₂)

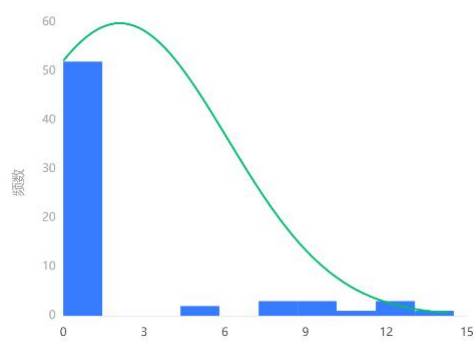


图 4-10 氧化钾(K₂O)

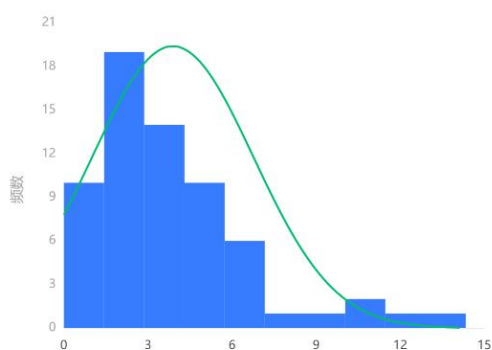


图 4-11 氧化铝(Al₂O₃)

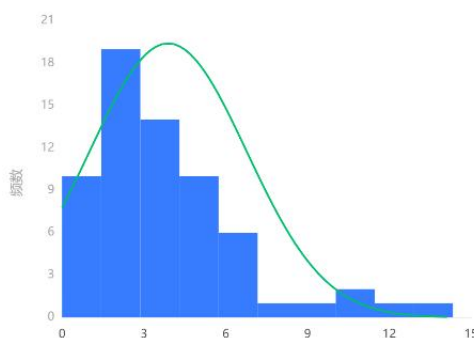


图 4-12 氧化铅(PbO)

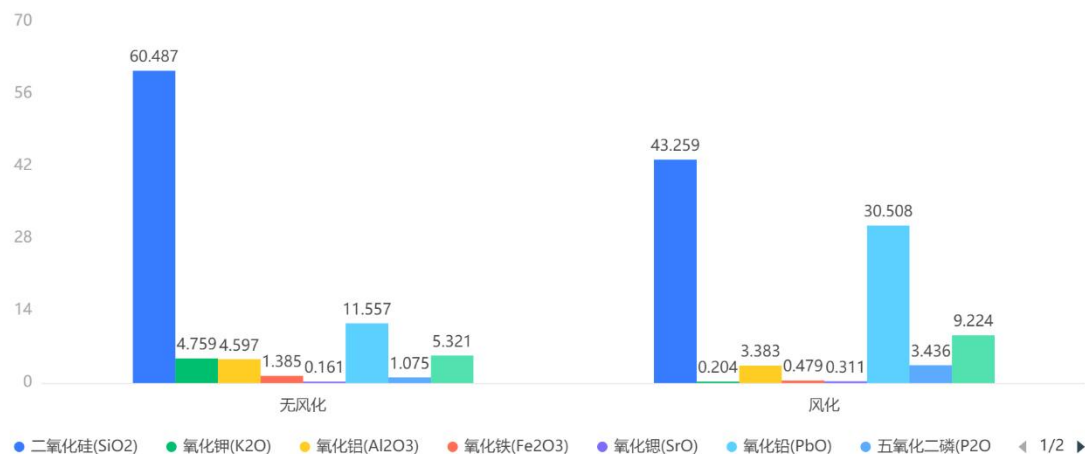


图 4-13 风化前后铅钡类型文物所含化学成分占比（左图为风化前，右图为风化后）

根据以上表格统计可得规律：风化前后二氧化硅、氧化铅、氧化钾、氧化钡发生显著性差异。详细数值参见附录 3。

对比以上有无风化及类别中的化学成分含量可以简略得出结论：高钾类文物制品中，风化前后二氧化硅、氧化铅、氧化铝、氧化钾含量发生较大量的交换，铅钡类型中，风化前后二氧化硅、氧化铅、氧化钡含量发生较大量的交换。

4.1.3 预测不同类别风化前化学成分含量：

由于风化前后化学成分含量发生变化，做出成分散点分布图：

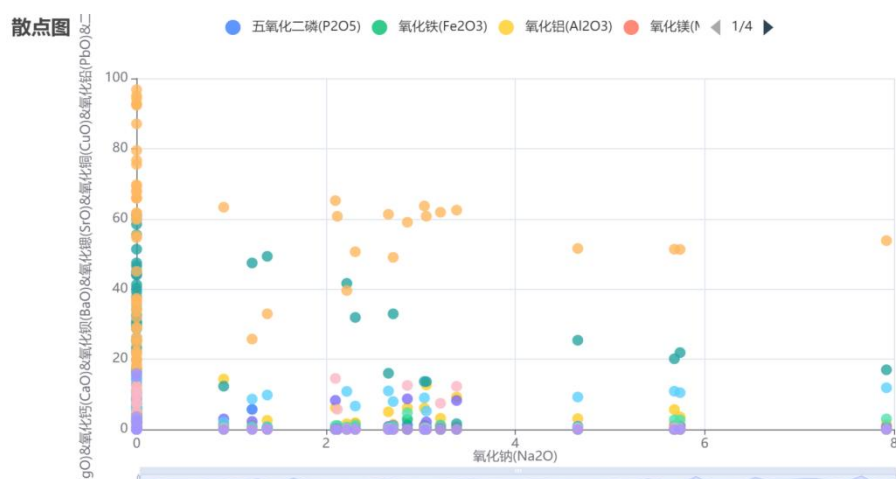


图 4-14 风化前后化学成分含量散点分布图

由散点图可知，同一化学成分分布大致呈现线性关系，将不同类别的文物进行分别计算（参见附录 4、附录 5），由于数据由检测得到，存在一定偏差，在此采用 Lasson 回归模型^[3]预测风化前各化学成分含量，由某一成分预测结果图可知，采用该模型预测拟合度较高：



图 4-14 Lasson 回归模型拟合曲线

铅钡类型：

采用 SPSS 软件计算得到以下方程^[4]：

$$Y(\text{二氧化硅}(\text{SiO}_2)) = 97.131 - 0.68 \times \text{氧化钾}(\text{K}_2\text{O}) - 1.104 \times \text{氧化铝}(\text{Al}_2\text{O}_3) - 1.409 \times \text{氧化钡}(\text{BaO}) - 1.02 \times \text{氧化铅}(\text{PbO}) - 0.97 \times \text{氧化铁}(\text{Fe}_2\text{O}_3) + 0.654 \times \text{氧化锶}(\text{SrO}) - 1.543 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)$$

$$Y(\text{氧化钾}(\text{K}_2\text{O})) = 0.174 + 0.005 \times \text{氧化铝}(\text{Al}_2\text{O}_3) - 0.001 \times \text{氧化铅}(\text{PbO})$$

$$Y(\text{氧化铝}(\text{Al}_2\text{O}_3)) = 33.384 + 1.462 \times \text{氧化钾}(\text{K}_2\text{O}) - 0.294 \times \text{二氧化硅}(\text{SiO}_2) - 0.512 \times \text{氧化钡}(\text{BaO}) - 0.368 \times \text{氧化铅}(\text{PbO}) - 0.267 \times \text{氧化铁}(\text{Fe}_2\text{O}_3) + 1.346 \times \text{氧化锶}(\text{SrO}) - 0.405 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)$$

$$Y(\text{氧化铁}(\text{Fe}_2\text{O}_3)) = 0.699 - 0.006 \times \text{氧化钡}(\text{BaO})$$

$$Y(\text{氧化铁}(\text{Fe}_2\text{O}_3)) = 0.699 - 0.006 \times \text{氧化钡}(\text{BaO})$$

$$Y(\text{氧化锶}(\text{SrO}))=0.536-0.006 \times \text{二氧化硅}(\text{SiO}_2)+0.001 \times \text{氧化铅}(\text{PbO})$$

$$Y(\text{氧化铅}(\text{PbO}))=90.472-0.897 \times \text{二氧化硅}(\text{SiO}_2)-0.094 \times \text{氧化钾}(\text{K}_2\text{O})+2.18 \times \text{氧化锶}(\text{SrO})-0.95 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-1.313 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)-1.216 \times \text{氧化铝}(\text{Al}_2\text{O}_3)-1.319 \times \text{氧化钡}(\text{BaO})$$

$$Y(\text{五氧化二磷}(\text{P}_2\text{O}_5))=45.085-0.47 \times \text{二氧化硅}(\text{SiO}_2)-0.865 \times \text{氧化钾}(\text{K}_2\text{O})+1.09 \times \text{氧化锶}(\text{SrO})-0.248 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-0.455 \times \text{氧化铅}(\text{PbO})-0.464 \times \text{氧化铝}(\text{Al}_2\text{O}_3)-0.649 \times \text{氧化钡}(\text{BaO})$$

$$Y(\text{氧化钡}(\text{BaO}))=64.283-0.639 \times \text{二氧化硅}(\text{SiO}_2)+0.053 \times \text{氧化钾}(\text{K}_2\text{O})+1.281 \times \text{氧化锶}(\text{SrO})-0.965 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)-0.815 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-0.679 \times \text{氧化铅}(\text{PbO})-0.87 \times \text{氧化铝}(\text{Al}_2\text{O}_3)$$

高钾类型:

$$Y(\text{二氧化硅}(\text{SiO}_2))=100.543-2.079 \times \text{氧化镁}(\text{MgO})-0.426 \times \text{氧化铅}(\text{PbO})-0.389 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)-0.363 \times \text{氧化钡}(\text{BaO})-1.087 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-1.024 \times \text{氧化钾}(\text{K}_2\text{O})-0.633 \times \text{氧化钠}(\text{Na}_2\text{O})-1.16 \times \text{氧化铝}(\text{Al}_2\text{O}_3)-1.115 \times \text{氧化钙}(\text{CaO})-1.338 \times \text{氧化铜}(\text{CuO})-1.709 \times \text{氧化锡}(\text{SnO}_2)$$

$$Y(\{\text{氧化钾}(\text{K}_2\text{O})=60.841-0.612 \times \text{二氧化硅}(\text{SiO}_2)-0.208 \times \text{氧化钡}(\text{BaO})-0.734 \times \text{氧化铝}(\text{Al}_2\text{O}_3)-1.064 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-0.113 \times \text{氧化镁}(\text{MgO})-0.251 \times \text{氧化钙}(\text{CaO})-0.573 \times \text{氧化铜}(\text{CuO})$$

$$Y(\text{氧化铝}(\text{Al}_2\text{O}_3))=48.639+0.361 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)-0.476 \times \text{二氧化硅}(\text{SiO}_2)-0.029 \times \text{氧化钡}(\text{BaO})-0.641 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)-0.469 \times \text{氧化钾}(\text{K}_2\text{O})+0.126 \times \text{氧化钠}(\text{Na}_2\text{O})-0.452 \times \text{氧化钙}(\text{CaO})-0.774 \times \text{氧化铜}(\text{CuO})-1.552 \times \text{氧化锡}(\text{SnO}_2)$$

$$Y(\{\text{氧化铁}(\text{Fe}_2\text{O}_3)=12.105-0.126 \times \text{二氧化硅}(\text{SiO}_2)-0.148 \times \text{氧化铝}(\text{Al}_2\text{O}_3)+0.649 \times \text{五氧化二磷}(\text{P}_2\text{O}_5)-0.228 \times \text{氧化钾}(\text{K}_2\text{O})-0.066 \times \text{氧化钠}(\text{Na}_2\text{O})+0.129 \times \text{氧化钙}(\text{CaO})$$

$$Y(\text{氧化锶}(\text{SrO}))=0.079-0.001 \times \text{二氧化硅}(\text{SiO}_2)$$

$$Y(\text{氧化铅}(\text{PbO}))=1.179-0.012 \times \text{二氧化硅}(\text{SiO}_2)$$

$$Y(\text{五氧化二磷}(\text{P}_2\text{O}_5))=5.336-0.054 \times \text{二氧化硅}(\text{SiO}_2)-0.162 \times \text{氧化钡}(\text{BaO})+0.513 \times \text{氧化铁}(\text{Fe}_2\text{O}_3)+0.144 \times \text{氧化铝}(\text{Al}_2\text{O}_3)-0.08 \times \text{氧化铅}(\text{PbO})-0.106 \times \text{氧化镁}(\text{MgO})-0.375 \times \text{氧化钙}(\text{CaO})$$

$$Y(\text{氧化钡}(\text{BaO}))=3.721-0.041 \times \text{二氧化硅}(\text{SiO}_2)+0.756 \times \text{氧化铅}(\text{PbO})-0.04 \times \text{氧化钾}(\text{K}_2\text{O})-0.144 \times \text{氧化钠}(\text{Na}_2\text{O})-0.122 \times \text{氧化钙}(\text{CaO})+0.185 \times \text{氧化铜}(\text{CuO})。$$

以上为化学成分含量的预测线性表达式，根据风化检测数据点的数据带入，即可求得风化前化学成分含量。

4.2 对问题二的求解:

4.2.1 分析高钾玻璃、铅钡玻璃分类规律:

由于古代玻璃制品埋藏时与环境发生风化，因此无法准确说明某个成分含量是否为该文物中原本的主要成分，故分析化学成分分类规律时应进行随机重估取样，直到将每个成分都计算分析，进而确定某个化学成分在不同类别时含量占比，从而得出分类规律，故此处采用随机森林分类算法进行分析^[5]：

以类型为变量 Y，主要化学成分为变量 X，使用 SPSS 软件导入附件数据进行分析得出成分含量占比：

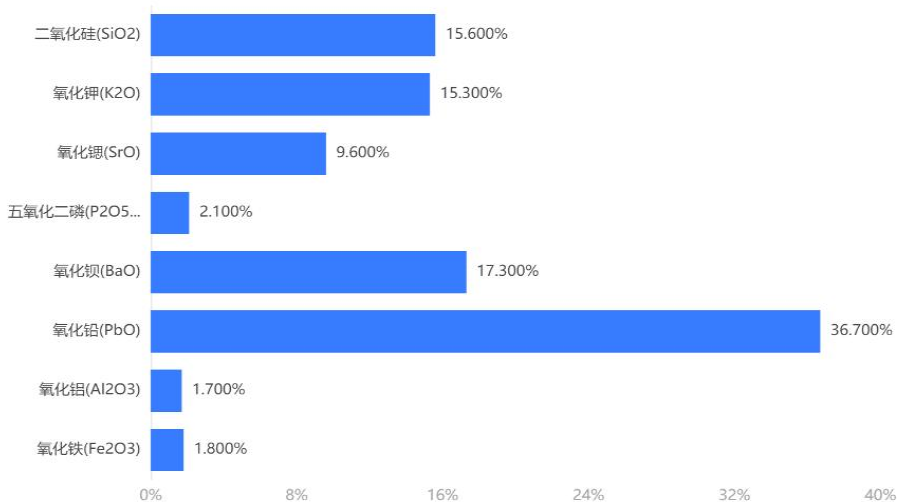


图 4-15 化学成分含量占比柱形图

根据以上不同化学成分含量占比，可以得出高钾玻璃与铅钡玻璃主要以氧化铅含量进行分类，当氧化铅含量占比高达 36.75 及以上时为铅钡玻璃，否则为高钾玻璃。

4.2.2 对每个类别进行亚分类，并分析结果的合理性与敏感性：

由于化学成分要进行的分类对象较多，对多种成分进行降维处理，将化学成分聚集为两类或多类。

高钾类型：
将高钾类型的数据表单独导出，导入 SPSS 软件进行聚类分析，将类型分为两类：

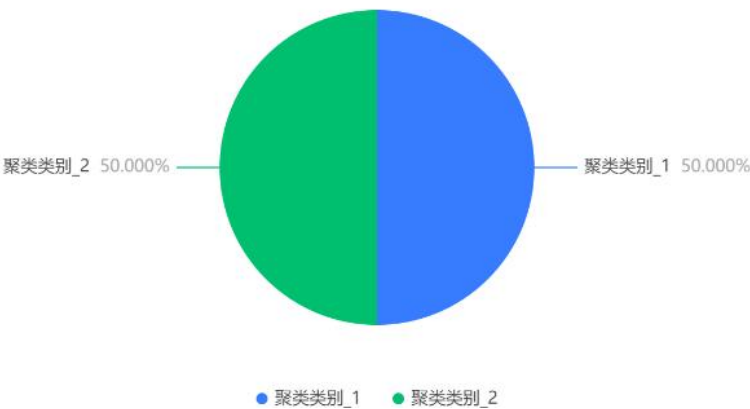


图 4-16

以不同化学成分含量所处期间作为划分类别的依据，具体分类细节参见附附录 6。根据以下表格数据将高钾类型不同化学成分含量分为两类：

表 2 高钾类型不同化学成分含量及其分类

类型	类别 1 (n=9)	类别 2 (n=9)
二氧化硅 (SiO ₂)	63.624±3.558	89.663±7.119
氧化铝 (Al ₂ O ₃)	7.349±2.346	2.764±1.67
氧化钾 (K ₂ O)	10.818±2.37	1.986±3.22
氧化铁 (Fe ₂ O ₃)	2.312±1.643	0.44±0.735
氧化铅 (PbO)	0.41±0.64	0.139±0.333
五氧化二磷 (P ₂ O ₅)	1.523±1.652	0.533±0.451
氧化钡 (BaO)	0.579±1.001	0.219±0.657
氧化锶 (SrO)	0.048±0.051	0.008±0.023

有表 2 分析可知，氧化铅(PbO)、五氧化二磷(P₂O₅)、氧化钡(BaO)、氧化锶(SrO)在两个类别类别中存在数值区间重叠的情况，导致分类此化学成分的敏感性较大，故分类时不将这些化学成分作为优先考虑对象。

铅钡类型：

将铅钡类型的数据表单独导出，导入 SPSS 软件进行聚类分析，将类型分为两类

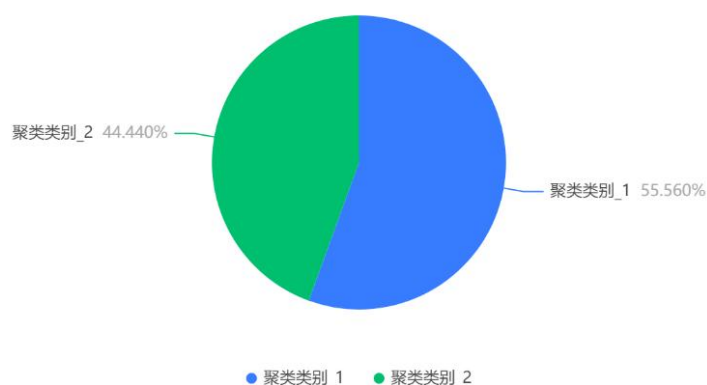


图 4-17

两类占比如上图所示。以不同化学成分含量 x 所处期间作为划分类别的依据，具体分类细节参见附录 7。根据以下表格数据将铅钡类型不同化学成分含量分为两类：

表 3 铅钡类型不同化学成分含量及其分类

类型	类别 1 (n=9)	类别 2 (n=9)
二氧化硅 (SiO ₂)	24.835±9.503	57.698±9.319
氧化铝 (Al ₂ O ₃)	2.633±1.502	4.635±3.441
氧化钾 (K ₂ O)	0.169±0.351	0.181±0.173
氧化铁 (Fe ₂ O ₃)	0.666±0.907	0.604±1.089
氧化铅 (PbO)	42.567±10.151	20.902±6.557

五氧化二磷 (P_2O_5)	4.723 ± 4.309	1.129 ± 1.958
氧化钡 (BaO)	13.074 ± 10.32	8.008 ± 4.738
氧化锶 (SrO)	0.45 ± 0.274	0.221 ± 0.205

由表 3 分析可得, 氧化钾(K_2O)、氧化锶(SrO)、氧化铁(Fe_2O_3)、五氧化二磷(P_2O_5)在两类分区中数据存在较大区域重叠, 导致分类此化学成分的敏感性较大, 故分类时不将这些化学成分作为优先考虑对象。

4.3 对问题三的求解

根据问题二中对不同类别的化学成分的亚分类, 比较附件表单 3 中的有效化学成分, 可得出其所属类型, 由于一份数据有多个化学成分, 故按标题 5.2.1 中含量占比较高的化学成分作为优先分类数据。

比较数据与分类依据可得:

表 4 鉴别类型

文物编号	类型
A1	高钾
A2	铅钡
A3	铅钡
A4	铅钡
A5	高钾
A6	高钾
A7	高钾
A8	铅钡

A1 到 A8 编号的文物类型分别为高钾、铅钡、铅钡、铅钡、高钾、高钾、高钾、铅钡。根据附录 8 中的分类数据, 以及 4.2.1 的图中化学成分含量占比高者依次由高到低进行比较, 从而鉴别其所属类型, 即由于高钾和铅钡类型的文物中, 氧化铅(PbO)占比最高, 但由于两种类型的文物中氧化铅(PbO)的含量区间有重叠部分, 故排除, 依次分析文物化学元素含量, 当分析至二氧化硅(SiO_2)时, 若其含量在 63.624 ± 3.558 范围内的, 则该文物类型为高钾, 若其含量在 89.663 ± 7.119 范围内的, 则为铅钡。

分析附录 8 中两种类型文物所含各化学成分的含量可以发现, 氧化镁(MgO)、氧化铁(Fe_2O_3)、氧化铜(CuO)、氧化锡(SnO_2)、二氧化硫(SO_2)、氧化钡(BaO)、氧化锶(SrO)这几种化学成分含量区间是存在较大重叠部分的, 因此在根据此方法鉴定文物类型时以上几种化学成分存在敏感性。

4.4 对问题四的求解

4.4.1 对铅钡玻璃和高钾玻璃的化学成分的关联性的分析

对铅钡玻璃使用相关性分析对仅为铅钡玻璃的化学成分分析,采用 person 相关系数分析, 得到每一份化学成分的相关系数热力图:

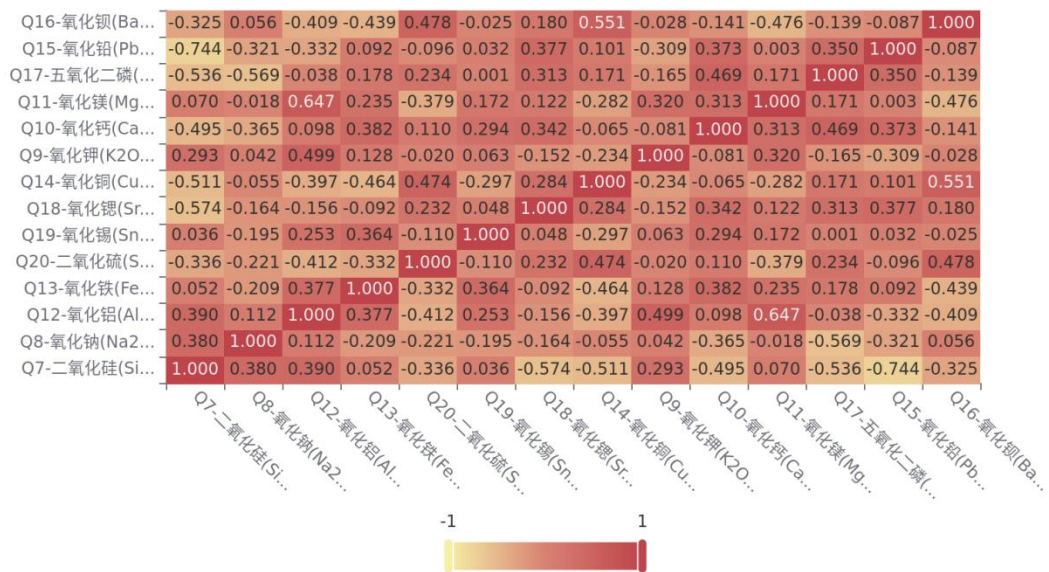


图 4-18 铅钡玻璃化学成分的相关系数热力图

从图中可以看出:氧化钡与氧化铜和二氧化硫存在明显的正相关, 相关系数分别为 0.739 与 0.632, 而氧化铅与二氧化硅存在明显的负相关, 相关系数分别为 0.728, 其他元素之前的相关性不强。

对高钾玻璃同样使用采用 person 相关系数分析, 得到每一份化学成分的相关系数热力图:

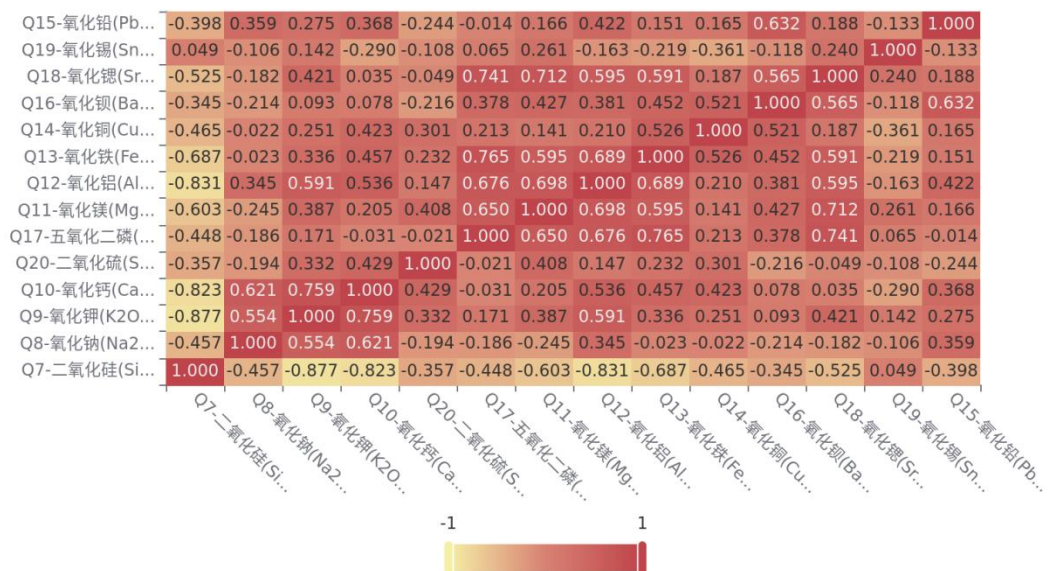
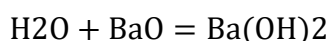


图 4-19 高钾玻璃化学成分的相关系数热力图

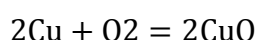
从图中可以看出:氧化锶与五氧化二磷、氧化镁存在明显的正相关关系,相关系数分别为:0.741 和 0.712,氧化铁与五氧化二磷、氧化铝也存在明显的正相关关系,相关系数为:0.765 和 0.689,而二氧化硅与氧化铝、氧化钙和氧化钾存在明显的负相关关系,相关系数分别为:-0.831、-0.823 和-0.877。

4.4.2 对不同类型的玻璃的化学成分的差异性讨论

从 4.1.1 中可以看出,两种类别的化学成分的关联性差异明显,考虑到两种类型玻璃的差异根源来自其组成成分的不同,铅钡玻璃的铅元素含量较高,高钾玻璃的钾元素含量高,结合 4.1.2 的风化前后的化学成分含量变化可知,由于两种玻璃的类型不同,查找相应文献后^[6],得知其风化的化学反应不同,对于铅钡玻璃,其主要发生的化学反应为钡离子的反应:



铜离子的反应:



而高钾玻璃则较多含有五氧化二磷,风化后与其他金属离子的含量成正相关。所以存在差异性。

5、模型的优缺点

5.1 模型的优点

- 1.在对相关性检验中,模型考虑了实际的风化的情况,以实际入手,保证了与实际的贴近。
- 2.模型的建立经过比对,让得到的结果更加有效。
- 3.模型以较基础简单的模型为基础,让模型更加符合波动较大的数据。

5.2 模型的缺点

- 1.由于采用比较基础的模型进行建立,对于数据的预测,无法给出非常精确的数据结果。
- 2.采用的数据仅限于所给的数据,对于更多的数据要求可能存在误差。

6、参考文献

- [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from <https://www.spsspro.com>.
- [2]徐维超. 相关系数研究综述[J]. 广东工业大学学报, 2012, 29(3):12-17.
- [3]柯郑林. Lasso 及其相关方法在多元线性回归模型中的应用[D]. 北京交通大学, 2011.
- [4] Fisher Box, Joan. Guinness, Gosset, Fisher, and Small Samples. Statistical Science. 1987, 2 (1): 45-52.
- [5]周志华. 机器学习[M]. 清华大学出版社, 2016.
- [6]周良知. 影响硅酸盐玻璃风化的主要因素[J]. 大连轻工业学院学报, 1984(01):34-44.

7、附录清单

- 附录 1: 相关成分化学分析:相关化学成分附件.xlsx
- 附录 2: 类别 T 检验:类别 T 检验附件.xlsx
- 附录 3: 风化 T 检验:风化.xlsx ,风化 T 检验附件.xlsx
- 附录 4: 高钾聚类.csv
- 附录 5: 铅钡聚类.csv
- 附录 6: 高钾聚类分析
- 附录 7: 铅钡聚类分析
- 附录 8: 类型聚类分析.xlsx
- 附录 9: 相关性分析.py ,相关性分析.xlsx
- 附录 10: 高钾_聚类分析.py,全是高钾.xlsx
- 附录 11: 铅钡_聚类分析.py ,全是铅钡.xlsx
- 附录 12: 随机森林.py, 随机森林分析.xlsx