

基于 K-均值聚类的古代玻璃制品化学成分分析模型

摘要

本文主要研究了古代玻璃制品的成分分析与鉴别。利用了决策树模型、主成分分析、K-均值聚类等方法，对玻璃文物表面风化与其基本信息和化学成分进行了分析，对风化点风化前化学成分含量进行预测，得到不同类型玻璃的分类规律并完成亚类划分以及未知类别玻璃文物类型鉴别，分析化学成分间以及不同类别间化学成分的相关关系，并对所得结果进行合理性与敏感性分析。

针对问题所给数据，我们首先进行了数据预处理。将化学成分总含量不在合理范围内及对分析结果影响程度弱的部分化学成分的数据进行剔除。对此，我们剔除了表单2中15、17号不合理样本数据，及表单2中氧化钠、氧化镁、氧化锶、氧化锡、二氧化硫五组化学含量对应数据。

针对问题一，首先借助Excel进行数据分析，根据决策树分类原理及折线图直接观察法，利用决策树模型求得玻璃文物表面风化与其基本信息的关系，根据折线图绘制得出高钾玻璃、铅钡玻璃有无风化及化学成分含量之间的统计规律，并基于线性原理预测了风化点风化前的化学成分含量（结果见附录）。

针对问题二，基于主成分分析法和K-均值聚类的原理，借助SPSS及Matlab编程，通过数据比对，我们经筛选采用先用主成分分析法降维，再进行K-均值聚类，得到高钾玻璃、铅钡玻璃的分类规律。亚类划分则直接对数据进行K-均值聚类的方法，给出了依据是否风化的划分方法并给出了分类结果（结果见正文6.3），并对划分方法给出了合理性较高、敏感性较低的分析结果。

针对问题三，先对表单2、3进行数据统一处理，分为风化与未风化两类，利用K-均值聚类法，对数据进行聚类，得出“A1为未风化高钾玻璃，A5、A6、A7为风化高钾玻璃，A3、A4、A8为未风化铅钡玻璃，A2为风化铅钡玻璃”的结果，并通过对数据进行扰动的方式分析其敏感性，得到其敏感性较低的结论。

针对问题四，通过相关系数矩阵分析的方式，借助Excel数据处理及SPSS编程，对高钾玻璃、铅钡玻璃各自化学成分关联关系进行分析，并通过差值矩阵的方式得出不同类别的化学成分关联关系的差异性（相关系数矩阵见附录）。

关键词：古文物研究 化学成分 决策树 K-均值聚类 主成分分析法

一、问题重述

玻璃贸易是古代丝绸之路中西方贸易的重要部分，而我国古代在吸收外来的玻璃制品技术后就地取材制作出玻璃，与外来玻璃成分不大相同。玻璃的主要化学成分是 SiO_2 ，往里面加入助熔剂与稳定剂，使得玻璃的化学成分组成多有不同。例如，铅钡玻璃的氧化铅、氧化钡的含量较高；钾玻璃的含钾量高。古代玻璃在长期埋藏的过程中容易受环境影响而风化，导致各化学成分含量发生变化，从而影响对其类型的判别。

本问题通过对至今留存的玻璃文物中化学成分进行分析，查资料得到玻璃文物的基本信息，并检测出玻璃文物的化学成分比例，针对下列问题得出相关结果，并最终得到相关关系。

附件给出了 3 份表单。表单 1 给出了 58 个玻璃文物的基本信息，表单 2 给出了已分类的文物表面共 69 个取点的化学物质成分含量表，表单 3 给出了未分类的玻璃文物共 8 个取点的化学物质成分含量表。现需完成如下问题：

1. 结合表单 1 数据，建立分析模型，分析与描述玻璃文物的表面风化状况及其玻璃类型、纹饰和颜色之间的关系；结合表单 2 数据，进行数据的分析处理，分析文物样品表面化学成分含量的统计规律，并且结合统计规律，根据现有检测数据，预测其风化前的成分含量。
2. 结合表单 2 数据，分析得出“高钾玻璃”“铅钡玻璃”的分类规律；对以上分类进行亚类划分，给出具体的划分方法及划分结果，并根据细分结果分析合理性和敏感性。
3. 基于表单 3 中未知文物化学成分的数据进行鉴别分析，鉴别其类型，并对分类结果进行敏感性分析。
4. 结合表单 2 数据分析不同类别的玻璃样品的化学成分之间的关联关系，并比较不同类别间化学成分关联关系的差异。

二、问题分析

古代玻璃制品的化学成分分析问题给出了不同文物基本信息及其表面取点

情况，题目要求据此完成不同文物的化学成分分析、分类与统计规律归纳。

2.1 问题一分析

问题一首先要求分析表面风化情况与玻璃类型、纹饰和颜色三个参数之间的关系。问题应当先处理数据，剔除个别未标明颜色的无效数值，然后通过建立决策树模型，形象直观地描述出表面风化情况与三者的关系并得出相关规律。其次通过统计图表、散点图、折线图观察对比得到有无风化与化学成分含量的统计规律，接着对每一种化学物质成分含量未风化和风化分别进行求平均值，根据平均值之间做差得出该化学物质成分含量在风化前后的平均变化量，将风化后数值加上变化数值即为预测出的风化前对应数值。

2.2 问题二分析

问题二要求找出高钾玻璃和铅钡玻璃的分类规律及对每个类别进行亚类划分，最后进行结果分析。首先处理数据，剔除无效数据并排序已知有效数据。接着将表单 2 数据先进行主成分分析降维处理，再进行 K-均值聚类，并将聚类结果与直接聚类所得结果进行对比，选择结果较为理想的一种聚类方式，并通过较为理想的聚类方式的途径得到高钾玻璃与铅钡玻璃的分类规律；也可通过直接聚类所得的两个聚类中心的化学成分含量数据进行对比，得出概括性的分类规律。对每个类别进行亚类划分，采用 K-均值聚类方法对每个玻璃类别聚成两类，并将分类结果与真实信息进行对比，通过计算正确率以及灵敏度来对结果的合理性和敏感性进行分析。

2.3 问题三分析

问题三要求对未知玻璃文物的化学成分进行分析鉴别，可依据 K-均值聚类方式解决需求，首先将数据分成风化和未风化两类，降低聚类难度，接着对每类数据进行 K-均值聚类，设置两个聚类中心，因为高钾玻璃与铅钡玻璃的差异性较为明显，故所得聚类应为高钾玻璃类和铅钡玻璃类，如此即可分析出未知玻璃文物的类别，最后在通过对数据进行扰动观察结果变化的方式来进行敏感性分析。

2.4 问题四分析

问题四要求分析不同类别玻璃化学成分之间的关联关系，可通过求解不同化学成分之间的相关系数矩阵来得出它们之间的关联关系，并通过矩阵做差来得出

不同类别的化学成分相关系数差值的矩阵，最后根据差值矩阵判断不同类别的化学成分关联关系的差异性。

三、基本假设与符号说明

3.1 基本假设

- (1) 附件中给出的数据均为真实测量数据，并且适用于本文章中的模型；
- (2) 样本量充分具有随机性与代表性，以至于能根据所给数据得出研究对象的一般规律；
- (3) 对部分干扰数据的合理剔除不会对最后的问题结果产生影响；
- (4) 参数分组的组内相应参数一致，忽略一定范围内的参数误差。

3.2 符号说明

符号	含义
x_1	古代玻璃样品纹饰量化值
x_2	古代玻璃样品类型量化值
x_3	古代玻璃样品颜色量化值
y_i	量化值 x_i 归一化后的规范值
p_k	当前样本集合中第 k 类样本所占的比例
$\text{Ent}(D)$	度量决策树结点纯度的指标信息熵
r_{ij}	数据标准化后指标 i 和指标 j 的相关系数
R	标准化后数据的相关系数矩阵
\bar{x}_j	j 个样本数据的平均值
$\text{Var}(x_j)$	j 个样本数据的标准差
β_j	贡献率
α_p	累计贡献率
Z	主成分根据贡献率加权的综合得分
C_i	K-均值聚类中心
S	K-均值聚类的样本数据集合

四、数据预处理

本题表单 1 给出了 58 个玻璃文物的基本信息，包括玻璃制品的编号、纹饰、类型、颜色、表面风化五个参数。在此先对数据进行预处理，以防止其中存在错误数据对计算结果产生影响。容易发现，表单 1 中存在信息缺失的数据，这些数据对模型建立无效，故需要剔除。

表格 1：对表单 1 的数据处理表

文物编号	纹饰	类型	颜色	表面风化
01	C	高钾	蓝绿	无风化
02	A	铅钡	浅蓝	风化
03	A	高钾	蓝绿	无风化
04	A	高钾	蓝绿	无风化
05	A	高钾	蓝绿	无风化
06	A	高钾	蓝绿	无风化
.....

经过数据分析，表单 1 中编号 19、40、48、58 无颜色标记，无法进行后期处理，属无效数据，故直接剔除。

表单 2 提供了 69 个玻璃文物化学成分含量的数据，这些数据的特点是成分性，即各成分比例的累加和应为 100%，但表单 2 中数据存在累加和不为 100% 的情况，本题中将成分比例累加和介于 85%~105% 之间的数据视为有效数据，并将不在 85%~105% 之间的数据视为无效数据并进行剔除。

表格 2：对表单 2 的数据处理表

文物采样点	二氧化硅	氧化钠	氧化钾	氧化钙
01	69.33		9.99	6.32
02	36.28		1.05	2.34
03 部位 1	87.05		5.19	2.01
03 部位 2	61.71		12.37	5.87
04	65.88		9.67	7.12

05	61.58		10.95	7.35
.....

经过数据分析，15 和 17 号数据为上文所述无效数据，对 15 及 17 号数据进行剔除。至此，通过无效数据的剔除，已完成数据的初步处理。接下来对各问题进行分析建模与求解。

五、问题一模型建立与求解

5.1 建立决策树模型分析玻璃纹饰、类型、颜色与风化情况的关系

5.1.1 表单 1 数据处理

表单 1 中共有 58 组数据，其中有 4 组未标明颜色的数据，因未标明颜色数据过少而对决策树建立的影响较小，为简化问题故将这四组未标明颜色的数据剔除。余下 54 组数据，随机取出 44 组为训练集，用以找出数据间的相关关系；余下的 10 组为测试集，用以鉴别关系式的正确性与最后结果的导出。

为了数据处理的方便，我们将纹饰、类型、颜色及风化情况等特征量化为数值，并使用归一化使数据更加合理。其中，将 A、B、C 共三种纹饰，分别替换为数值 0、1、2；将铅钡、高钾玻璃两种类型，分别替换为数值 0、1；将浅绿、浅蓝、蓝绿、绿、深绿、深蓝、紫、黑八种颜色，分别表示为 0，1，2，3，4，5，6，7；将未风化、风化情况，分别表示为-1，1。

下面对每组编号进行归一化处理，使结果更加准确合理。

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

表格 3：玻璃制品基本信息归一化处理

参数	原数据	归一化处理后编码
纹饰	0、1、2	0、0.5、1
类型	0、1	0、1
颜色	0、1、2、3、4、5、 6、7	0、0.1428、0.2857、0.4285、 0.5714、0.7142、0.8571、1

5.1.2 决策树模型原理

解决此问题，我们选取决策树中的分类树模型。决策树分为分类树与回归树。决策树原理是机器学习中一种基本的分类和回归算法，本质是从训练数据集中归纳出一组分类规则，可以依据判决规则来预测未知样本。

一棵决策树包含一个根结点、若干个内部结点和若干个叶结点；叶结点对应于决策结果，其他每个结点则对应于一个属性测试；每个结点包含的样本集合根据属性测试的结果被划分到子节点中；根结点包含样本全集。从根结点到每个叶结点的路径对应了一个判定测试序列，最后一层“叶子”作为分类标签。

5.1.3 分类树模型建立应用步骤

首先依据信息熵选择最优划分属性，属性总共有三种，分别为类型、纹饰及颜色。依据以下公式，求得各属性对应的信息熵。

$$\text{Ent}(D) = - \sum_{k=1}^{|y|} p_k \log_2 p_k \quad (2)$$

表格 4：属性点的信息熵及其增益

	x_1	x_2	x_3
Ent(D)	1.3897	0.8905	2.4299

结果为：类型信息熵最小，纯度最高，其次为纹饰，最后是颜色。

依据划分属性的优先度确定根结点、内部结点以及叶结点，根据结点不断划分数据，逐渐建立决策树模型。再通过剪枝策略，自底而上地对非叶结点进行考查，若将该结点对应的指数替换为叶结点能使决策树泛化性能提升，则将该指数替换为叶结点，最终得到剪枝后的决策树模型。

5.1.4 使用决策树求解问题一的结果与分析

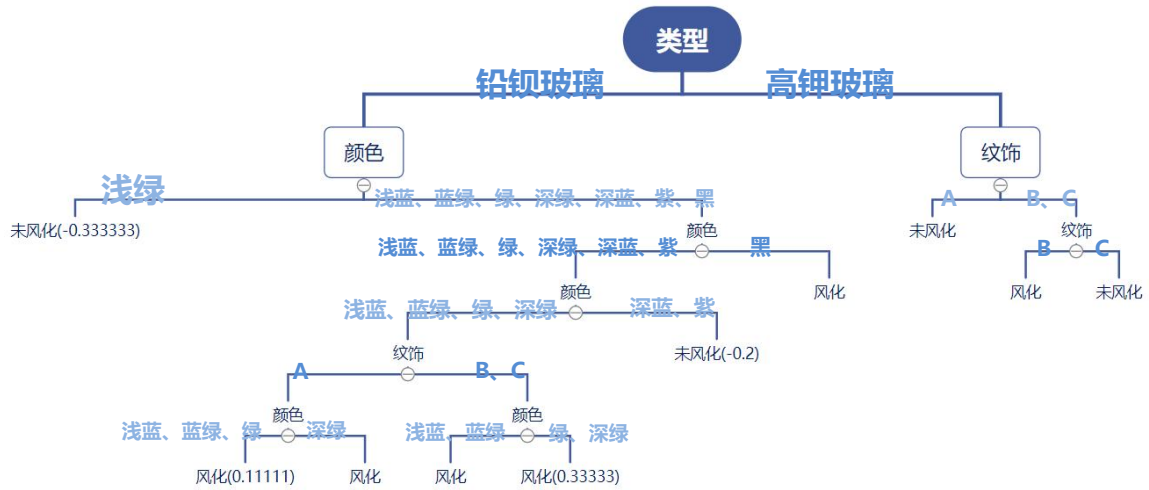


图 1：决策树求解结果

表格 5：决策树文字解析

类别	类型	纹饰	颜色	风化情况
1	铅钡	A、B、C	浅绿	未风化(0.33333)
2	铅钡	A	浅蓝、蓝绿、绿色	风化(0.11111)
3	铅钡	A	深绿	风化
4	铅钡	B、C	浅蓝、蓝绿	风化
5	铅钡	B、C	绿色、深绿	风化(0.33333)
6	铅钡	A、B、C	深蓝、紫色	未风化(0.2)
7	铅钡	A、B、C	黑色	风化
8	高钾	A	所有颜色	未风化
9	高钾	B	所有颜色	风化
10	高钾	C	所有颜色	未风化

其中，风化情况一栏中，“风化”“未风化”表示对应类别的玻璃制品更趋于的风化情况。若其后加括号数值 x ，表示对应概率 $P = \frac{x+1}{2}$ 。如“未风化(0.33333)”，表示此类趋于未风化，且未风化概率 $P = \frac{0.33333+1}{2} = 0.666665$ 。若其后未加括号，表示所给数据范围内对应风化情况一定成立。

5.2 通过线性回归与拟合分析风化与化学成分统计规律

5.2.1 数据处理

表单 2 提供了玻璃制品表面取点化学成分含量数据，其中有些化学成分数据对解决问题作用不大，我们选择直接剔除相关数据，以便统计规律能更直观地反映问题的本质。其中，氧化钠、氧化镁、氧化锡、二氧化硫数值过少，氧化锆随不同取点变化不大，均不利于问题分析，故直接剔除以上五种化学成分。接下来进行模型建立。

5.2.2 题目分析与模型建立

为得到风化与化学成分含量的统计关系，也即其内在联系，可通过找它们之间的函数关系，故可以先将不同化学成分含量与对应的文物样本做出对应的散点图，得到直观的印象，再根据散点图确定所要使用的预测函数，再将函数拟合，即得到统计规律，也可通过对同种化学成分未风化与风化后的数据做折线图，经对比得到统计规律。其次对每一种化学物质成分含量未风化和风化分别进行求平均值，根据平均值之间做差得出该化学物质成分含量在风化前后的平均变化量，将风化后数值加上变化数值即为预测出的风化前对应数值。

(1) 用散点图及拟合线描述与求解不同化学成分含量的关系

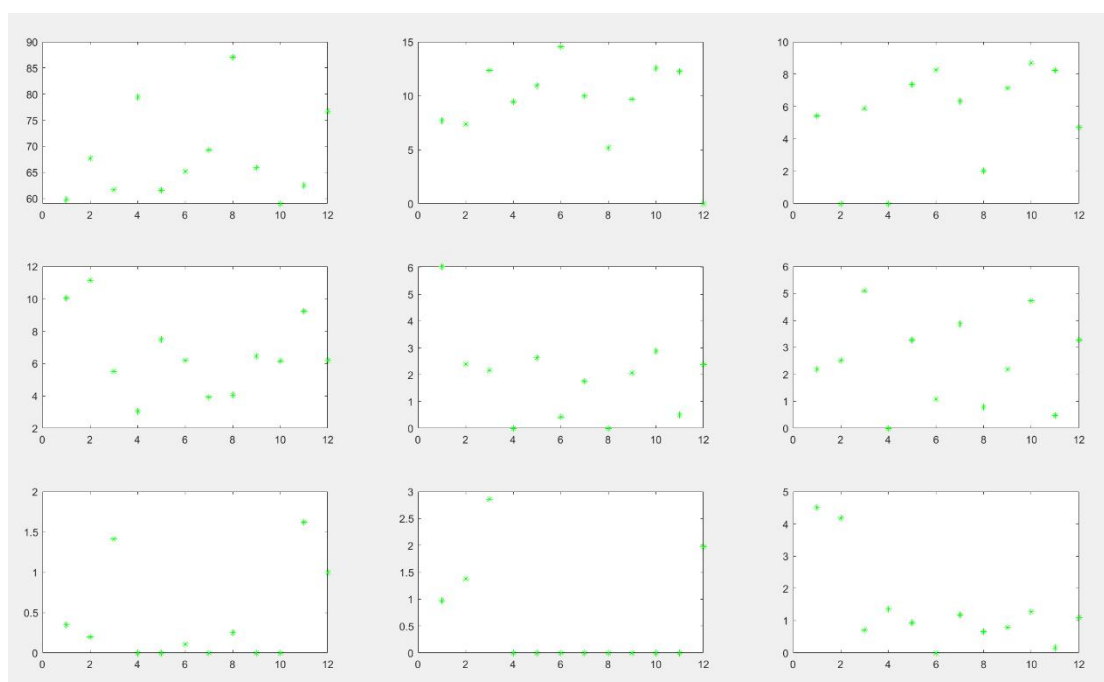


图 2：高钾玻璃各化学成分散点图

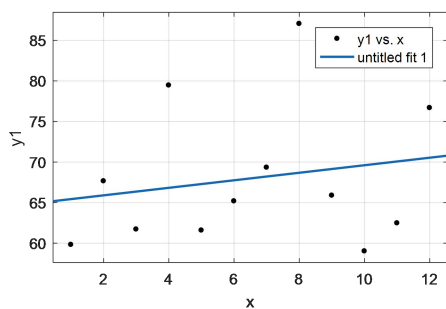


图 3：高钾玻璃二氧化硅散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.464 (-1.215, 2.143)$$

$$k_2 = 64.97 (52.61, 77.33)$$

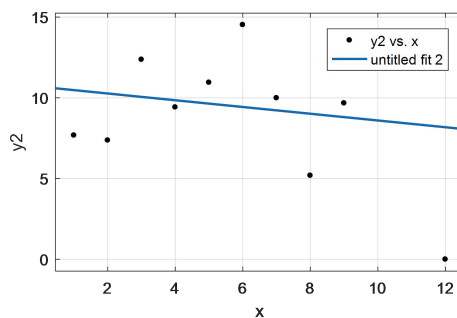


图 4：高钾玻璃氧化钾散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.2088 (-0.9607, 0.543)$$

$$k_2 = 10.69 (5.155, 16.22)$$

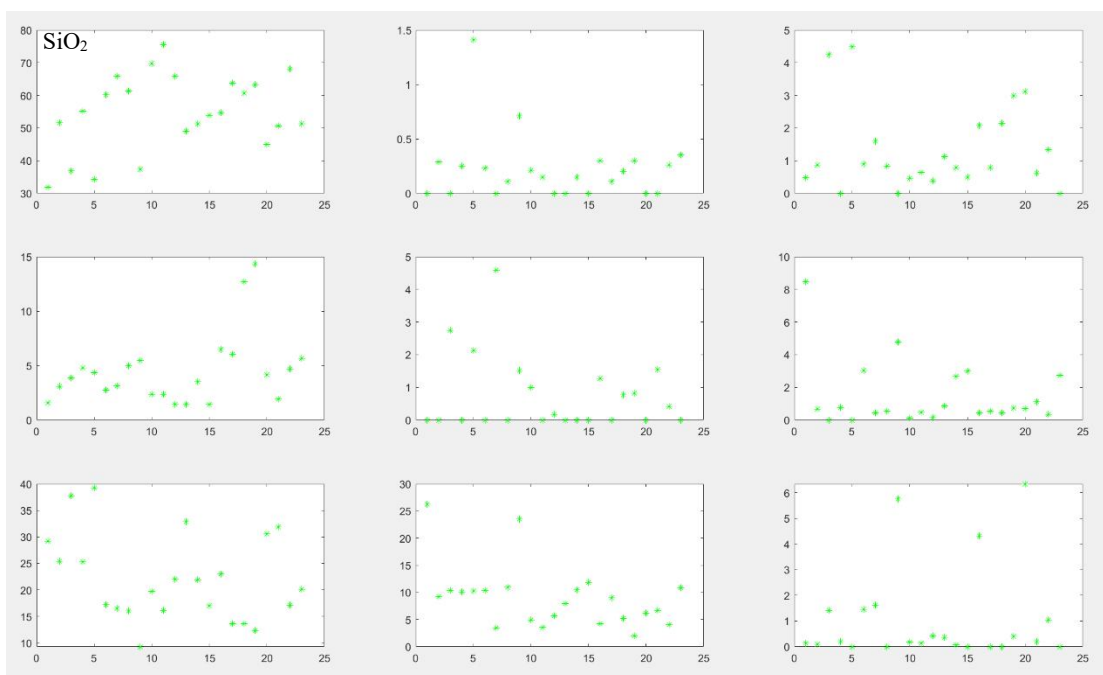


图 5：铅钡玻璃各化学成分散点图

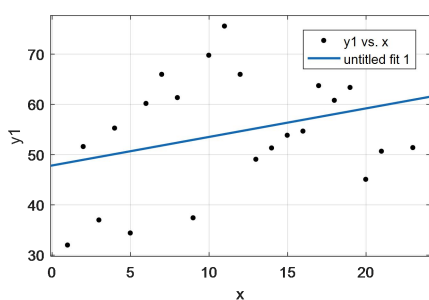


图 6：铅钡玻璃二氧化硅散点及拟合线

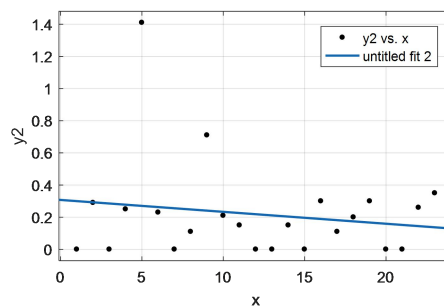


图 7：铅钡玻璃氧化钾散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.5679 (-0.1804, 1.316)$$

$$k_2 = 47.85 (37.58, 58.11)$$

$$f(x) = k_1x + k_2$$

$$k_1 = -0.007401 (-$$

$$-0.02787, 0.01307)$$

$$k_2 = 0.3075 (0.0268, 0.5882)$$

(2) 借助折线图判断统计规律

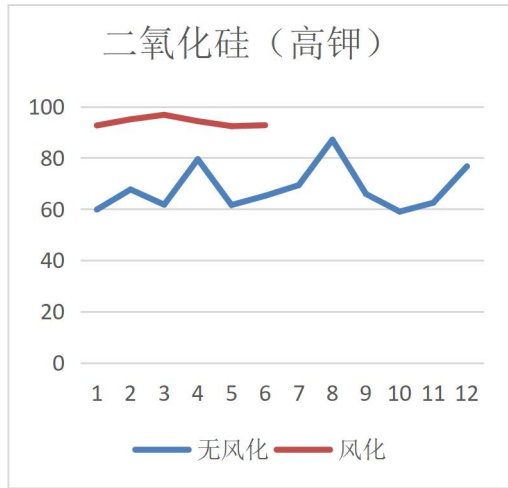


图 8：高钾玻璃二氧化硅含量折线图

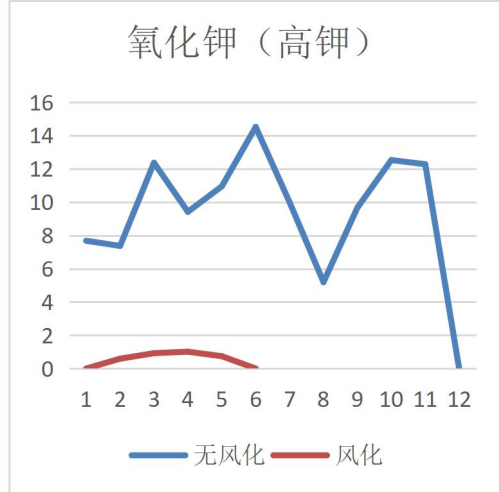


图 9：高钾玻璃氧化钾含量折线图

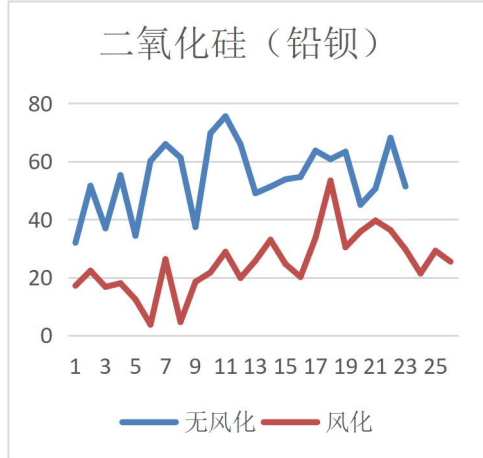


图 10：铅钡玻璃二氧化硅含量折线图

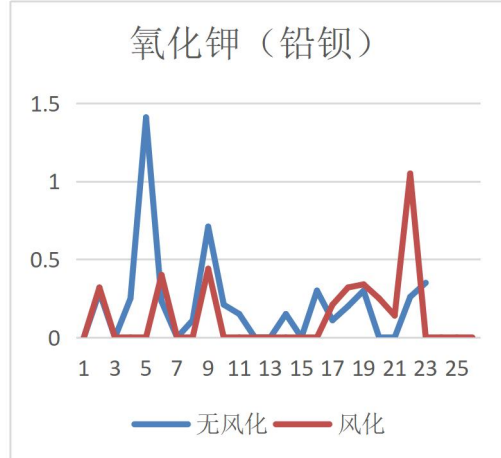


图 11：铅钡玻璃氧化钾含量折线图

注：所有化学成分散点图及折线图见附录图 13~47。

(3) 数值代入计算

利用函数式：

$$w'_\alpha = (\overline{w}_\beta - \overline{w}_\gamma) + w_\alpha \quad (3)$$

其中， w_α 表示风化后玻璃某种化学成分含量的测量值， w'_α 表示风化前

玻璃该化学成分含量的预测值， \overline{w}_β 表示无风化玻璃该化学成分含量的平均值， \overline{w}_γ 表示风化后玻璃该化学成分含量的平均值。

下表为计算出各化学成分含量的平均值：

表格 6：高钾玻璃各化学成分含量部分平均值（一）

高钾 玻璃	二氧化硅 (SiO_2)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)
风化平均值	67.98	9.33	5.33	6.62	1.93
无风化平均值	93.96	0.54	0.87	1.93	0.27
差值	-25.98	8.79	4.46	4.69	1.67

表格 7：高钾玻璃各化学成分含量部分平均值（二）

高钾 玻璃	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (Al_2O_3)
风化平均值	2.45	0.41	0.60	1.40
无风化平均值	1.56	0	0	0.28
差值	0.89	0.41	0.60	1.12

表格 8：铅钡玻璃各化学成分含量部分平均值（一）

铅钡 玻璃	二氧化硅 (SiO_2)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)
风化平均值	54.66	0.22	1.32	4.46	0.74
无风化平均值	24.91	0.13	2.70	2.97	0.58
差值	29.75	0.09	-1.37	1.49	0.15

表格 9：铅钡玻璃各化学成分含量部分平均值（二）

	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (Al_2O_3)
风化平均值	1.43	22.08	9.00	1.05
无风化平均值	2.28	43.31	11.81	5.28
差值	-0.84	-21.23	-2.81	-4.23

据此预测风化高钾与风化铅钡的化学成分含量部分结果如下表（完整结

果见附录)：

表格 10：高钾玻璃风化样本点风化前部分化学成分预测值（一）

预测值	二氧化硅	氧化钾	氧化钙	氧化铝
07	66.65	8.79	5.53	6.67
09	69.04	9.38	5.08	6.01
.....
27	66.74	8.79	5.40	7.20

表格 11：铅钡玻璃风化样本点风化前部分化学成分预测值（一）

预测值	二氧化硅	氧化钾	氧化钙	氧化铝
54 严重	46.86	0.09	0	5.14
54	52.03	0.41	1.82	5.64
40	46.46	0.09	0.5	1.94
.....
56	58.9	0.09	0	3.34
57	55.17	0.09	0	3.67

注：其中负值均修正为零值。完整表单见附录表格 33~36。

5.3 问题一结论

（1）玻璃文物的表面风化与其玻璃类型、纹饰和颜色有如下关系：

- ①对铅钡玻璃，颜色为浅绿色，有 66.67%的概率未风化；
- ②对铅钡玻璃，纹饰为 A 型，颜色为浅绿、蓝绿及绿色的，有 55.56%的概率风化；
- ③对铅钡玻璃，纹饰为 A 型，颜色为深绿色的，一定风化；
- ④对铅钡玻璃，纹饰为 B、C 型，颜色为浅蓝、蓝绿色的，一定风化；
- ⑤对铅钡玻璃，纹饰为 B、C 型，颜色为深绿、绿色的，有 55.56%的概率风化；
- ⑥对铅钡玻璃，颜色为深蓝、紫色的，有 60%的概率未风化；
- ⑦对铅钡玻璃，颜色为黑色的，一定风化；
- ⑧对高钾玻璃，纹饰为 A 型的，一定未风化；

⑨对高钾玻璃，纹饰为 B 型的，一定风化；

⑩对高钾玻璃，纹饰为 C 型的，一定未风化。

(2) 参考如上高钾玻璃与铅钡玻璃各化学成分含量折线图，文物样品表面有无风化与化学成分含量的统计规律如下：

对于高钾玻璃，

①风化的高钾玻璃二氧化硅含量比无风化的显著提高，变化更小。

风化的高钾玻璃二氧化硅含量波动在 95% 附近，而无风化的二氧化硅含量在 70% 左右波动；

②风化的高钾玻璃氧化钾含量比无风化的含量更低，变化更小。风化的高钾玻璃氧化钾含量在 0.5% 附近波动，无风化的氧化钾含量在 8% 附近波动。

③风化的高钾玻璃氧化钙含量比无风化的含量更低，变化更小。风化的高钾玻璃氧化钙含量在 1% 附近波动，无风化的氧化钾含量在 5% 附近波动。

④风化的高钾玻璃氧化铝含量比无风化的含量更低，变化更小。风化的高钾玻璃氧化铝含量在 2% 附近波动，无风化的氧化铝含量在 6% 附近波动。

⑤风化的高钾玻璃氧化铁含量比无风化的含量更低，变化更小。风化的高钾玻璃氧化铁含量在 0.5% 附近波动，无风化的氧化铁含量在 1.5% 附近波动。

⑥风化的高钾玻璃氧化铜含量比无风化的含量更低，变化更小。风化的高钾玻璃氧化铜含量在 1.5% 附近波动，无风化的氧化铜含量在 2.5% 附近波动。

对于铅钡玻璃，

①风化的铅钡玻璃二氧化硅含量比无风化的含量更低。风化的铅钡玻璃二氧化硅含量在 20% 附近波动，无风化的二氧化硅含量在 50% 附近上下波动。

②风化的铅钡玻璃氧化钾含量比无风化的变化更小。铅钡玻璃氧化

钾含量在 0.2%附近波动。

③风化的铅钡玻璃氧化钙含量比无风化的含量更高，变化更大。风化的铅钡玻璃氧化钙含量在 3%附近上下波动，无风化的含量在 2%附近波动。

④风化的铅钡玻璃氧化铝含量与无风化的趋势类似。铅钡玻璃的氧化铝含量在 3%附近波动。

⑤风化的铅钡玻璃氧化铁含量与无风化的趋势类似。铅钡玻璃的氧化铁含量在 1%附近波动。

⑥风化的铅钡玻璃氧化铜含量比无风化的含量更低，变化更小。风化的铅钡玻璃氧化铜含量在 4%上下波动，未风化的氧化铜含量在 2%附近波动。

⑦风化的铅钡玻璃氧化铅含量比无风化的含量更高，变化更大。风化的铅钡玻璃氧化铅含量在 50%上下波动，未风化的氧化铅含量在 20%附近波动。

⑧风化的铅钡玻璃氧化钡含量比无风化的含量更高，变化更大。风化的铅钡玻璃氧化钡含量在 15%上下波动，未风化的氧化钡含量在 10%附近波动。

⑨风化的铅钡玻璃五氧化二磷含量比无风化的含量更高，变化更大。风化的铅钡玻璃五氧化二磷含量在 6%上下波动，未风化的五氧化二磷含量在 2%附近波动。

(3) 预测风化前的化学成分含量见表格 6、表格 7、表格 8、表格 9。.

六、问题二模型建立与求解

6.1 使用 K-均值聚类分析高钾玻璃与铅钡玻璃的分析规律

为得到高钾玻璃与铅钡玻璃的分类规律，我们先使用主成分分析法降维，再进行 K-均值聚类，并将聚类结果与直接进行 K-均值聚类所得结果进行对比，最后选择较好的分类结果。最后利用较好的方式，给出对应的分类规律。

6.1.1 主成分分析法降维原理

(1) 主成分分析模型建立

① 原始数据标准化

利用各化学成分数据构建原始的数据矩阵 X , 设第 i 个化学成分含量对应的第 j 个指标的值为 x_{ij} , 并用以下公式对数据进行标准化处理:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}} (i = 1, 2, \dots, n; j = 1, 2, \dots, m), \quad (4)$$

其中, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$; $\text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$, ($j = 1, 2, \dots, m$).

② 计算相关系数矩阵

用以下公式计算化学成分之间的相关系数 r_{ij} :

$$r_{ij} = \frac{\sum_{k=1}^n x'_{ki} x'_{kj}}{n-1} \quad (5)$$

其中, $n > 1, i, j = 1, 2, \dots, m$,

③ 计算特征值和特征向量

解特征方程 $|\lambda I - R| = 0$, 求出特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ 及对应的特征向量 u_1, u_2, \dots, u_m , 其中 $u_j = (u_{1j}, u_{2j}, \dots, u_{mj})^T$, 且 $\|u_j\| = 1, j = 1, 2, \dots, m$, 则第 j 个主成分为

$$y_j = u_{1j}x_1 + u_{2j}x_2 + \dots + u_{mj}x_m$$

其中, $x_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T, j = 1, 2, \dots, m$

④ 计算贡献率及累计贡献率

$$\beta_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k}, (j = 1, 2, \dots, m) \quad (6)$$

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_k} \quad (7)$$

⑤ 计算综合得分

$$Z = \sum_{j=1}^p \beta_j y_j \quad (8)$$

6.1.2 K-均值聚类法分析规律

我们可以通过主成分分析得到的主成分来进行 K-均值聚类,也可以直接选取所有的化学成分来进行 K-均值聚类,对比聚类结果,选取最合适的聚类样本。

(1) K-均值聚类法模型建立

我们需要使用 K-均值聚类法选择两个初始的聚类中心 C_1 、 C_2 ,使得目标函数 F 最小。目标函数 F 公式如下:

$$F = \sum_{s \in S}^k \min \|s - C_i\|^2 \quad (9)$$

具体步骤:

- ① 随机选择两个初始聚类中心 $C = \{c_1, c_2\}$;
- ② $\forall s \in S$, 如果 s 与 c_i 的距离小于到 c_j 的距离,则将 s 划分到聚类簇 C_i 的中心;
- ③ 重新计算 C_i 中心:

$$c_i = \sum_{s \in C_i} \frac{s}{\|C_i\|}$$

- ④ 不断迭代至聚类中心 C 不变。

(2) 利用两种基于 K-均值聚类法的模型求解

① 直接用原本的数据进行 K-均值聚类

直接 K-均值聚类结果如下表 (完整表见附录):

表格 12: 直接用 K-均值聚类法数据分析表

个案号	文物采样点	聚类	距离
1	06部位2	2	15.501
2	06部位1	2	12.538
3	03部位2	2	14.328
4	18	2	16.974
5	05	2	15.245
6	16	2	16.367
7	01	2	13.029
8	03部位1	2	21.737
.....
66	28未风化点	2	8.835

表格中，聚类为“1”表示铅钡玻璃，“2”表示高钾玻璃。程序共处理67份数据，共判定出31个“1”，即铅钡玻璃；判定出36个“2”，即高钾玻璃。经比对，其中有18份错误数据。

正确率计算公式：

$$\alpha = \frac{n_c}{n_t} \times 100\% \quad (10)$$

其中， n_t 表示判断总量， n_c 表示判断正确的数量。

灵敏度计算公式：

$$\beta_i = \frac{n_{wi}}{n_{Ki}} \times 100\% \quad (11)$$

其中， n_{Ki} 表示判断出聚类*i*的数量， n_{wi} 表示判断正确的该聚类的数量。

故计算得正确率

$$\alpha = \frac{n_c}{n_t} \times 100\% = \frac{67 - 18}{67} \times 100\% = 73.13\%$$

灵敏度：

$$\beta_1 = \frac{n_{w1}}{n_{K1}} \times 100\% = \frac{31 - 18}{31} \times 100\% = 41.94\%$$

$$\beta_2 = \frac{n_{w2}}{n_{K2}} \times 100\% = \frac{36 - 18}{36} \times 100\% = 50.00\%$$

②先用主成分分析法降维，所得数据进行 K-均值聚类

先用主成分分析法降维，所得数据进行 K-均值聚类的结果如下表（完整表见附录）：

表格 13：先用主成分分析法降维再 K-均值聚类法数据分析表

个案号	文物采样点	聚类	距离
1	06部位2	1	3.836
2	06部位1	1	1.027
3	03部位2	1	2.786
4	18	1	1.256
5	05	1	2.913
6	16	1	2.354
7	01	1	2.031
8	03部位1	1	1.137
.....
66	28未风化点	1	1.669

表格中，聚类为“1”表示高钾玻璃，“2”表示铅钡玻璃。程序共处理 67 份数据，共判定出 28 个“1”，即高钾玻璃；判定出 39 个“2”，即铅钡玻璃。经比对，其中有 10 份错误数据。

正确率与灵敏度计算公式如上 6.1.2 (1) ①所示。

故正确率

$$\alpha = \frac{n_c}{n_t} \times 100\% = \frac{67 - 10}{67} \times 100\% = 85.07\%$$

灵敏度：

$$\beta_1 = \frac{n_{w1}}{n_{K1}} \times 100\% = \frac{28 - 10}{28} \times 100\% = 64.29\%$$

$$\beta_2 = \frac{n_{w2}}{n_{K2}} \times 100\% = \frac{39 - 10}{39} \times 100\% = 74.36\%$$

③ 总结

经比对，以上两种方法中，方法②“先采用主成分分析法降维，再基于 K-均值聚类法判别分析”的正确率更高、敏感性更低，故方法②更优。

利用先采用主成分分析法降维，再基于 K-均值聚类法判别分析的分析结果，得到高钾玻璃、铅钡玻璃的分析规律。

6.1.3 结论

经过以上数据分析得知高钾玻璃、铅钡玻璃分类规律：

离高钾玻璃聚类中心簇的距离近的大概率属高钾玻璃类，离高钾玻璃聚类中心簇距离远的大概率属铅钡玻璃类。具体分类规律见 6.3 问题二结论。

6.2 根据 K-均值聚类方法进行亚类划分

6.2.1 问题分析与模型建立

对于每个类别选择合适的化学成分进行亚类划分，我们可以基于先前选取合适化学成分数据进行 K-均值聚类，并对聚类所得的结果进行分析，判断所得类是哪些类型，并对分类的结果的合理性、敏感性通过正确率和灵敏度指标进行评价。

将玻璃分为两类，高钾玻璃和铅钡玻璃。接着对高钾玻璃进行 K-均值聚类分析。

K-均值聚类具体步骤依照 6.1.2 中进行。

6.2.2 模型求解与结果分析

选取风化与未分化数据进行 K-均值聚类，得到如下结果（完整表单见附录）：

表格 14：高钾玻璃 K-均值聚类法分析表

个案号	文物采样点	聚类	距离
1	06部位2	1	7.478
2	06部位1	1	9.533
3	03部位2	1	4.669
4	18	2	12.822
5	05	1	2.517
6	16	1	5.524
7	01	1	6.855
8	03部位1	2	4.474
.....
17	22	2	3.239
18	27	2	3.696

表格中，聚类为“1”表示未风化玻璃，“2”表示风化玻璃。程序共处理 18 组数据，其中有 9 个“1”，即未风化玻璃；9 个“2”，即风化玻璃。其中存在 3 个错误数据。

正确率与灵敏度计算公式如上 6.1.2（1）①所示。

故正确率

$$\alpha = \frac{n_c}{n_t} \times 100\% = \frac{18-3}{18} \times 100\% = 83.33\%$$

灵敏度：

$$\beta_1 = \frac{n_{w1}}{n_{K1}} \times 100\% = \frac{9-3}{9} \times 100\% = 66.67\%$$

$$\beta_2 = \frac{n_{w2}}{n_{K2}} \times 100\% = \frac{9-3}{9} \times 100\% = 66.67\%$$

表格 15：铅钡玻璃 K-均值聚类法分析表

个案号	文物采样点	聚类	距离
1	24	1	22.628
2	47	2	8.557
3	30部位2	1	14.397
4	46	2	6.413
5	30部位1	1	12.280

6	37	2	5.382
7	31	2	11.105
8	45	2	6.372
.....
48	56	1	6.626
49	57	1	7.461

表格中，聚类为“1”表示风化玻璃，“2”表示未风化玻璃。程序共处理 49 份数据，共判定出 27 个“1”，即风化玻璃；判定出 22 个“2”，即未风化玻璃。经比对，其中有 4 份错误数据。

正确率与灵敏度计算公式如上 6.1.2 (1) ①所示。

故正确率

$$\alpha = \frac{n_c}{n_t} \times 100\% = \frac{49 - 4}{49} \times 100\% = 89.80\%$$

灵敏度：

$$\beta_1 = \frac{n_{w1}}{n_{K1}} \times 100\% = \frac{27 - 4}{27} \times 100\% = 85.19\%$$

$$\beta_2 = \frac{n_{w2}}{n_{K2}} \times 100\% = \frac{22 - 4}{22} \times 100\% = 81.82\%$$

6.3 问题二结论

(1) 高钾玻璃、铅钡玻璃分类规律：

①离高钾玻璃聚类中心簇的距离近的属高钾玻璃类，离高钾玻璃聚类中心簇距离远的属铅钡玻璃类。

设成分得分系数矩阵为 $A_{9 \times 3}$ ，需要判断的玻璃的化学成分含量矩阵为 $B_{1 \times 9}$ ，用矩阵乘法： $B \times A = C$ ，得到矩阵 $C_{1 \times 3}$ ，矩阵 C 即为需要判断的玻璃的三个主成分的矩阵。再把 C 看作有三个元素的行向量，并设其三个元素分别为 c_{i1}, c_{i2}, c_{i3} ，通过欧氏距离计算公式 $d_i = \sqrt{\sum_{j=1}^n (c_{ij} - e_{ij})^2}$ ，计算得出玻璃 i 距离两个聚类中心的距离 d_1, d_2 。通过比较，若 $d_1 < d_2$ ，则对应元素被划分到聚类 1，即高钾玻璃中；若 $d_1 > d_2$ ，则对应元素被划分到聚类 2，即铅钡玻璃中。

表格 17：成分得分系数矩阵

化学成分种类	主成分 1	主成分 2	主成分 3
--------	-------	-------	-------

二氧化硅(SiO_2)	0.764	-0.585	0.063
氧化钾(K_2O)	0.732	0.343	0.295
氧化钙(CaO)	0.425	0.768	0.054
氧化铝(Al_2O_3)	0.617	0.282	-0.137
氧化铁(Fe_2O_3)	0.501	0.555	-0.099
氧化铜(CuO)	-0.238	0.251	0.868
氧化铅(PbO)	-0.795	0.257	-0.412
氧化钡(BaO)	-0.722	0.092	0.531
五氧化二磷(P_2O_5)	-0.399	0.63	-0.276

表格 18：主成分降维后进行 K-均值聚类所得聚类中心坐标

	聚类中心 i	
	$i = 1$	$i = 2$
主成分 e_{i1}	-1.780504057	1.278310605
主成分 e_{i2}	-0.351139318	0.252100023
主成分 e_{i3}	0.122353788	-0.087843745

②也可直接用 K-均值聚类所得聚类中心来得到高钾玻璃和铅钡玻璃的分类规律，直接聚类所得的聚类 1 为铅钡玻璃类，聚类 2 为高钾玻璃类，可从下表格对比得出，铅钡玻璃和高钾玻璃在二氧化硅、氧化钾、氧化铅、氧化钡等化学成分含量上有较为明显的差异，可以得到以下规律：

- 二氧化硅含量在 45.09% 上下波动的玻璃更有可能是铅钡玻璃；二氧化硅含量更高，在 67.33% 上下波动的玻璃更有可能是高钾玻璃。
- 氧化钾含量在 0.37% 上下波动的玻璃更有可能是铅钡玻璃；二氧化硅含量更高，在 4.94% 上下波动的玻璃更有可能是高钾玻璃。
- 氧化铅含量在 26.79% 上下波动的玻璃更有可能是铅钡玻璃；氧化铅含量更低，在 7.37% 上下波动的玻璃更有可能是高钾玻璃。
- 氧化钡含量在 10.77% 上下波动的玻璃更有可能是铅钡玻璃；氧化钡含量更低，在 2.88% 上下波动的玻璃更有可能是高钾玻璃。

表格 19：未风化直接 K-均值聚类所得聚类中心

	聚类中心 1 (铅钡)	聚类中心 2 (高钾)
二氧化硅(SiO_2)	45.09	67.33
氧化钾(K_2O)	0.37	4.94
氧化钙(CaO)	1.83	3.57
氧化铝(Al_2O_3)	3.74	6.15
氧化铁(Fe_2O_3)	1.42	1.44
氧化铜(CuO)	2.27	1.67
氧化铅(PbO)	26.79	7.37
氧化钡(BaO)	10.77	2.88
五氧化二磷(P_2O_5)	1.97	1.01

表格 20：风化直接 K-均值聚类所得聚类中心

	聚类中心 1 (铅钡)	聚类中心 2 (高钾)
二氧化硅(SiO_2)	24.31	86.54
氧化钾(K_2O)	0.12	0.63
氧化钙(CaO)	2.88	1.14
氧化铝(Al_2O_3)	2.53	4.46
氧化铁(Fe_2O_3)	0.55	0.39
氧化铜(CuO)	2.28	1.32
氧化铅(PbO)	44.03	2.79
氧化钡(BaO)	11.53	0.95
五氧化二磷(P_2O_5)	5.78	0.33

(2) 亚类划分：

对高钾玻璃、铅钡玻璃两个类别选择二氧化硅、氧化钾、氧化钙、氧化铝、氧化铁、氧化铜、氧化铅、氧化钡、五氧化二磷 9 种化学物质的数据对

其进行亚类划分。

具体划分方法：对每个类别玻璃分别采用 K-均值聚类法，亚类划分为类 1 与类 2。

划分结果：

(1) 对高钾玻璃，采用 K-均值聚类法，得到类 1 和类 2。我们认为，类 1 为未风化玻璃，类 2 为风化玻璃。

类 1 = {06 部位 2 , 06 部位 1 , 03 部位 2, 05, 16, 01, 04, 13, 14};

类 2 = {18, 03 部位 1 , 21, 07, 09, 10, 12, 22, 27}。

(2) 对铅钡玻璃，采用 K-均值聚类法，得到类 1 和类 2。我们认为，类 1 为风化玻璃，类 2 为未风化玻璃。

类 1 = {24, 30 部位 2 , 30 部位 1 , 54 严重风化点
, 54, 40, 50, 43 部位 1 , 26 严重风化点 , 39, 08 严重风化点
, 41, 43 部位 2, 49, 26, 52, 38, 51 部位 1 , 08, 11, 58, 34, 36, 02, 19, 51 部位 2 , 56, 57}
;

类 2 =
{47, 46, 37, 31, 45, 20, 32, 33, 35, 55, 42 未风化点 1 , 23 未风化点
, 49 未风化点 , 53 未风化点 , 44 未风化点 , 29 未风化点 , 50 未风化点
, 25 未风化点 , 28 未风化点 , 42 未风化点 2 , 48}

合理性分析：

由上，高钾玻璃亚类划分的正确率为 83.33%，铅钡玻璃亚类划分的正确率为 89.80%，数据合理性高。

由灵敏度计算结果 β_1, β_2 ，可得到敏感性较低的结论。

七、问题三模型建立与求解

7.1 利用 K-均值聚类鉴别未知类别玻璃文物所属类型

7.1.1 数据处理

将表单 2 与表单 3 中数据全部分为两类，分别为风化玻璃、未风化玻璃。采用分情况讨论的方式，用于后续 K-均值聚类过程。

7.1.2 问题分析与模型建立

将分类后的数据分别用 K-均值聚类方法选取两个聚类中心进行聚类，一共得到四类，据分析可知应为风化高钾玻璃、风化铅钡玻璃、未风化高钾玻璃、未风化铅钡玻璃，据此可将未知玻璃文物进行类型鉴别，接着通过数据扰动并观察结果变动来进行敏感性分析。

K-均值聚类具体步骤依照 6.1.2 中进行。

7.1.3 模型求解与结果分析

经 K-均值聚类数据分析，得到下表（完整表见附录）：

表格 21：未风化玻璃制品类别预测表

个案号	文物采样点	聚类	距离
1	06部位2	2	13.005
2	06部位1	2	10.444
3	03部位2	2	11.845
4	18	2	16.039
5	05	2	12.321
.....
35	42未风化点2	1	9.822
36	A1	2	14.794
37	A3	1	20.983
38	A4	1	13.652
39	A8	1	10.891

表格中，聚类为“1”表示铅钡玻璃，“2”表示高钾玻璃。程序共处理 39 份数据。分析数据，容易得到以下信息：

表格 22：未分类部分玻璃制品预测结果

文物编号	风化情况	类型
A1	未风化	高钾玻璃

A3	未风化	铅钡玻璃
A4	未风化	铅钡玻璃
A8	未风化	铅钡玻璃

同理，得到如下表单：

表格 23：风化玻璃制品类别预测表

个案号	文物采样点	聚类	距离
1	07	2	7.492
2	09	2	9.528
3	10	2	11.312
4	12	2	8.844
5	22	2	6.653
.....
32	57	1	8.562
33	A2	1	22.542
34	A5	2	25.593
35	A6	2	7.890
36	A7	2	5.262

表格中，聚类为“1”表示铅钡玻璃，“2”表示高钾玻璃。程序共处理 36 份数据。分析所得数据，容易得到下列信息：

表格 24：未分类部分玻璃制品预测结果

文物编号	风化情况	类型
A2	风化	铅钡玻璃
A5	风化	高钾玻璃
A6	风化	高钾玻璃
A7	风化	高钾玻璃

7.2 结论

经 K-均值聚类法，对未分类玻璃做出以下鉴别：

表格 25：未分类部分玻璃制品鉴别结果

文物编号	风化情况	类型
A1	未风化	高钾玻璃
A2	风化	铅钡玻璃
A3	未风化	铅钡玻璃

A4	未风化	铅钡玻璃
A5	风化	高钾玻璃
A6	风化	高钾玻璃
A7	风化	高钾玻璃
A8	未风化	铅钡玻璃

对此结果做出如下敏感性分析：

参考上 6.3 表格 16、表格 17，通过每个数据向相反聚类中心方向移动的方式扰动数据，观察其抗干扰能力，得到结果如下表：

表格 26：无风化数据扰动结果				表格 27：风化数据扰动结果			
	0%	5%	10%		0%	5%	10%
A1	高钾	高钾	高钾	A2	铅钡	铅钡	铅钡
A3	铅钡	铅钡	铅钡	A5	高钾	高钾	高钾
A4	铅钡	铅钡	铅钡	A6	高钾	高钾	高钾
A8	铅钡	铅钡	铅钡	A7	高钾	高钾	高钾

通过观察可得，在扰动数据为 0%~10% 时，聚类预测结果没有发生改变，得到“敏感性较低，聚类模型较稳定”的结论。

八、问题四模型建立与求解

8.1 基于相关系数矩阵求解化学成分之间的关联关系

8.1.1 相关系数矩阵分析法方法具体步骤

(1) 原始数据标准化

利用各组玻璃制品的化学成分构造数据矩阵如下：

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

接着运用如下公式对数据进行标准化处理：

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (12)$$

其中，

$$x_j = \frac{1}{n} \sum_{i=1}^n x_{ij}; Var(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2 \quad (j = 1, 2, \dots, m).$$

(2) 计算相关系数矩阵

计算标准化后的相关系数矩阵 R

$$R = \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix}$$

$$\text{其中, } r_{ij} = \frac{\sum_{k=1}^n x'_{ki} x'_{kj}}{n-1}, n > 1, i, j = 1, 2, \dots, m.$$

8.1.3 模型的建立与求解

依据上式, 对铅钡玻璃的化学成分计算相关系数矩阵, 得铅钡玻璃相关系数矩阵 (相关系数在-1~1 范围内, 数值绝对值越大, 对应的化学成分含量之间的关联越强, 数值为正则是正相关, 数值为负则为负相关):

表格 28: 铅钡玻璃相关系数部分矩阵 (一)

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)
二氧化硅(SiO_2)	1	0.362	0.087	-0.488
氧化钠(Na_2O)	0.362	1	-0.059	-0.373
氧化钾(K_2O)	0.087	-0.059	1	0.096
.....
氧化锶(SrO)	-0.503	-0.093	-0.121	0.171
氧化锡(SnO_2)	0.085	-0.083	0.195	0.21
二氧化硫(SO_2)	-0.386	-0.13	0.01	0.104

注: 完整表见附录表格 43~表格 45。

同理, 对高钾玻璃化学物质成分计算相关系数矩阵, 得高钾玻璃相关系数矩阵 (相关系数在-1~1 范围内, 数值绝对值越大, 对应的化学成分含量之间的关联越强, 数值为正则是正相关, 数值为负则为负相关):

表格 29: 高钾玻璃相关系数部分矩阵 (一)

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)
二氧化硅(SiO_2)	1	-0.457	-0.877	-0.823
氧化钠(Na_2O)	-0.457	1	0.554	0.621
氧化钾(K_2O)	-0.877	0.554	1	0.759
.....
氧化锡(SnO_2)	0.049	-0.106	0.142	-0.29

二氧化硫(SO_2)	-0.357	-0.194	0.332	0.429
----------------	--------	--------	-------	-------	-------

注：完整表见附录表格 46~表格 48。

8.2 依据相关系数矩阵比较不同类别之间化学成分关联关系

用铅钡玻璃的相关系数矩阵减去高钾玻璃的相关系数矩阵，得差值结果相关系数矩阵（相关系数在-2~2 范围内，数值绝对值越大，对应化学成分含量的不同类别间的相关性的差异越大）：

表格 30：差值结果相关系数部分矩阵（一）

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)
二氧化硅(SiO_2)	0	0.819	0.964	0.335
氧化钠(Na_2O)	0.819	0	-0.613	-0.994
氧化钾(K_2O)	0.964	-0.613	0	-0.663
.....
氧化锶(SrO)	0.022	0.089	-0.542	0.136
氧化锡(SnO_2)	0.036	0.023	0.053	0.5
二氧化硫(SO_2)	-0.029	0.064	-0.322	-0.325

注：完整表见附录表格 49~表格 51。

八、模型评价

8.1 模型的优缺点

- （1）本模型综合利用了主成分分析、K-均值聚类等模型算法，并利用了决策树模型使得结果更加地形象清晰。
- （2）本模型考虑得比较全面细致，对每种化学成分都进行了风化与未风化之间的比较，给出了大量的折线图、散点图作为比较依据。
- （3）本模型对所得结果进行了分析检验，并具有较好的准确率以及不错的稳定性。
- （4）本模型在进行聚类时分别用了直接聚类和先主成分分析降维后再聚类两种方式的比对，选出较优的聚类方式。
- （5）本模型在选用数据时带有一定的主观性，剔除了一部分认为是不必要的的数据，这可能会带来结果上一定的偏差。

(6) 本模型在进行敏感度分析时选用的灵敏度这一指标，在说服力上有所欠缺。

8.2 模型的改进方向

我们的模型是在附件的数据支持下，通过一些合理的假设建立的。通过假设使得模型的建立更为简单直观。但是这可能会使结果与实际上的有所偏差。除此之外，还有一些其他的因素未合理考虑，比如玻璃类型、纹饰、颜色之间可能存在一些关联关系。因时间以及人力关系，我们只是讨论了比较简单的情形，而上述问题都是有待于改进的方向。

九、参考文献

- [1]周志华. 机器学习[M]. 清华大学出版社. 2016-01.
- [2]刘红良,李成福. 数学模型与建模算法[M]. 科学出版社. 2016-08.
- [3]余胜威. MATLAB 数学建模经典案例实战[M]. 清华大学出版社. 2015-1.
- [4]韩中庚,宋明武,邵广纪. 数学建模竞赛——获奖论文精选与点评[M]. 科学出版社. 2007-05.

十、附录

附录 1：支撑材料文件列表

附录 2：补充表格和图片

附录 3：代码

附录 1：支撑材料文件列表

表格 31：支撑材料文件列表

date.mat	problem1_1.m
problem.1_2_1.m	problem1_2_2.m
tree.fig	x.mat
y1.mat	习题 1.1.xlsx
习题 1.2(1).xlsx	习题 1.2(2).xlsx
习题 1.2（统计）.xlsx	预测数据.xlsx
2.1 直接聚类_1.sav	2.1 主成分加聚类（优）.sav
2.2.1 高钾.sav	2.2.2 铅钼.sav
problem2_2.m	问题 2.1 直接聚类.spv
问题 2.1 主成分+聚类（优）.spv	问题 2.2.1 高钾.spv
问题 2.2.2 铅钼.spv	习题 2.1（主）.xlsx
习题 2.1.xlsx	习题 2.2.1（高钾）.xlsx
习题 2.2.2（铅钼）.xlsx	
问题 3.1.1（无风化）.spv	问题 3.1.2（风化）
习题 3.1.1（无风化）.xlsx	习题 3.1.2（风化）
problem4_2.m	差值计算.xlsx
习题 4.1.1 铅钼.spv	习题 4.1.2 高钾.spv
习题 4.1.xlsx	习题 4.2.xlsx

附录 2：补充表格和图片

1、结构树原图

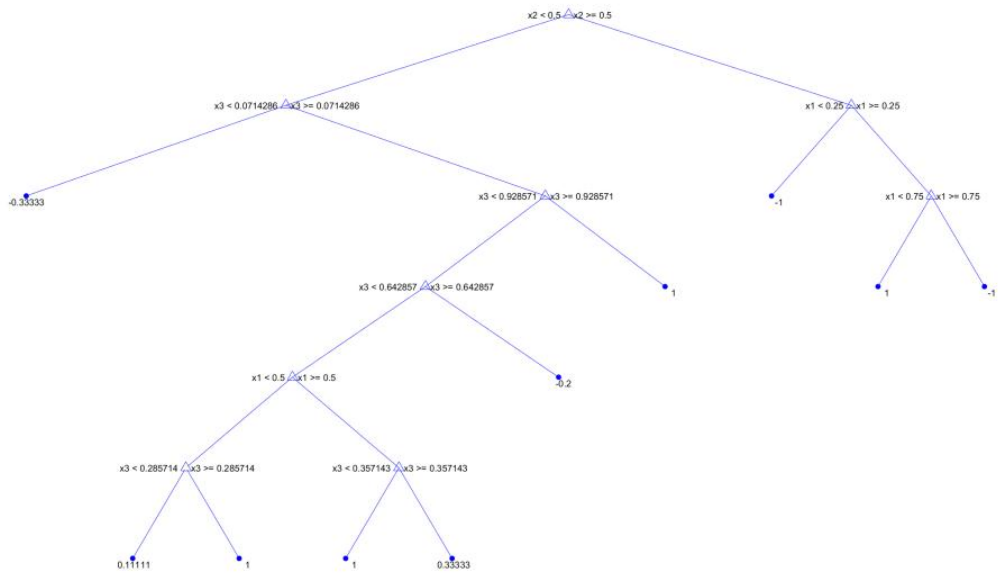


图 12：结构树程序分析原图

2、问题一结果

表格 32：决策树文字解析

类别	类型	纹饰	颜色	风化情况
1	铅钋	A、B、C	浅绿	未风化(0.33333)
2	铅钋	A	浅蓝、蓝绿、绿色	风化(0.11111)
3	铅钋	A	深绿	风化
4	铅钋	B、C	浅蓝、蓝绿	风化
5	铅钋	B、C	绿色、深绿	风化(0.33333)
6	铅钋	A、B、C	深蓝、紫色	未风化(0.2)
7	铅钋	A、B、C	黑色	风化
8	高钾	A	所有颜色	未风化
9	高钾	B	所有颜色	风化
10	高钾	C	所有颜色	未风化

(1) 用散点图及拟合线描述与求解不同化学成分含量的关系

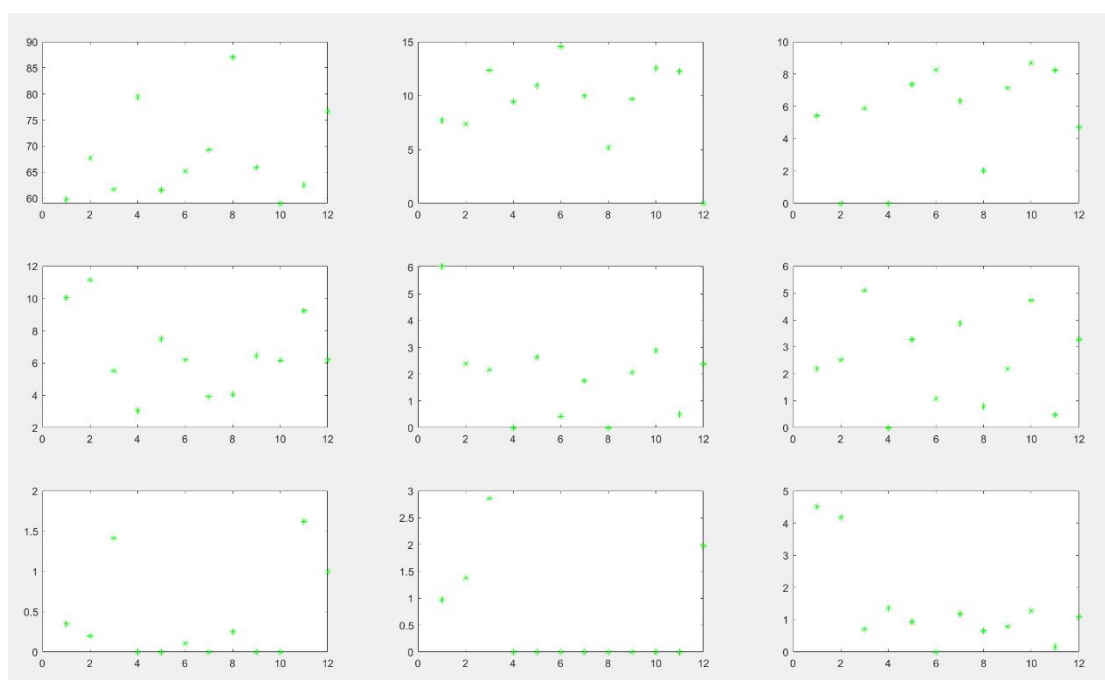


图 13: 高钾玻璃各化学成分散点图

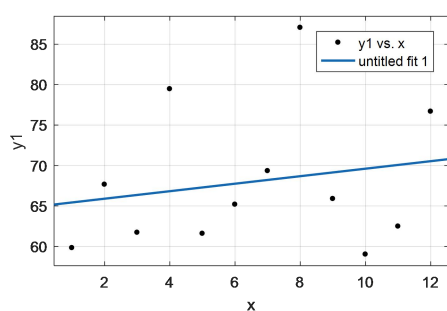


图 14: 高钾玻璃二氧化硅散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.464 \text{ (- 1.215, 2.143)}$$

$$k_2 = 64.97 \text{ (52.61, 77.33)}$$

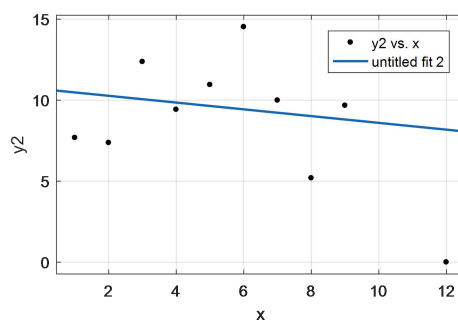


图 15: 高钾玻璃氧化钾散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.2088 \text{ (- 0.9607, 0.543)}$$

$$k_2 = 10.69 \text{ (5.155, 16.22)}$$

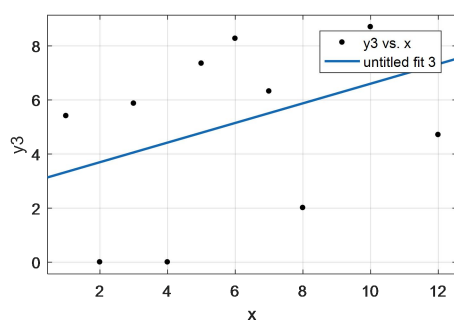


图 16: 高钾玻璃氧化钙散点及拟合线

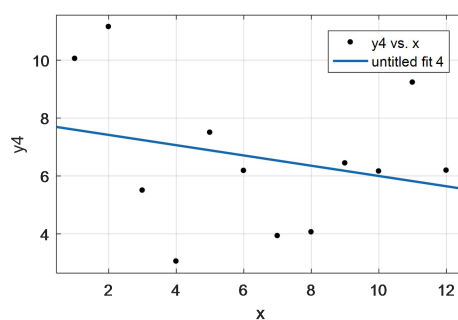


图 17: 高钾玻璃氧化铝散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.363 (-0.1846, 0.9105)$$

$$k_2 = 2.973 (-1.057, 7.003)$$

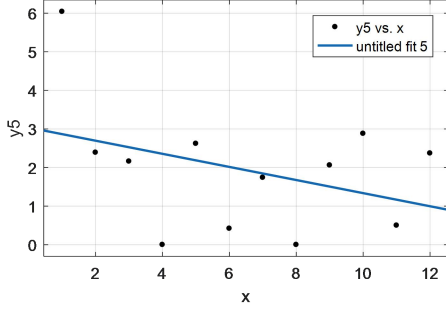


图 18: 高钾玻璃氧化铁散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.1774 (-0.648, 0.2932)$$

$$k_2 = 7.773 (4.31, 11.24)$$

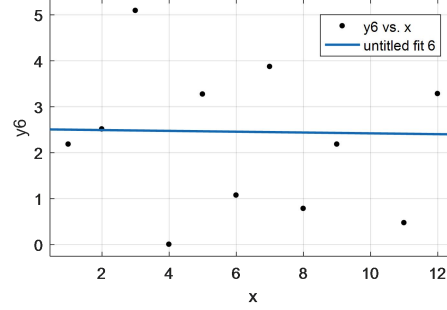


图 19: 高钾玻璃氧化铜散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.1699 (-0.4728, 0.1331)$$

$$k_2 = 3.036 (0.8063, 5.265)$$

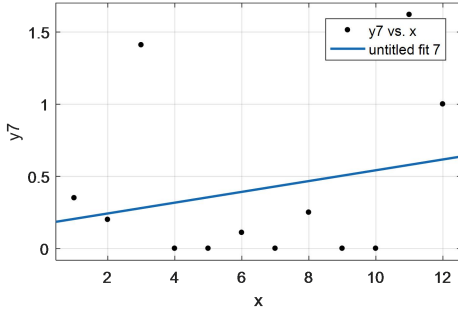


图 20: 高钾玻璃氧化铅散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.008916 (-0.3333, 0.3154)$$

$$k_2 = 2.51 (0.1234, 4.897)$$

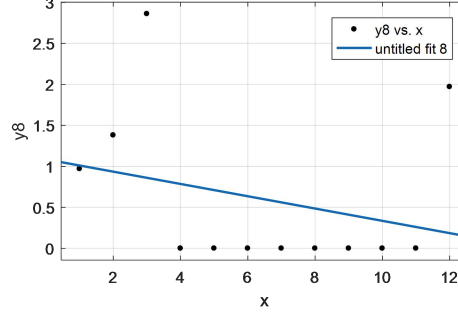


图 21: 高钾玻璃氧化钡散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.03741 (-0.07463, 0.1495)$$

$$k_2 = 0.1685 (-0.6361, 0.9931)$$

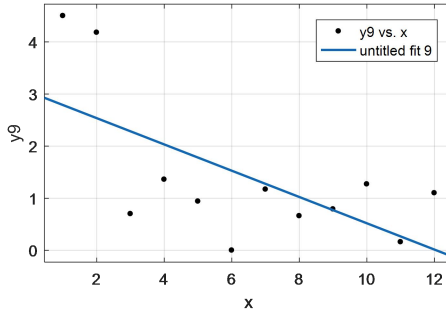


图 22: 高钾玻璃五氧化二磷散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.07497 (-0.2595, 0.1095)$$

$$k_2 = 1.086 (-0.2724, 2.444)$$

$$f(x) = k_1x + k_2$$

$$k_1 = -0.2521 (-0.4688, -0.03542)$$

$$k_2 = 3.041 (1.446, 4.636)$$

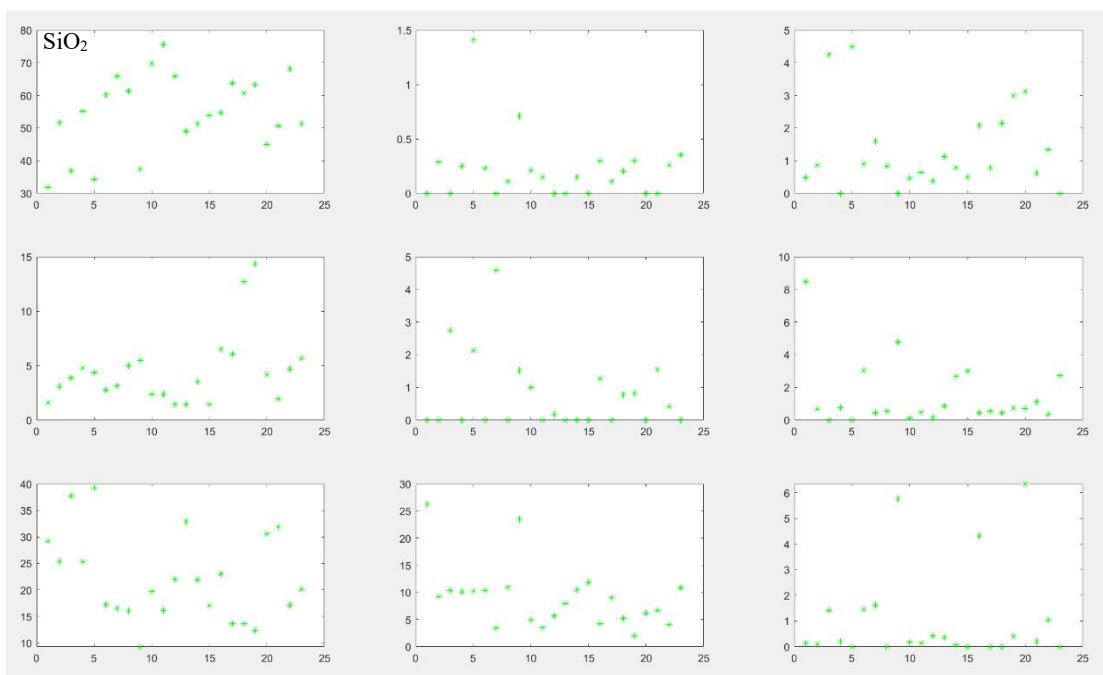


图 23: 铅钡玻璃各化学成分散点图

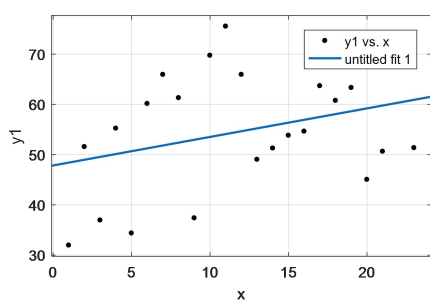


图 24: 铅钡玻璃二氧化硅散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.5679 (-0.1804, 1.316)$$

$$k_2 = 47.85 (37.58, 58.11)$$

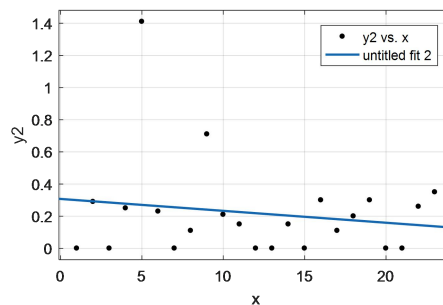


图 25: 铅钡玻璃氧化钾散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.007401 (-$$

$$-0.02787, 0.01307)$$

$$k_2 = 0.3075 (0.0268, 0.5882)$$

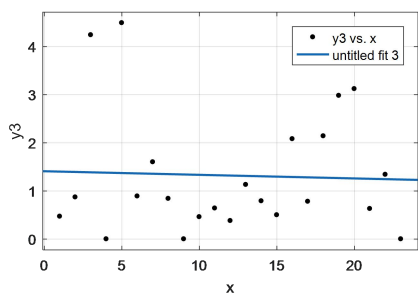


图 26: 铅钡玻璃氧化钙散点及拟合线

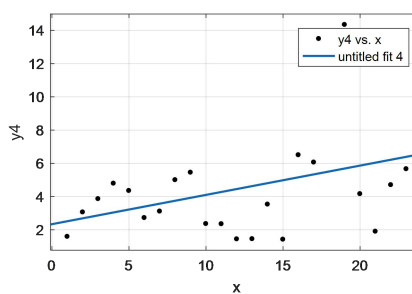


图 27: 铅钡玻璃氧化铝散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.00747 \text{ (} -0.09336, 0.07842 \text{)}$$

$$k_2 = 1.41 \text{ (} 0.2324, 2.588 \text{)}$$

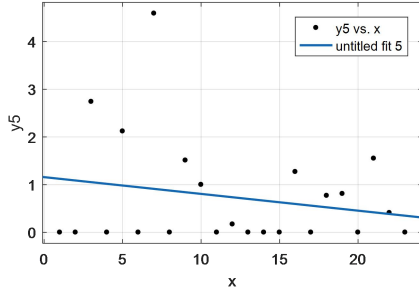


图 28: 铅钡玻璃氧化铁散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = 0.1762 \text{ (} -0.02689, 0.3793 \text{)}$$

$$k_2 = 2.341 \text{ (} -0.4436, 5.126 \text{)}$$

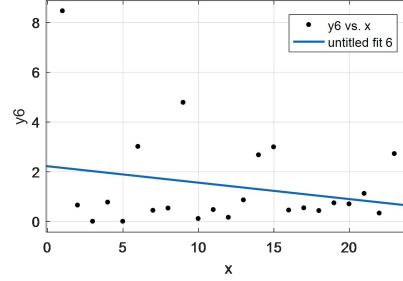


图 29: 铅钡玻璃氧化铜散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.03514 \text{ (} -0.1107, 0.04046 \text{)}$$

$$k_2 = 1.158 \text{ (} 0.1216, 2.195 \text{)}$$

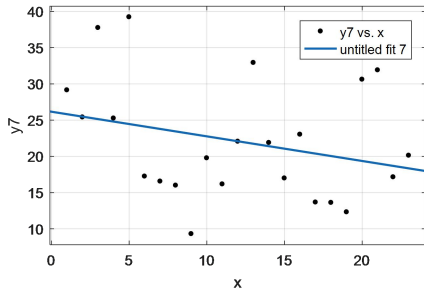


图 30: 铅钡玻璃氧化铅散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.06601 \text{ (} -0.1944, 0.06235 \text{)}$$

$$k_2 = 2.224 \text{ (} 0.4639, 3.984 \text{)}$$

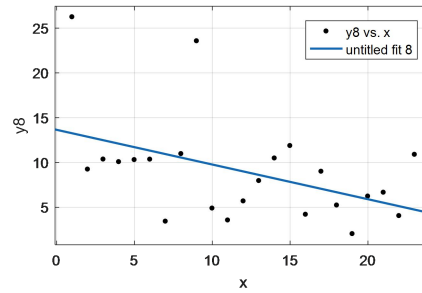


图 31: 铅钡玻璃氧化钡散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.3399 \text{ (} -0.8675, 0.1877 \text{)}$$

$$k_2 = 26.16 \text{ (} 18.93, 33.4 \text{)}$$

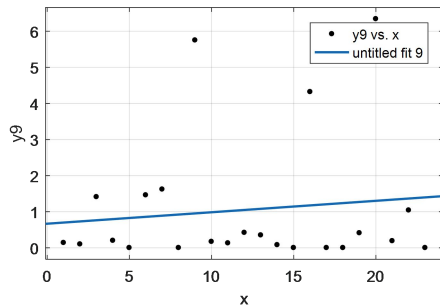


图 32: 铅钡玻璃五氧化二磷散点及拟合线

$$f(x) = k_1x + k_2$$

$$k_1 = -0.3874 \text{ (} -0.7353, 0.0395 \text{)}$$

$$k_2 = 13.65 \text{ (} 8.88, 18.42 \text{)}$$

$$f(x) = k_1x + k_2$$

$$k_1 = 0.0317 \text{ (} -0.09105, 0.1544 \text{)}$$

$$k_2 = 0.6687 \text{ (} -1.014, 2.352 \text{)}$$

(2) 借助折线图判断统计规律

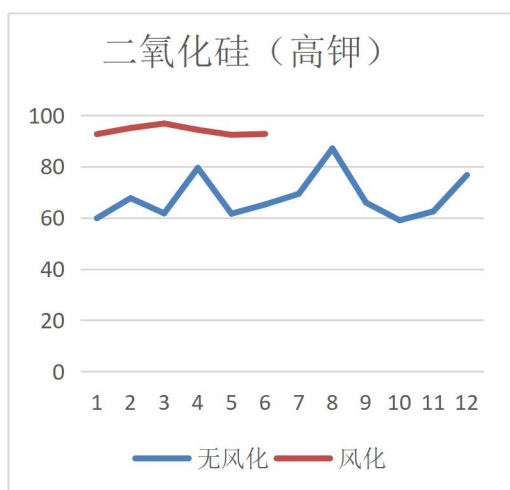


图 33: 高钾玻璃二氧化硅含量折线图

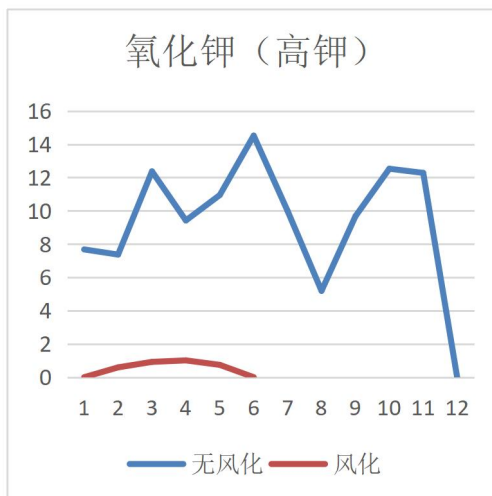


图 34: 高钾玻璃氧化钾含量折线图

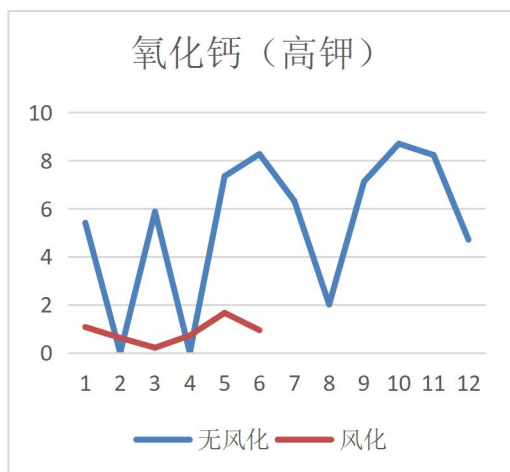


图 35: 高钾玻璃氧化钙含量折线图

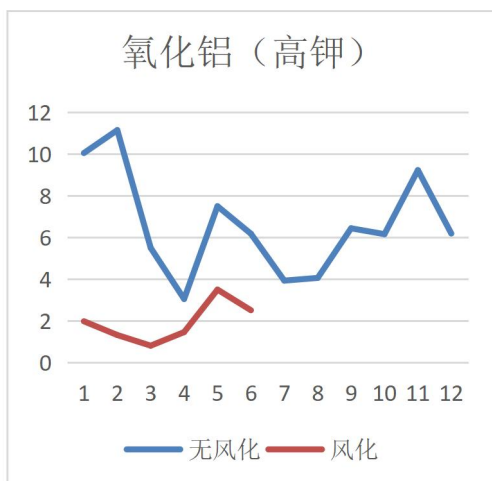


图 36: 高钾玻璃氧化铝含量折线图

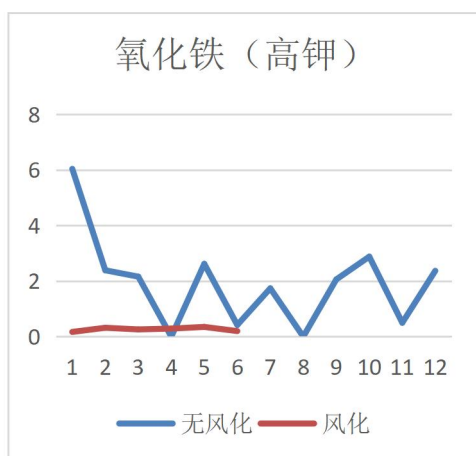


图 37: 高钾玻璃氧化铁含量折线图

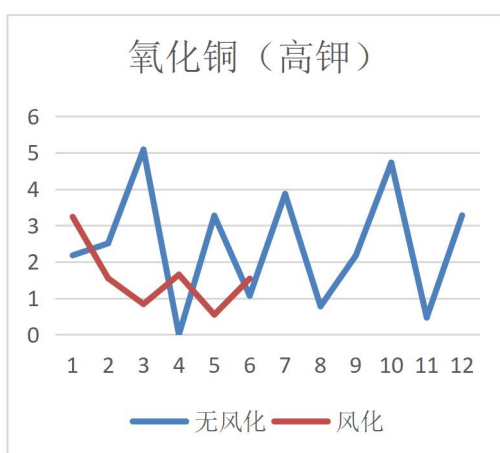


图 38: 高钾玻璃氧化铜含量折线图

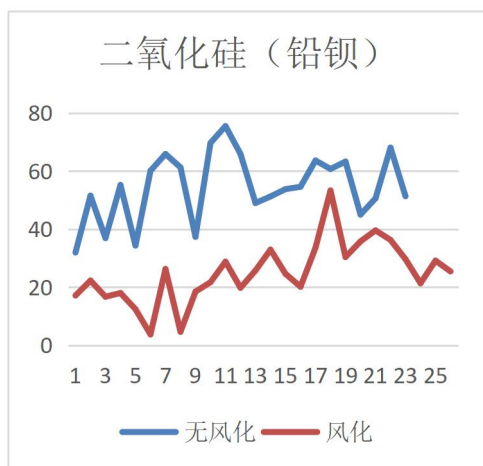


图 39：铅钡玻璃二氧化硅含量折线图

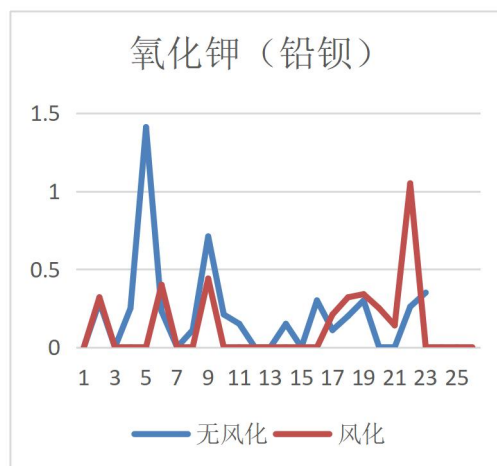


图 40：铅钡玻璃氧化钾含量折线图

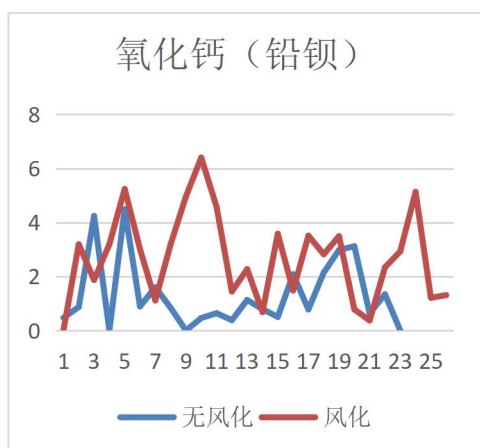


图 41：铅钡玻璃氧化钙含量折线图

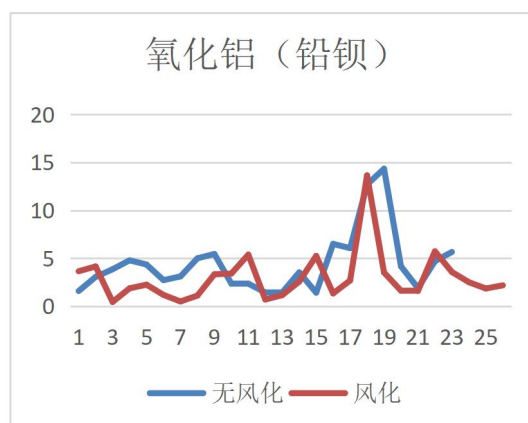


图 42：铅钡玻璃氧化铝含量折线图

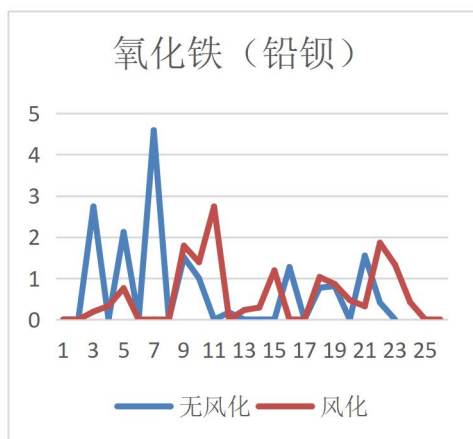


图 43：铅钡玻璃氧化铁含量折线图

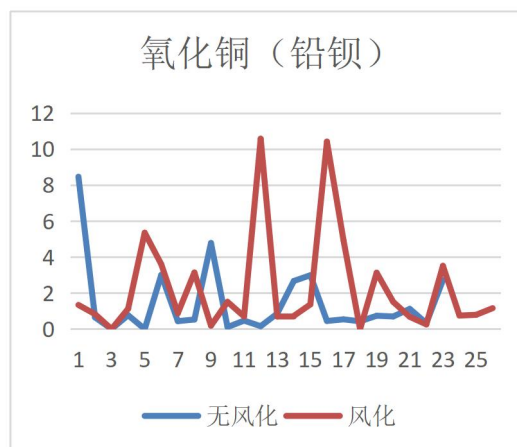


图 44：铅钡玻璃氧化铜含量折线图

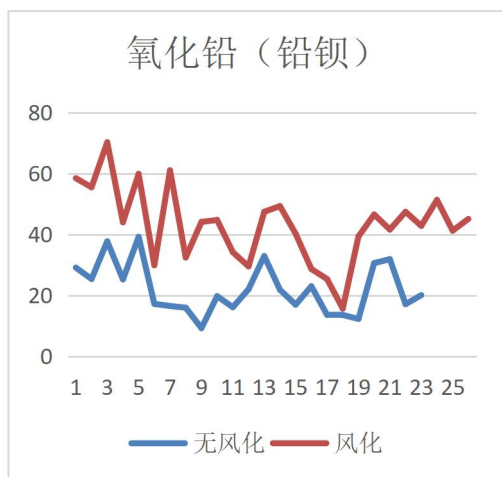


图 45：铅钡玻璃氧化铅含量折线图

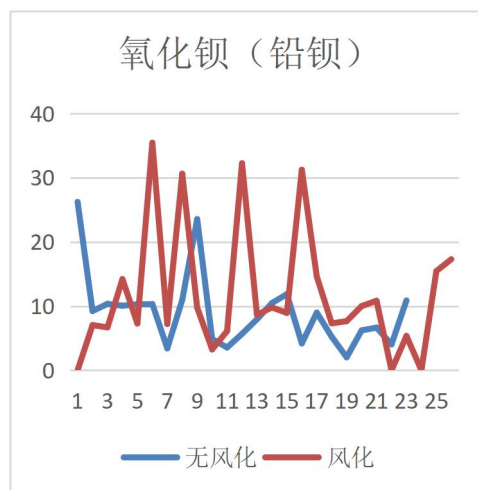


图 46：铅钡玻璃氧化钡含量折线图

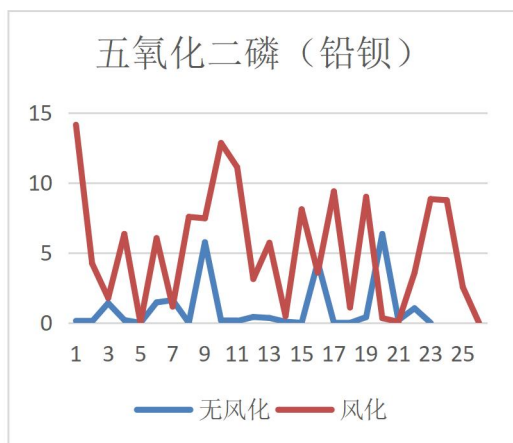


图 47：铅钡玻璃五氧化二磷含量折线图

表格 33：高钾玻璃风化样本点风化前部分化学成分预测值（一）

预测值	二氧化硅	氧化钾	氧化钙	氧化铝	氧化铁
07	66.65	8.79	5.53	6.67	1.84
09	69.04	9.38	5.08	6.01	1.99
10	70.79	9.71	4.67	5.50	1.93
12	68.31	9.80	5.18	6.15	1.96
22	66.37	9.53	6.12	8.19	2.02
27	66.74	8.79	5.40	7.20	1.87

表格 34：高钾玻璃风化样本点风化前部分化学成分预测值（二）

预测值	氧化铜	氧化铅	氧化钡	五氧化二磷
07	4.13	0.41	0.60	1.73
09	2.44	0.41	0.60	1.47
10	1.73	0.41	0.60	1.12
12	2.54	0.41	0.60	1.27
22	1.44	0.41	0.60	1.33
27	2.43	0.41	0.60	1.48

表格 35：铅钡玻璃风化样本点风化前部分化学成分预测值（一）

预测值	二氧化硅	氧化钾	氧化钙	氧化铝	氧化铁
54 严重	46.86	0.09	0	5.14	0.15
54	52.03	0.41	1.82	5.64	0.15
40	46.46	0.09	0.5	1.94	0.34
50	47.73	0.09	1.82	3.36	0.48
43 部位 1	42.16	0.09	3.87	3.74	0.91
26 严重	33.47	0.49	1.64	2.67	0.15
39	56	0.09	0	1.99	0.15
08 严重	34.36	0.09	1.82	2.6	0.15
41	48.21	0.53	3.59	4.82	0.94
43 部位 2	51.45	0.09	5.03	4.9	1.54
49	58.54	0.09	3.21	6.87	2.89
26	49.54	0.09	0.07	2.19	0.15
52	55.49	0.09	0.9	2.65	0.38
38	62.68	0.09	0	4.06	0.44
51 部位 1	54.36	0.09	2.21	6.74	1.34
08	49.89	0.09	0.11	2.83	0.15
11	63.34	0.3	2.14	4.18	0.15

48	83.08	0.41	1.45	15.14	1.18
58	60.14	0.43	2.12	5.01	1.01
34	65.53	0.34	0	3.11	0.62
36	69.32	0.23	0	3.09	0.47
02	66.03	1.14	0.97	7.22	2.01
19	59.39	0.09	1.56	5.06	1.48
51 部位 2	51.1	0.09	3.76	4	0.57
56	58.9	0.09	0	3.34	0.15
57	55.17	0.09	0	3.67	0.15

表格 36：铅钡玻璃风化样本点风化前部分化学成分预测值（二）

预测值	氧化铜	氧化铅	氧化钡	五氧化二磷
54 严重	0.5	37.23	0	9.9
54	0	34.23	4.23	0.01
40	0	48.98	3.88	0
50	0.29	22.77	11.39	2.11
43 部位 1	4.51	38.62	4.48	0
26 严重	2.76	8.69	32.64	1.81
39	0.04	39.8	4.41	0
08 严重	2.3	11.22	27.81	3.33
41	0	22.89	6.95	3.23
43 部位 2	0.67	23.52	0.45	8.6
49	0	12.95	3.29	6.87
26	9.73	8.3	29.44	0
52	0	26.19	5.83	1.48
38	0	28.08	6.98	0
51 部位 1	0.53	19.01	6.13	3.87
08	9.57	7.45	28.42	0

11	4.09	4.16	11.8	5.15
48	0	0	4.5	0
58	2.29	18.12	4.85	4.76
34	0.67	25.32	7.19	0
36	0	20.38	8.02	0
02	0	26.2	0	0
19	2.67	21.59	2.54	4.6
51 部位 2	0	30.11	0	4.52
56	0	20.02	12.64	0
57	0.32	23.87	14.49	0

注：其中负值均设为零值。

2、直接 K-均值聚类所得表单：

表格 37：直接 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	06部位2	2	15.501
2	06部位1	2	12.538
3	03部位2	2	14.328
4	18	2	16.974
5	05	2	15.245
6	16	2	16.367
7	01	2	13.029
8	03部位1	2	21.737
9	04	2	13.024
10	13	2	17.789
11	14	2	15.320
12	21	2	13.043
13	07	2	27.246
14	09	2	29.442
15	10	2	31.155
16	12	2	28.713
17	22	2	26.725
18	27	2	27.235
19	54严重风化点	1	24.515
20	54	1	15.002

21	40	1	30.541
22	50	1	9.894
23	43部位1	1	24.168
24	26严重风化点	1	35.595
25	39	1	19.841
26	08严重风化点	1	31.159
27	41	1	10.146
28	43部位2	1	13.869
29	49	1	12.717
30	26	1	26.665
31	52	1	6.658
32	38	1	10.491
33	51部位1	1	6.338
34	08	1	26.113
35	11	1	19.134
36	48	2	18.594
37	58	1	7.609
38	34	1	10.915
39	36	1	13.607
40	02	1	16.335
41	19	1	8.307
42	51部位2	1	16.763
43	56	1	5.268
44	57	1	8.113
45	24	1	21.638
46	47	2	23.996
47	30部位2	1	11.948
48	46	2	21.734
49	30部位1	1	10.029
50	37	2	13.578
51	31	2	9.556
52	45	2	12.315
53	20	2	36.689
54	32	2	11.959
55	33	2	11.595
56	35	2	14.390
57	55	1	24.602
58	42未风化点1	2	22.262
59	23未风化点	2	18.724
60	49未风化点	2	19.816
61	53未风化点	2	8.809
62	44未风化点	2	11.796

63	29未风化点	2	11.351
64	50未风化点	1	22.269
65	25未风化点	1	26.678
66	28未风化点	2	8.835
67	42未风化点2	2	21.354

3、主成分加聚类所得表单

表格 38：先主成分分析降维，后 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	06部位2	1	3.836
2	06部位1	1	1.027
3	03部位2	1	2.786
4	18	1	1.256
5	05	1	2.913
6	16	1	2.354
7	01	1	2.031
8	03部位1	1	1.137
9	04	1	2.230
10	13	1	3.651
11	14	1	2.378
12	21	1	.663
13	07	1	1.918
14	09	1	2.045
15	10	1	2.292
16	12	1	1.989
17	22	1	1.640
18	27	1	1.925
19	54严重风化点	2	2.474
20	54	2	1.136
21	40	2	1.670
22	50	2	1.037
23	43部位1	2	1.233
24	26严重风化点	2	2.714
25	39	2	1.330
26	08严重风化点	2	2.372
27	41	2	2.084
28	43部位2	2	2.940
29	49	2	2.816
30	26	2	4.503
31	52	2	1.026

32	38	2	1.120
33	51部位1	2	1.592
34	08	2	4.326
35	11	2	1.367
36	48	1	1.142
37	58	2	1.239
38	34	2	.962
39	36	2	1.337
40	02	2	2.068
41	19	2	1.319
42	51部位2	2	2.113
43	56	2	.853
44	57	2	.968
45	24	2	3.469
46	47	2	1.760
47	30部位2	2	1.635
48	46	2	1.928
49	30部位1	2	1.574
50	37	2	1.827
51	31	1	1.209
52	45	1	2.050
53	20	2	1.702
54	32	1	1.989
55	33	1	2.210
56	35	2	2.350
57	55	2	1.549
58	42未风化点1	2	1.712
59	23未风化点	2	2.016
60	49未风化点	1	1.775
61	53未风化点	1	1.833
62	44未风化点	1	.832
63	29未风化点	1	.963
64	50未风化点	2	1.057
65	25未风化点	2	1.462
66	28未风化点	1	1.669
67	42未风化点2	2	1.850

4、问题二亚类划分 K-均值聚类结果

表格 39：直接 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	06部位2	1	7.478

2	06部位1	1	9.533
3	03部位2	1	4.669
4	18	2	12.822
5	05	1	2.517
6	16	1	5.524
7	01	1	6.855
8	03部位1	2	4.474
9	04	1	3.052
10	13	1	5.970
11	14	1	4.791
12	21	2	14.378
13	07	2	4.077
14	09	2	5.775
15	10	2	7.584
16	12	2	4.969
17	22	2	3.239
18	27	2	3.696

表格 40：直接 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	24	1	22.628
2	47	2	8.557
3	30部位2	1	14.397
4	46	2	6.413
5	30部位1	1	12.280
6	37	2	5.382
7	31	2	11.105
8	45	2	6.372
9	20	2	28.227
10	32	2	12.995
11	33	2	19.174
12	35	2	9.814
13	55	2	16.001
14	42未风化点1	2	7.433
15	23未风化点	2	7.479
16	49未风化点	2	6.821
17	53未风化点	2	9.001
18	44未风化点	2	10.886
19	29未风化点	2	14.718
20	50未风化点	2	17.498
21	25未风化点	2	14.336

22	28未风化点	2	11.666
23	42未风化点2	2	7.225
24	54严重风化点	1	23.119
25	54	1	13.628
26	40	1	28.936
27	50	1	7.505
28	43部位1	1	22.147
29	26严重风化点	1	34.204
30	39	1	19.015
31	08严重风化点	1	29.607
32	41	1	8.196
33	43部位2	1	13.161
34	49	1	14.029
35	26	1	26.277
36	52	1	6.047
37	38	1	11.498
38	51部位1	1	6.430
39	08	1	25.801
40	11	1	20.825
41	48	2	10.642
42	58	1	9.358
43	34	1	12.647
44	36	1	15.931
45	02	1	17.795
46	19	1	9.487
47	51部位2	1	15.859
48	56	1	6.626
49	57	1	7.461

5、问题三 K-均值聚类划分结果

表格 41：未风化 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	06部位2	2	13.005
2	06部位1	2	10.444
3	03部位2	2	11.845
4	18	2	16.039
5	05	2	12.321
6	16	2	13.487
7	01	2	10.468
8	03部位1	2	21.397
9	04	2	10.022

10	13	2	15.074
11	14	2	12.371
12	21	2	12.571
13	24	1	21.631
14	47	1	7.441
15	30部位2	1	14.141
16	46	1	10.820
17	30部位1	1	16.986
18	37	2	15.818
19	31	2	11.660
20	45	2	14.581
21	20	1	23.579
22	32	2	14.594
23	33	2	13.963
24	35	2	16.924
25	55	1	8.567
26	42未风化点1	1	8.312
27	23未风化点	1	13.663
28	49未风化点	1	12.829
29	53未风化点	2	11.216
30	44未风化点	2	12.572
31	29未风化点	2	11.485
32	50未风化点	1	7.803
33	25未风化点	1	9.110
34	28未风化点	2	11.368
35	42未风化点2	1	9.822
36	A1	2	14.794
37	A3	1	20.983
38	A4	1	13.652
39	A8	1	10.891

表格 42：风化 K-均值聚类所得结果

个案号	文物采样点	聚类	距离
1	07	2	7.492
2	09	2	9.528
3	10	2	11.312
4	12	2	8.844
5	22	2	6.653
6	27	2	7.156
7	54严重风化点	1	21.757
8	54	1	12.745

9	40	1	28.168
10	50	1	7.034
11	43部位1	1	21.414
12	26严重风化点	1	34.634
13	39	1	18.500
14	08严重风化点	1	29.887
15	41	1	7.153
16	43部位2	1	11.824
17	49	1	13.894
18	26	1	27.235
19	52	1	5.170
20	38	1	11.857
21	51部位1	1	5.967
22	08	1	26.757
23	11	1	21.533
24	48	2	37.422
25	58	1	9.288
26	34	1	13.254
27	36	1	16.779
28	02	1	17.598
29	19	1	8.971
30	51部位2	1	14.536
31	56	1	7.925
32	57	1	8.562
33	A2	1	22.542
34	A5	2	25.593
35	A6	2	7.890
36	A7	2	5.262

6、问题四结论

表格 43：铅钡玻璃相关系数部分矩阵（一）

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化镁 (MgO)
二氧化硅(SiO_2)	1	0.362	0.087	-0.488	0.088
氧化钠(Na_2O)	0.362	1	-0.059	-0.373	0.026
氧化钾(K_2O)	0.087	-0.059	1	0.096	0.269
氧化钙(CaO)	-0.488	-0.373	0.096	1	0.417
氧化镁(MgO)	0.088	0.026	0.269	0.417	1
氧化铝(Al_2O_3)	0.401	0.103	0.306	0.135	0.45
氧化铁(Fe_2O_3)	0.082	-0.244	0.248	0.387	0.293
氧化铜(CuO)	-0.354	-0.058	-0.146	-0.068	-0.309

氧化铅(PbO)	-0.738	-0.349	-0.131	0.356	-0.036
氧化钡(BaO)	-0.436	-0.056	-0.041	-0.124	-0.453
五氧化二磷 (P_2O_5)	-0.565	-0.396	-0.106	0.535	0.272
氧化锶(SrO)	-0.503	-0.093	-0.121	0.171	0.087
氧化锡(SnO_2)	0.085	-0.083	0.195	0.21	0.246
二氧化硫(SO_2)	-0.386	-0.13	0.01	0.104	-0.266

表格 44：铅钡玻璃相关系数部分矩阵（二）

	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)
二氧化硅(SiO_2)	0.401	0.082	-0.354	-0.738	-0.436
氧化钠(Na_2O)	0.103	-0.244	-0.058	-0.349	-0.056
氧化钾(K_2O)	0.306	0.248	-0.146	-0.131	-0.041
氧化钙(CaO)	0.135	0.387	-0.068	0.356	-0.124
氧化镁(MgO)	0.45	0.293	-0.309	-0.036	-0.453
氧化铝(Al_2O_3)	1	0.23	-0.282	-0.441	-0.339
氧化铁(Fe_2O_3)	0.23	1	-0.247	-0.048	-0.295
氧化铜(CuO)	-0.282	-0.247	1	-0.085	0.719
氧化铅(PbO)	-0.441	-0.048	-0.085	1	-0.14
氧化钡(BaO)	-0.339	-0.295	0.719	-0.14	1
五氧化二磷 (P_2O_5)	-0.074	0.153	0.096	0.331	-0.02
氧化锶(SrO)	-0.155	-0.091	0.196	0.434	0.164
氧化锡(SnO_2)	0.469	0.226	-0.184	-0.126	-0.077
二氧化硫(SO_2)	-0.206	-0.18	0.241	-0.065	0.634

表格 45：铅钡玻璃相关系数部分矩阵（三）

	五氧化二磷 (P_2O_5)	氧化锶 (SrO)	氧化锡 (SnO_2)	二氧化硫 (SO_2)
二氧化硅(SiO_2)	-0.565	-0.503	0.085	-0.386
氧化钠(Na_2O)	-0.396	-0.093	-0.083	-0.13
氧化钾(K_2O)	-0.106	-0.121	0.195	0.01
氧化钙(CaO)	0.535	0.171	0.21	0.104
氧化镁(MgO)	0.272	0.087	0.246	-0.266
氧化铝(Al_2O_3)	-0.074	-0.155	0.469	-0.206
氧化铁(Fe_2O_3)	0.153	-0.091	0.226	-0.18
氧化铜(CuO)	0.096	0.196	-0.184	0.241
氧化铅(PbO)	0.331	0.434	-0.126	-0.065
氧化钡(BaO)	-0.02	0.164	-0.077	0.634
五氧化二磷(P_2O_5)	1	0.28	-0.082	0.173

氧化锶(SrO)	0.28	1	-0.038	0.181
氧化锡(SnO_2)	-0.082	-0.038	1	-0.071
二氧化硫(SO_2)	0.173	0.181	-0.071	1

表格 46：高钾玻璃相关系数部分矩阵（一）

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化镁 (MgO)
二氧化硅(SiO_2)	1	-0.457	-0.877	-0.823	-0.603
氧化钠(Na_2O)	-0.457	1	0.554	0.621	-0.245
氧化钾(K_2O)	-0.877	0.554	1	0.759	0.387
氧化钙(CaO)	-0.823	0.621	0.759	1	0.205
氧化镁(MgO)	-0.603	-0.245	0.387	0.205	1
氧化铝(Al_2O_3)	-0.831	0.345	0.591	0.536	0.698
氧化铁(Fe_2O_3)	-0.687	-0.023	0.336	0.457	0.595
氧化铜(CuO)	-0.465	-0.022	0.251	0.423	0.141
氧化铅(PbO)	-0.398	0.359	0.275	0.368	0.166
氧化钡(BaO)	-0.345	-0.214	0.093	0.078	0.427
五氧化二磷 (P_2O_5)	-0.448	-0.186	0.171	-0.031	0.65
氧化锶(SrO)	-0.525	-0.182	0.421	0.035	0.712
氧化锡(SnO_2)	0.049	-0.106	0.142	-0.29	0.261
二氧化硫(SO_2)	-0.357	-0.194	0.332	0.429	0.408

表格 47：高钾玻璃相关系数部分矩阵（二）

	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)
二氧化硅(SiO_2)	-0.831	-0.687	-0.465	-0.398	-0.345
氧化钠(Na_2O)	0.345	-0.023	-0.022	0.359	-0.214
氧化钾(K_2O)	0.591	0.336	0.251	0.275	0.093
氧化钙(CaO)	0.536	0.457	0.423	0.368	0.078
氧化镁(MgO)	0.698	0.595	0.141	0.166	0.427
氧化铝(Al_2O_3)	1	0.689	0.21	0.422	0.381
氧化铁(Fe_2O_3)	0.689	1	0.526	0.151	0.452
氧化铜(CuO)	0.21	0.526	1	0.165	0.521
氧化铅(PbO)	0.422	0.151	0.165	1	0.632
氧化钡(BaO)	0.381	0.452	0.521	0.632	1
五氧化二磷 (P_2O_5)	0.676	0.765	0.213	-0.014	0.378
氧化锶(SrO)	0.595	0.591	0.187	0.188	0.565
氧化锡(SnO_2)	-0.163	-0.219	-0.361	-0.133	-0.118

二氧化硫(SO ₂)	0.147	0.232	0.301	-0.244	-0.216
------------------------	-------	-------	-------	--------	--------

表格 48：高钾玻璃相关系数部分矩阵（三）

	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
二氧化硅(SiO ₂)	-0.448	-0.525	0.049	-0.357
氧化钠(Na ₂ O)	-0.186	-0.182	-0.106	-0.194
氧化钾(K ₂ O)	0.171	0.421	0.142	0.332
氧化钙(CaO)	-0.031	0.035	-0.29	0.429
氧化镁(MgO)	0.65	0.712	0.261	0.408
氧化铝(Al ₂ O ₃)	0.676	0.595	-0.163	0.147
氧化铁(Fe ₂ O ₃)	0.765	0.591	-0.219	0.232
氧化铜(CuO)	0.213	0.187	-0.361	0.301
氧化铅(PbO)	-0.014	0.188	-0.133	-0.244
氧化钡(BaO)	0.378	0.565	-0.118	-0.216
五氧化二磷(P ₂ O ₅)	1	0.741	0.065	-0.021
氧化锶(SrO)	0.741	1	0.24	-0.049
氧化锡(SnO ₂)	0.065	0.24	1	-0.108
二氧化硫(SO ₂)	-0.021	-0.049	-0.108	1

表格 49：差值结果相关系数部分矩阵（一）

	二氧化硅 (SiO ₂)	氧化钠 (Na ₂ O)	氧化钾 (K ₂ O)	氧化钙 (CaO)	氧化镁 (MgO)
二氧化硅(SiO ₂)	0	0.819	0.964	0.335	0.691
氧化钠(Na ₂ O)	0.819	0	-0.613	-0.994	0.271
氧化钾(K ₂ O)	0.964	-0.613	0	-0.663	-0.118
氧化钙(CaO)	0.335	-0.994	-0.663	0	0.212
氧化镁(MgO)	0.691	0.271	-0.118	0.212	0
氧化铝(Al ₂ O ₃)	1.232	-0.242	-0.285	-0.401	-0.248
氧化铁(Fe ₂ O ₃)	0.769	-0.221	-0.088	-0.07	-0.302
氧化铜(CuO)	0.111	-0.036	-0.397	-0.491	-0.45
氧化铅(PbO)	-0.34	-0.708	-0.406	-0.012	-0.202
氧化钡(BaO)	-0.091	0.158	-0.134	-0.202	-0.88
五氧化二磷 (P ₂ O ₅)	-0.117	-0.21	-0.277	0.566	-0.378
氧化锶(SrO)	0.022	0.089	-0.542	0.136	-0.625
氧化锡(SnO ₂)	0.036	0.023	0.053	0.5	-0.015
二氧化硫(SO ₂)	-0.029	0.064	-0.322	-0.325	-0.674

表格 50：差值结果相关系数部分矩阵（二）

	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)
二氧化硅(SiO_2)	1.232	0.769	0.111	-0.34	-0.091
氧化钠(Na_2O)	-0.242	-0.221	-0.036	-0.708	0.158
氧化钾(K_2O)	-0.285	-0.088	-0.397	-0.406	-0.134
氧化钙(CaO)	-0.401	-0.07	-0.491	-0.012	-0.202
氧化镁(MgO)	-0.248	-0.302	-0.45	-0.202	-0.88
氧化铝(Al_2O_3)	0	-0.459	-0.492	-0.863	-0.72
氧化铁(Fe_2O_3)	-0.459	0	-0.773	-0.199	-0.747
氧化铜(CuO)	-0.492	-0.773	0	-0.25	0.198
氧化铅(PbO)	-0.863	-0.199	-0.25	0	-0.772
氧化钡(BaO)	-0.72	-0.747	0.198	-0.772	0
五氧化二磷 (P_2O_5)	-0.75	-0.612	-0.117	0.345	-0.398
氧化锶(SrO)	-0.75	-0.682	0.009	0.246	-0.401
氧化锡(SnO_2)	0.632	0.445	0.177	0.007	0.041
二氧化硫(SO_2)	-0.353	-0.412	-0.06	0.179	0.85

表格 51：差值结果相关系数部分矩阵（三）

	五氧化二磷 (P_2O_5)	氧化锶 (SrO)	氧化锡 (SnO_2)	二氧化硫 (SO_2)
二氧化硅(SiO_2)	-0.117	0.022	0.036	-0.029
氧化钠(Na_2O)	-0.21	0.089	0.023	0.064
氧化钾(K_2O)	-0.277	-0.542	0.053	-0.322
氧化钙(CaO)	0.566	0.136	0.5	-0.325
氧化镁(MgO)	-0.378	-0.625	-0.015	-0.674
氧化铝(Al_2O_3)	-0.75	-0.75	0.632	-0.353
氧化铁(Fe_2O_3)	-0.612	-0.682	0.445	-0.412
氧化铜(CuO)	-0.117	0.009	0.177	-0.06
氧化铅(PbO)	0.345	0.246	0.007	0.179
氧化钡(BaO)	-0.398	-0.401	0.041	0.85
五氧化二磷(P_2O_5)	0	-0.461	-0.147	0.194
氧化锶(SrO)	-0.461	0	-0.278	0.23
氧化锡(SnO_2)	-0.147	-0.278	0	0.037
二氧化硫(SO_2)	0.194	0.23	0.037	0

附录 3 代码

1.1 处理数据，画出决策树（适用于 matlab）

```
clear
clc

%% 读取数据

data=xlsread('习题 1.1.xlsx',1,'A2:D55');
save date

%%前 n-1 维为输入数据，最后一维为输出数据
int=data(:,1:end-1);
out=data(:,end);

N=length(out); %样本总数
test_num=5; %测试集数据个数
train_num=N-test_num; %训练集数据个数

%% 划分训练集和测试集
int_train = int(1:train_num,:)';
out_train =out(1:train_num)';
int_test =int(train_num+1:train_num+test_num,:)';
out_test =out(train_num+1:train_num+test_num)';

%% 数据归一化
[inputn,inputps]=mapminmax(int_train,0,1);
[outputn,outputps]=mapminmax(out_train);
inputn_test=mapminmax('apply',int_test,inputps);
inputn=inputn';
outputn=outputn';
inputn_test=inputn_test';

%% 创建决策树
Mdl = fitrtree(inputn,outputn);
view(Mdl,'Mode','graph')
```

1.2.1 散点图拟合（高钾的化学成分）（适用于 matlab）

```
x=xlsread('习题 1.2 (1).xlsx','Sheet2','A2:A13');
save x;
```

```

y1=xlsread('习题 1.2 (1).xlsx','Sheet2','B2:B13');
y2=xlsread('习题 1.2 (1).xlsx','Sheet2','C2:C13');
y3=xlsread('习题 1.2 (1).xlsx','Sheet2','D2:D13');
y4=xlsread('习题 1.2 (1).xlsx','Sheet2','E2:E13');
y5=xlsread('习题 1.2 (1).xlsx','Sheet2','F2:F13');
y6=xlsread('习题 1.2 (1).xlsx','Sheet2','G2:G13');
y7=xlsread('习题 1.2 (1).xlsx','Sheet2','H2:H13');
y8=xlsread('习题 1.2 (1).xlsx','Sheet2','I2:I13');
y9=xlsread('习题 1.2 (1).xlsx','Sheet2','J2:J13');
save y1,y2,y3,y4,y5,y6,y7,y8,y9;
%画出散点图
subplot(3,3,1),plot(x,y1,'g*'),
subplot(3,3,2),plot(x,y2,'g*'),
subplot(3,3,3),plot(x,y3,'g*'),
subplot(3,3,4),plot(x,y4,'g*'),
subplot(3,3,5),plot(x,y5,'g*'),
subplot(3,3,6),plot(x,y6,'g*'),
subplot(3,3,7),plot(x,y7,'g*'),
subplot(3,3,8),plot(x,y8,'g*'),
subplot(3,3,9),plot(x,y9,'g*')
%%进行拟合
cftool

```

1.2.2 散点图拟合（铅钨的化学成分）（适用于 matlab）

```

x=xlsread('习题 1.2 (2).xlsx','Sheet2','A2:A24');
save x;
y1=xlsread('习题 1.2 (2).xlsx','Sheet2','B2:B24');
y2=xlsread('习题 1.2 (2).xlsx','Sheet2','C2:C24');
y3=xlsread('习题 1.2 (2).xlsx','Sheet2','D2:D24');
y4=xlsread('习题 1.2 (2).xlsx','Sheet2','E2:E24');
y5=xlsread('习题 1.2 (2).xlsx','Sheet2','F2:F24');
y6=xlsread('习题 1.2 (2).xlsx','Sheet2','G2:G24');
y7=xlsread('习题 1.2 (2).xlsx','Sheet2','H2:H24');
y8=xlsread('习题 1.2 (2).xlsx','Sheet2','I2:I24');
y9=xlsread('习题 1.2 (2).xlsx','Sheet2','J2:J24');
save y1,y2,y3,y4,y5,y6,y7,y8,y9;
%画出散点图
subplot(3,3,1),plot(x,y1,'g*'),
subplot(3,3,2),plot(x,y2,'g*'),
subplot(3,3,3),plot(x,y3,'g*'),
subplot(3,3,4),plot(x,y4,'g*'),

```

```

subplot(3,3,5),plot(x,y5,'g*'),
subplot(3,3,6),plot(x,y6,'g*'),
subplot(3,3,7),plot(x,y7,'g*'),
subplot(3,3,8),plot(x,y8,'g*'),
subplot(3,3,9),plot(x,y9,'g*')
%进行拟合
cftool

```

2.1 直接聚类 (SPSS)

```

GET
  FILE='C:\Users\86159\Desktop\数模编程资料\问题2\2.1直接聚类_1.sav'.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

问题 2.1 主成分加聚类

计算主成分 (适用于 matlab) :

```

A=xlsread('习题 2.1 (主).xlsx','sheet2','B2:J68');

% 数据标准化处理
a=size(A,1);
b=size(A,2);
for i=1:b
    s1(:,i)=(A(:,i)-mean(A(:,i)))/std(A(:,i));
end

%%计算相关系数矩阵的特征值和特征向量
s2=corrcoef(s1); %计算相关系数矩阵
[V,D]=eig(s2); %计算特征值和特征向量

for j=1:b

```



```

        DS(j, 1)=D(b+1-j, b+1-j); %对特征值进
行降序
    end

    for i=1:b
        DS(i, 2)=DS(i, 1)/sum(DS(:, 1)); %计算贡献率
        DS(i, 3)=sum(DS(1:i, 1))/sum(DS(:, 1)); %计算累计贡
献率
    end

    %%选择主成分及对应的特征向量
    T=0.74 %主成分信息保留率（根据
结果保留三个）
    for n=1:b
        if DS(n, 3)>=T
            Com_num=n;
            break;
        end
    end

    %%提取主成分对应的特征向量
    for j=1:Com_num
        PV(:, j)=V(:, b+1-j);
    end

    %%计算各评价对象的主成分得分
    new_score=s1*PV;
    for i=1:a
        score(i, 1)=sum(new_score(i, :));
        score(i, 2)=i;
    end

    result=[new_score, score]; %将各主成分得分与总分放
    在同一个矩阵中
    result=sortrows(result, -4); %按总分降序排序

    %%输出模型和结果报告
    disp('特征值及其贡献率、累计贡献率：')
    DS
    disp('信息保留率 T 对应的主成分数与特征向量')
    Com_num
    PV

```

```
disp('主成分得分及其排序（按第四列的总分进行降序排序，前三列为各主成分得分，第五列为编号）')
result
```

聚类(SPSS 计算聚类)

```
GET
  FILE='C:\Users\86159\Desktop\数模编程资料\问题2\2.1直接聚类_1.sav'.
GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\数模编程资料\问题2\习题2.1(主).xlsx'
  /SHEET=name 'Sheet3'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集3 WINDOW=FRONT.
```

```
GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\数模编程资料\问题2\习题2.1(主).xlsx'
  /SHEET=name 'Sheet3'
  /CELLRANGE=FULL
  /READNAMES=OFF
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集4 WINDOW=FRONT.
QUICK CLUSTER V1 V2 V3
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(V6) INITIAL ANOVA CLUSTER DISTAN.
```

2.2.1高钾亚类（SPSS）

```
GET DATA
  /TYPE=XLSX
```

```

/FILE='E:\bin\附件.xlsx'
/SHEET=name 'Sheet2'
/CELLRANGE=FULL
/READNAMES=OFF
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集4 WINDOW=FRONT.

GET DATA
  /TYPE=XLSX
  /FILE='E:\bin\附件.xlsx'
  /SHEET=name 'Sheet2'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集5 WINDOW=FRONT.
DATASET ACTIVATE 数据集1.
DATASET CLOSE 数据集4.
DATASET ACTIVATE 数据集5.
DATASET CLOSE 数据集1.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化铝Al2O3 氧化钙CaO 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(20) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

2.2.2 铅钡亚类 (SPSS)

```

GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\习题2.2.2.xlsx'
  /SHEET=name 'Sheet2'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.

```

```

DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
/MISSING=LISTWISE
/CRITERIA=CLUSTER(2) MXITER(20) CONVERGE(0)
/METHOD=KMEANS(NOUPDATE)
/SAVE CLUSTER DISTANCE
/PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

问题 3.1.1 无风化（SPSS）

```

GET DATA
/TYPE=XLSX
/FILE='C:\Users\86159\Desktop\习题3.1.1（无风化）.xlsx'
/SHEET=name 'Sheet2'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
/MISSING=LISTWISE
/CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
/METHOD=KMEANS(NOUPDATE)
/SAVE CLUSTER DISTANCE
/PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

问题 3.1.2 风化（SPSS）

```

GET DATA
/TYPE=XLSX
/FILE='C:\Users\86159\Desktop\习题3.1.2（风化）.xlsx'
/SHEET=name 'Sheet2'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁

```

```

Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
/MISSING=LISTWISE
/CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
/METHOD=KMEANS(NOUPDATE)
/SAVE CLUSTER DISTANCE
/PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

问题3.2.1 对高钾扰动后的数据进行聚类分析 (SPSS)

+0.05

```

GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\习题3.1.1（无风化）.xlsx'
  /SHEET=name '+0.05'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

+0.1

```

GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\习题3.1.1（无风化）.xlsx'
  /SHEET=name '+0.1'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)

```

```

/METHOD=KMEANS(NOUPDATE)
/SAVE CLUSTER DISTANCE
/PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

3.2.2 对铅钡扰动后的数据进行聚类分析 (SPSS)

+0.05

```

GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\习题3.1.2 (风化).xlsx'
  /SHEET=name '+0.05'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE
  /PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

```

+0.1

```

GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\86159\Desktop\习题3.1.2 (风化).xlsx'
  /SHEET=name '+0.1'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
QUICK CLUSTER 二氧化硅SiO2 氧化钾K2O 氧化钙CaO 氧化铝Al2O3 氧化铁
Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(2) MXITER(25) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER DISTANCE

```

/PRINT ID(文物采样点) INITIAL ANOVA CLUSTER DISTAN.

问题 4. 1. 1 铅钡化学成分相关性 (SPSS)

FACTOR

/VARIABLES 二氧化硅SiO2 氧化钠Na2O 氧化钾K2O 氧化钙CaO 氧化镁MgO 氧化铝Al2O3 氧化铁Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5

氧化锶SrO 氧化锡SnO2 二氧化硫SO2

/MISSING LISTWISE

/ANALYSIS 二氧化硅SiO2 氧化钠Na2O 氧化钾K2O 氧化钙CaO 氧化镁MgO 氧化铝Al2O3 氧化铁Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5

氧化锶SrO 氧化锡SnO2 二氧化硫SO2

/PRINT INITIAL CORRELATION EXTRACTION

/CRITERIA FACTORS(7) ITERATE(25)

/EXTRACTION PC

/ROTATION NOROTATE

/SAVE REG(ALL)

/METHOD=CORRELATION.

问题 4. 1. 2 高钾化学成分相关性 (SPSS)

GET DATA

/TYPE=XLSX

/FILE='C:\Users\86159\Desktop\数模编程资料\问题4\习题4.1.xlsx'

/SHEET=name '高钾'

/CELLRANGE=FULL

/READNAMES=ON

/DATATYPEMIN PERCENTAGE=95.0

/HIDDEN IGNORE=YES.

EXECUTE.

DATASET NAME 数据集2 WINDOW=FRONT.

FACTOR

/VARIABLES 二氧化硅SiO2 氧化钠Na2O 氧化钾K2O 氧化钙CaO 氧化镁MgO 氧化铝Al2O3 氧化铁Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5

氧化锶SrO 氧化锡SnO2 二氧化硫SO2

/MISSING LISTWISE

/ANALYSIS 二氧化硅SiO2 氧化钠Na2O 氧化钾K2O 氧化钙CaO 氧化镁MgO 氧化铝Al2O3 氧化铁Fe2O3 氧化铜CuO 氧化铅PbO 氧化钡BaO 五氧化二磷P2O5

氧化锶SrO 氧化锡SnO2 二氧化硫SO2
/PRINT INITIAL CORRELATION EXTRACTION
/CRITERIA FACTORS(7) ITERATE(25)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL)
/METHOD=CORRELATION.

问题4.2计算差值

```
x1=xlsread('差值计算.xlsx',1,'B2:O15');  
x2=xlsread('差值计算.xlsx',2,'B2:O15');  
save x1 , x2;  
xlswrite('差值计算.xlsx',x1-x2,3,'B2:O15');
```