

基于层次聚类对玻璃文物的

分类鉴定与成分分析

摘要

本文基于层次聚类模型对玻璃文物分类问题进行分析,对文物各成分含量作出分析,并依据纹饰、颜色、某一元素成分含量等因素对此作聚类分析。

对于问题一,首先针对数据进行整理。使用斯皮尔曼相关系数(Spearman correlation)分析相关性,同时采用 Pearson 卡方检验其差异性。以此分析玻璃类型、纹饰和颜色与表面风化的关系。

为结合玻璃类型分析风化前后化学成分的规律,将玻璃进一步分化为高钾未风化、高钾已风化、铅钡已风化、铅钡未风化四大类,求各类中成分元素均值,做出对比。

预测风化点风化前的化学成分含量,依据玻璃类型,分别将已知风化和未风化的样本各成分均值作为期望,做风化前后浮动比例。如若成分存在 0 的情况,则不适用比例法,进而采用差值法预测。预测结果见文中 5.2.3。

对于问题二,采用控制变量法。分别对纹饰、颜色和成分控制变量后,分析得出高钾玻璃与铅钡玻璃分类规律。采用层次聚类法分别对高钾玻璃与铅钡玻璃作亚分类。以共性分类相关系数做亚分类合理性分析,以均值差的线性回归做亚分类敏感性分析。

对于问题三,对待检测文物进行分类可以采用欧氏空间方法。将高钾已风化、高钾未风化、铅钡已风化、铅钡未风化的样品的各类成分均值作 14 元素,以此作欧氏空间,在欧氏空间中确定 4 个大类中心点,再将待测文物对应的点与 4 个中心点作欧氏距离,以最小值并且结合斯皮尔曼相关系数确定其所属类别(详见 7.2.1)。以距离均值、方差等因素判断其敏感性。

对于问题四,要求分析对于不同类别的文物样品其化学成分之间的相关关系,以及差异关系。故我们采用了和问题一中相同的斯皮尔曼相关性分析模型以及卡方检验模型来分析它们之间的相关性和差异性。

关键词: 层次聚类 Ward 方法 斯皮尔曼相关系数 Pearson 卡方检验 欧氏空间

一、问题重述

1.1 问题背景

古代玻璃极易受埋藏环境的影响而风化。在风化过程中，内部元素与环境元素进行大量交换，导致其成分比例发生变化，从而影响对其类别的正确判断。现有一批我国古代玻璃制品的相关数据，考古工作者依据这些文物样品的化学成分和其他检测手段已将其分为高钾玻璃和铅钡玻璃两种类型。

1.2 问题提出

- **问题一：**分析玻璃类型、纹饰和颜色与表面风化的关系；结合玻璃的类型，对表面有无风化化学成分含量进行统计分析，并根据风化点检测数据，预测其风化前的化学成分含量。
- **问题二：**对高钾玻璃、铅钡玻璃分别选择合适的化学成分对其进行亚类划分，并对分类结果的合理性和敏感性进行分析。
- **问题三：**对附件表单 3 中未知类别玻璃文物的化学成分进行分析并归类，并对分类结果的敏感性进行分析。
- **问题四：**对不同类别的玻璃间化学成分的关联进行分析，并比较不同类别之间的化学成分关联关系的差异性。

二、问题假设

1、因检测手段等原因可能导致其成分比例的累加和非 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；

2、数据中所有玻璃的各成分含量都具有其所属类型玻璃的特征。即：通过斯皮尔曼相关性检验不存在不可消除的异常值；

3、由于每一个玻璃都是独立的，在数量较大的情况下，消除极小数量的数据不完整的玻璃对应数据，对玻璃聚类等分析影响甚微。

三、符号说明

符号	含义说明	单位
ρ	斯皮尔曼相关系数	无
d_i	第 i 个数据对的位次值之差	位次
α_i	第 i 种文物纹饰	无
β_j	第 j 种文物类型	元素
γ_k	第 k 种文物颜色	无

j_i	实际频数	次数
jp_i	理论频数	次数
χ^2	卡方检验统计值	-
x_i	第 i 个风化前的化学变量	元素
x_i'	第 i 个风化后的化学变量	元素
\bar{x}	风化前所有化学变量的平均值	-
\bar{x}'	风化后所有化学变量的平均值	-
∂_i	第 i 个指标风化前后的占比	%
Δ_i	第 i 个风化前后指标的差值	-
$D_{avg}(c_i, c_j)$	第 i 个样本到第 j 个样本的距离的值的平均值	-
Γ	损失函数	-
$d_{\max(t)}(i, j)$	在簇中 i×j 个点中距离最远的两个点	dist
ω_{im}	编号为 i 的物品的第 m 种化学成分	元素
ω_{jm}	类型为 j 的物品的第 m 种化学成分	元素
$D_{ed}(i, j)$	欧氏距离函数	-
e_j	第 j 个聚类中心群的中心	-
SSE	误差平方和	-

四、问题分析

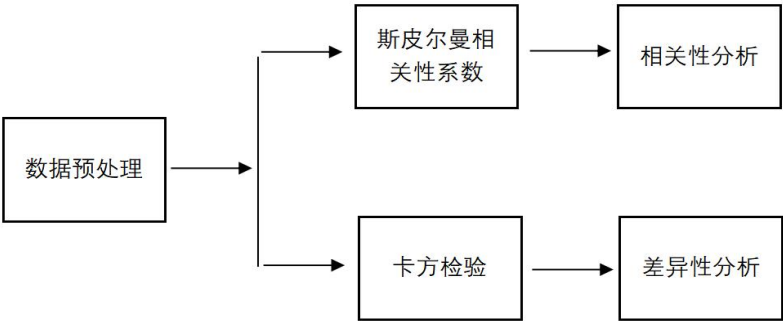
4.1 问题一的分析

该问主要对风化的相关因素进行分析。分别将三个变量与风化程度进行相关性、差异性分析。由于数据是离散型变量而非连续性变量，可以采用 Spearman correlation 分析相关性，同时采用 Pearson 卡方检验其差异性。

而后我们结合玻璃类型分析风化前后化学成分的规律。将表单一、二进行结合，将玻璃进一步分化为高钾未风化、高钾已风化、铅钡已风化、铅钡未风化四大类，求各类中成分元素均值，做出对比。

而预测风化点风化前的化学成分含量，依据玻璃类型，分别将已知风化和未

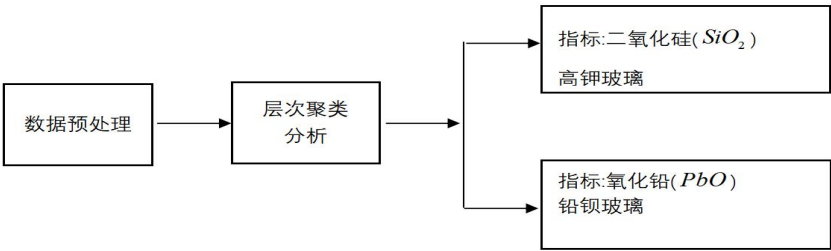
风化的样本各成分均值作为期望,做风化前后浮动比例。如若成分存在 0 的情况,则不适用比例法,进而采用差值法预测。



4.2 问题二的分析

对于玻璃类型分类的规律,即高钾和铅钡玻璃划分的规律,我们结合附件中数据,采用了控制变量法分别对纹饰、颜色和成分进行相应分析。

对于玻璃各类型的亚分类,我们用层次聚类分析法中的离差平方和方法(ward 方法)作逐步聚类。再以共性分类相关系数做亚分类合理性分析,以均值差的线性回归做亚分类敏感性分析。

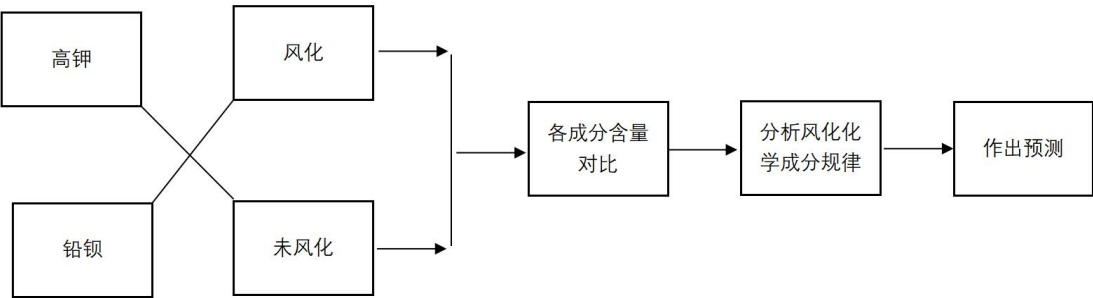


4.3 问题三的分析

在以上问题中,我们已经对各类型玻璃的成分规律做出分析。我们要对待检测文物进行分类可以将高钾已风化、高钾未风化、铅钡已风化、铅钡未风化的样品的各类成分均值作 14 元素,以此作欧氏空间,在欧式空间中确定 4 个大类中心点,再将待测文物对应的点与 4 个中心点作欧氏距离,以最小值确定其所属类别。

对于鉴定的敏感性分析,以测试点与其余四个中心点间的距离均值与标准差,和与所属类别中心点间距离做对比。以距离均值与最小值差值、方差来量化敏感性。差值越大、方差越大,敏感性越大。

如下图:



4.4 问题四的分析

分析不同类别的文物样品其化学成分之间的相关关系，类似于问题一，采用斯皮尔曼相关性分析模型。对相关系数、p 值作出分析，讨论其相关性。

比较不同类别之间文物样品其化学成分的差异关系，以卡方检验模型分析。

五、 问题一模型的建立与求解

5.1 问题一模型的建立

5.1.1 数据预处理

1.将表格中文本类型数据进行处理，对玻璃类型、纹饰和颜色及是否风化赋予对应的数据代号（如：纹饰 A 记为 1，纹饰 B 记为 2 等）。并依照类型、纹饰、颜色、是否风化的顺序导入矩阵中。

2.对于部分数据缺失，由于缺失量较少，对于表格 1 中缺失颜色的四行数据进行删除，不影响以下分析结果。

3.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，由此对于表单 2 中部分不达标数据进行删除处理。

4.对于风化与未分化成分分析统计，将表单 1 和表单 2 按照玻璃编号进行合并。

5.1.2 相关性分析（Spearman correlation）

由于数据是离散型变量而非连续性变量，可以采用 Spearman correlation 分析相关性。

斯皮尔曼相关系数是根据原始数据的等级排序进行求解，是两个变量之间变化趋势的方向以及程度，取值范围为[-1, 1]。当接近 1 时，表示两者具有强烈的正相关性；当接近-1 时，表示有强烈的负相关性；而值接近 0，则表示相关性很低。

由于数据来源于不同各类玻璃，可能出现异常值。而斯皮尔曼相关系数对于异常值不太敏感，因为它基于排序位次进行计算，实际数值之间的差异大小对于计算结果没有直接影响。

由斯皮尔曼相关系数定义，得到以下公式：

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

设纹饰为 $\alpha_i (i=1,2,3,4)$ ，记 A=1,B=2,C=3,D=4；设类型为 $\beta_j (j=1,2)$ ，记高钾=1,铅钡=2；设颜色为 $\gamma_k (k=1,2,\dots,8)$ ，记蓝绿=1,浅蓝=2,紫=3,深绿=4,深蓝=5,浅绿=6,黑=7,绿=8； d_i 表示第 i 个数据对的位次值之差；n 为总的观测样本数； ρ 为 Spearman 相关系数；设表面风化为 λ ， $\lambda=d(\alpha_i, \beta_j, \gamma_k)$ 。

得到斯皮尔曼的相关系数及其相应 P 值后便可进行下一步的分析检验。

5.1.3 差异性分析（卡方检验）

卡方检验：比较定类变量与定类变量之间的差异性，采用独立性检验：验证从两个变量抽出的配对观察值组是否互相独立。实际观测值与理论推断值之间的偏离程度就决定卡方值的大小，如果卡方值越大，二者偏差程度越大；反之，二者偏差越小；若两个值完全相等时，卡方值就为 0，表明理论值完全符合。

玻璃类型、纹饰和颜色及是否风化均为定类变量，故采用卡方检验。比较定类变量与定类变量之间的差异性，采用独立性检验：验证从两个变量抽出的配对观察值组是否互相独立。

设文物表面风化为理论值 E, 纹饰为观察值 O, 文物类型为观察值 S, 文物颜色为观察值 L. 实际频数为 j_i , 理论频数为 jp_i .

首先计算卡方检验的统计值 χ^2 ：把每一个观察值和理论值的差做平方后，除以理论值，再加总；然后计算 χ^2 统计值的自由度；最后我们依据设定的置信水准，查出自由度为 df 的卡方分配临界值，计算得出 χ^2 的统计值，如下：

$$\left\{ \begin{array}{l} \chi_1^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \\ \chi_2^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(S_{i,j} - E_{i,j})^2}{E_{i,j}} \\ \chi_3^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(L_{i,j} - E_{i,j})^2}{E_{i,j}} \end{array} \right.$$
$$E_{i,j} = j \times p_i (j = 1, \dots, c; i = 1, \dots, r)$$
$$P_i = P(\zeta \in A_i) = P(a_{i-1} \leq \zeta \leq a_i) = F_0(a_i) - F_0(a_{i-1})$$

由此，得到卡方值及其 p 值后，便可进行下一步的卡方检验分析。

5.2 问题一模型的求解

5.2.1 分析玻璃类型、纹饰和颜色与表面风化的关系

首先，我们进行数据预处理。对于表单一中颜色缺失的玻璃进行数据删除而后对纹饰、类型、颜色和表面是否风化的文本信息改写为数值形式，将此数据导入 matlab 中形成矩阵。矩阵的第一列到第五列分别对应文物编号、纹饰、类型、颜色、表面是否风化。然后对数据进行相关性分析和差异性分析。

利用 matlab 我们可以求解出文物样品表面风化与纹饰、类型、颜色之间的 Spearman 相关系数以及对应的 p 值。在分析过程中，我们首先对 p 值进行检验：若 p 值小于 0.05，则证明该相关系数在 5%显著水平下显著，然后进行对相关系数进行分析；若 p 值大于 0.05，则证明该相关系数在 5%显著水平下不显著，故相关系数无效，即二者不存在相关性。上述三组的相关系数及 p 值详见表 1.

由此，通过分析我们得出如下结论：

斯皮尔曼相关系数检验的结果显示，基于表面风化和纹饰相关系数 r 为 0.0479， p 值为 0.7311、基于表面风化和颜色相关系数 r 为 -0.1119， P 值为 0.4206 因此对于表面风化和纹饰、颜色在 5%显著水平下不显著，不存在相关性；同时表面风化和类型相关系数 r 为 0.3162， p 为 0.0198。即对于表面风化和类型在 5%显著水平下显著，具有相关性。

表 1 Spearman 相关系数表

因素 1	因素 2	Spearman 相关系数	p 值
表面风化	纹饰	0.0479	0.7311
	类型	0.3162**	0.0198
	颜色	-0.1119	0.4206

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。

由于样本数量较小，我们除了进行相关性分析，同时进行了卡方检验差异性分析，以此补足可能出现的系统误差。

在卡方检验中，我们针对表面有无风化下纹饰不同、类型不同、颜色不同的样本数量进行了统计，然后分别针对有无风化与纹饰 ABC、有无风化与高钾铅钼类型、有无风化与颜色进行了卡方检验，利用 matlab 求解出了其卡方值和 p 值。若 p 值小于 0.05，则说明其在 5%显著水平下显著，该组因素具有显著性差异，若 p 值大于 0.05，则说明其在 5%显著水平下不显著，即该组因素之间不具有显著性差异。每组因素之间的卡方值、 p 值及相关统计详见表 2。

卡方检验的结果显示，基于表面风化和纹饰显著性 p 值为 0.0565、基于表面风化和颜色显著性 P 值为 0.5066，故在 5%显著水平下不存在显著性差异（表面风化和纹饰在 10%显著水平下存在显著性差异），因此对于表面风化和纹饰、颜色不存在显著性差异；同时表面风化和类型 p 值为 0.0201，即对于表面风化和类型数据在 5%显著水平下存在显著性差异。

表 2 卡方检验表

因素	分类	表面风化		χ^2	p 值
		无风化	风化		
纹饰	A	11	9	5.7471*	0.0565
	B	0	6		
	C	13	15		
类型	高钾	12	6	5.4000**	0.0201
	铅钼	12	24		
	蓝绿	6	9		
	浅蓝	8	12		
颜色	紫	2	2	6.2871	0.5066
	深绿	3	4		
	深蓝	2	0		
	浅绿	2	1		

黑	0	2
绿	1	0

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。

综上，卡方分布检验与斯皮尔曼相关系数检验结论一致：表面风化和类型有较强相关性；与纹饰、颜色无相关性。

5.2.2 结合玻璃类型分析风化化学成分规律

首先我们对表二中数据进行处理(详见 5.1.1)，而后将表单一、表单二进行合并。题目要求按照玻璃类型对是否风化的成分类型做规律分析，由此结合表单一，表单二作出新的一列指标——【表面风化 2.0】，在此指标中，对于在表一中样本整体风化而表二中部分点或部位未风化的玻璃，我们将其归类为未风化。由此产生的新的指标。

再结合玻璃的类型进一步划分，由此我们得到高钾未风化、高钾已风化、铅钡已风化、铅钡未风化四大类，并求解出在每一类数据中每个化学成分的均值及标准差，详见表 3。

表 3 化学成分描述性统计表

化学成分	类型							
	高钾				铅钡			
	风化		无风化		风化		无风化	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差
二氧化硅	93.9633	1.7336	67.9842	8.7551	24.9127	10.6055	54.6596	11.8286
氧化钠	0	0	0.6950	1.2869	0.2162	0.5566	1.6826	2.3716
氧化钾	0.5433	0.4452	9.3308	3.9203	0.1335	0.2400	0.2187	0.3101
氧化钙	0.8700	0.4878	5.3325	3.0925	2.6954	1.6598	1.3204	1.2847
氧化镁	0.1967	0.3063	1.0792	0.6761	0.6500	0.7064	0.6404	0.5468
氧化铝	1.9300	0.9645	6.6200	2.4915	2.9700	2.6343	4.4561	3.2624
氧化铁	0.2650	0.0695	1.9317	1.6667	0.5846	0.7365	0.7365	1.1547
氧化铜	1.5617	0.9348	2.4525	1.6600	2.2758	2.8205	1.4317	1.9699
氧化铅	0	0	0.4117	0.5890	43.3138	12.2302	22.0848	8.2151
氧化钡	0	0	0.5983	0.9821	11.8073	9.9783	9.0017	5.8253
五氧化二磷	0.2800	0.2100	1.4025	1.4340	5.2773	4.1967	1.0491	1.8471
氧化锶	0	0	0.0417	0.0484	0.4185	0.2648	0.2683	0.2435
氧化锡	0	0	0.1967	0.6813	0.0685	0.2695	0.0465	0.1273
二氧化硫	0	0	0.1017	0.1855	1.3662	4.2061	0.1591	0.7632

通过对数据进行整理与分析，我们得到如下结论。

对于高钾玻璃，二氧化硅含量在风化后大幅提升至 90%以上；氧化钾含量在风化后由约 10%降至 1%以下；此外氧化钠、氧化铅、氧化锶、氧化钡、氧化锡、二氧化硫在风化后含量为 0，其余成分含量均有所下降；

对于铅钡玻璃，二氧化硅含量在风化后由 50%以上下降至 30%左右；氧化铅

含量在风化后由约 20%增长至 40%左右；其余成分略有浮动，但各成分含量均大于 0。

5.2.3 预测据风化点风化前的化学成分含量

为预测风化点风化前的化学成分含量，我们对风化前后的化学成分进行对比分析，结合表三所得数据，我们依照玻璃类型分别分析。以各成分均值作为期望值作浮动比例。我们首先对高钾玻璃进行分析，对于其风化前后各化学成分求出均值，然后对以均值求变化比例。

由于在高钾玻璃中有许多成分在分化后含量为零，我们无法使用比例做预测。由于此类成分物质在未风化前含量占比同样较低，我们采用风化前后的差值做均值作为风化前此成分含量的期望。

然后对预测数据进行进一步的筛选。首先对成分含量做求和。如若符合各成分总含量在 85%~105%即符合预测要求，完成预测。如若不符合要求则对数据进行剔除。铅钡玻璃同理。

设风化前的 14 个化学变量二氧化硅、氧化钠、氧化钾、...、二氧化硫为 $x_1, x_2, x_3, \dots, x_{14}$ ，风化后的 14 个化学变量二氧化硅、氧化钠、氧化钾、...、二氧化硫为 $x'_1, x'_2, x'_3, \dots, x'_{14}$ 。

生成数据列： $X = (x_1, x_2, x_3, \dots, x_{14})$ $X' = (x'_1, x'_2, x'_3, \dots, x'_{14})$

得到风化前和风化后的化学变量的平均值

$$\begin{cases} \bar{x} = \frac{\sum_{i=1}^{14} x_i}{14} \\ \bar{x}' = \frac{\sum_{i=1}^{14} x'_i}{14} \end{cases}$$

设 ∂_i 为第 i 个指标风化前后的占比，则： $\partial_i = \frac{\bar{x}_i}{x_i} (i = 1, 2, \dots, 14)$

由于文物类型为高钾的部分化学成分如氧化钠、氧化钡、五氧化二磷、二氧化硫等的均值为 0， ∂_i 无法得出预测结果，此种情况可用均值的差值 Δ_i 进行预测，

则： $\Delta_i = \bar{x}_i - \bar{x}'_i (i = 1, 2, \dots, 14)$

由于预测的数据数目较多且化学成分较多，本文中只对部分样本的重要化学成分进行罗列，见表 4。

表 4 部分预测数据

样本编号	二氧化硅 (SiO ₂)	...	氧化铁 (Fe ₂ O ₃)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	...	二氧化硫 (SO ₂)
9	68.75	...	2.33	2.43	0.41	0.60	...	0.10
10	70.01	...	1.90	1.32	0.41	0.60	...	0.10

12	68.22	...	2.11	2.59	0.41	0.60	...	0.10
27	67.08	...	1.46	2.42	0.41	0.60	...	0.10
8	44.19	...	0.00	6.55	14.62	23.81	...	0.30
19	65.03	...	1.68	2.21	21.83	4.08	...	0.00
26	43.42	...	0.00	6.65	15.06	24.59	...	0.23

注：具体预测数值详见代码结果中 **Predicts1** 和 **Predicts2** 或附录——全部预测数据。

六、问题二模型的建立与求解

6.1 问题二模型的建立

6.1.1 数据预处理（控制变量筛选）

1. 合并表单 1 和表单 2。

2. 依据玻璃类型、是否风化、纹饰作自变量（固定变量），对其颜色和各成分含量进行筛选处理。

6.1.2 层次聚类法建模

对于不同类型的玻璃进行亚分类，我们可以采取层次聚类分析法层次聚类具有不需要预先制定聚类数的优点，同时可以发现类的层次关系，并且可以聚类形成其他形状。这有利于我们玻璃较为复杂的数据。

$$D_{avg}(c_i, c_j) = \frac{1}{|c_i|} \frac{1}{|c_j|} \sum_{x_1 \in c_i} \sum_{x_2 \in c_j} Dist(\sigma_1, \sigma_2)$$

平均连接：

$$\Gamma = loss(M) = \sum_{i=1}^{67} \sum_{j \in M_i} (K_i - j)^2$$

损失函数：

M_i 表示簇中样本的集合， K_i 表示簇 M_i 的中心点

将 67 个样本（有 58 件文物采样，某些文物采样有部位 1 和部位 2）初始化为一个簇， $c_1 = \{\sigma_i\}, i=1, 2, \dots, 67$ ；计算样本两两间的距离

$d(i, j), i=1, 2, \dots, 67; j=1, 2, \dots, 67; i \neq j$ ；再计算当前所有簇 $c_i, i \in 1, 2, \dots, n$ 的损失函数

Γ ，选择损失最大的簇进行二分；设该损失最大的簇为 $(c_i)_{max}$ ，计算该簇下两

点间距离 $d_{max(i)}(i, j), i=1, 2, \dots, n_i, j=1, 2, \dots, n_j$ ，选择簇中最远的两个点作为类中心将簇二分，直到将簇分到 67 个。

6.2 问题二模型的求解

6.2.1 玻璃类型分类规律

结合 6.1.1 数据分析，见下表 5。

表 5 玻璃类型分类规律表

类型	是否风化	纹饰	颜色	二氧化硅 (SiO ₂)	氧化钾 (K ₂ O)	氧化铅 (PbO)
高钾	未风化	A、C	多为1，且有2、4、5	60%~80%	约10%	0
	已风化	B	1	90%以上	1%以下	0
铅钡	未风化	A、C	2、3、4、5、6、8	约55%	约0.21%	约20%
	已风化	A、C	多为2，同时有1、3、4、6、7	约25%	约0.13%	约40%

注：【1 蓝绿】【2 浅蓝】【3 紫】【4 深绿】【5 深蓝】【6 浅绿】【7 黑】【8 绿】

结合表中数据进行分析，可以得出以下几条明显分类规律：

- (1) 纹饰 B 出现时，必为高钾已风化玻璃；
- (2) 颜色为【1 蓝绿】，大概率为高钾玻璃；
- (4) 颜色为【3 紫】、【6 浅绿】、【7 黑】、【8 绿】，大概率为铅钡玻璃；
- (5) 二氧化硅含量大于 70%，大概率为高钾玻璃；
- (6) 化钾含量大于 1%，大概率为高钾玻璃；
- (7) 氧化铅含量不为 0，为铅钡玻璃。

6.2.2 玻璃类型亚分类（层次聚类分析）

对于不同类型的玻璃进行亚分类，我们采取层次聚类分析法。本文中，我们利用 matlab 对筛选后的数据进行层次聚类，其中在以“距离”为指标进行分类时，为了提高准确性和合理性，我们采用了离差平方和方法（Ward 方法）分别对铅钡类样本高钾类样本进行聚类分析，得到的系统聚类树如图 1 图 2。

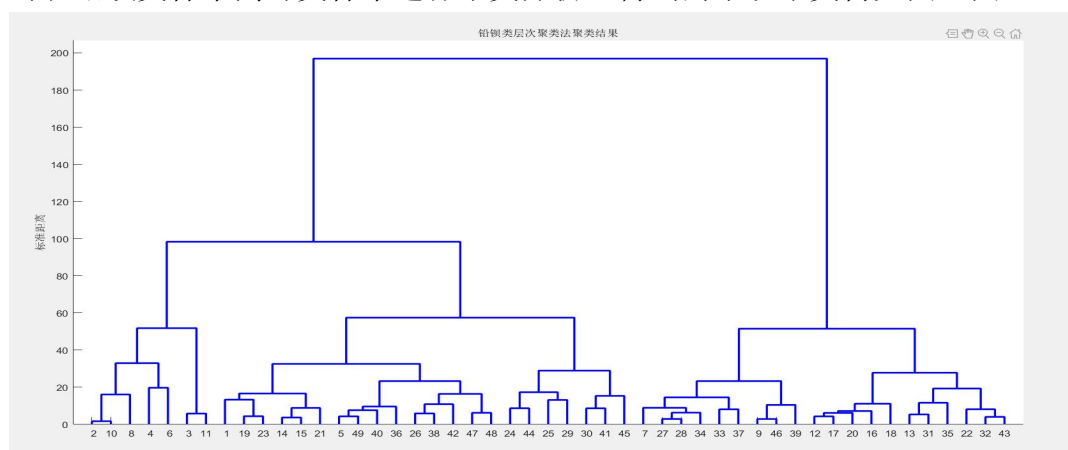


图 1 铅钡类系统聚类树

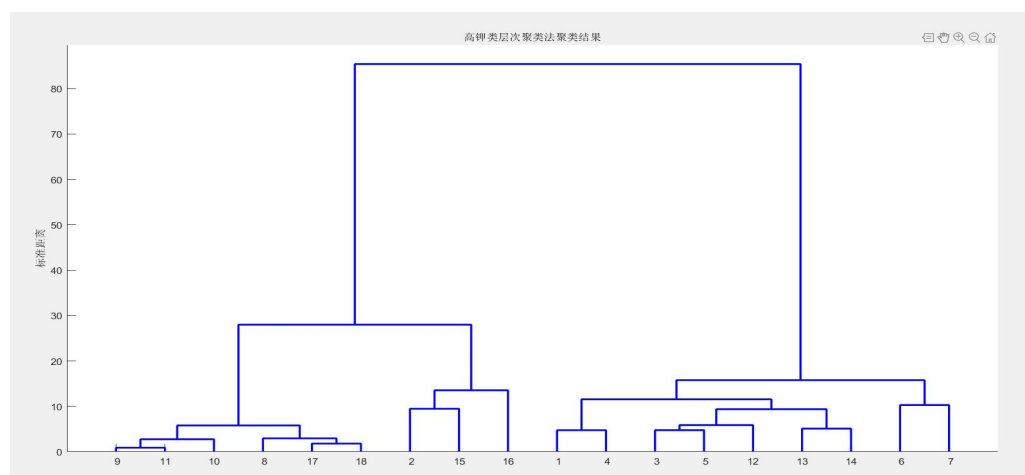


图 2 高钾类系统聚类树

由于在铅钡玻璃中，氧化铅含量较大且具有差异性，所以对于铅钡玻璃的分类，我们采用氧化铅作为指标进行层次聚类分析。我们在使用层次聚类的时候可以依照指标在不同类中差异是否明显来判定此分类是否合理。步骤如下：

（1）我们首先以 $n=6$ （ n 指的是分类数）进行聚类分析。在这 6 个类中求取各类中[指标]（氧化铅含量）均值。由于在第 5 类和第 6 类中[指标]差异并不明显，所以我们继续进行聚类。

（2）在聚类 $n=5$ 的时候，我们重复以上步骤。如若在某两类中指标（氧化铅含量均值）差异依旧不明显，我们则继续进行聚类。

（3）重复以上步骤，直至各类中指标差异明显。

以此类推，我们得到 n 等于 3 的时候，有铅钡玻璃的亚分类如下表 6。

表 6 铅钡玻璃亚分类表

	铅钡玻璃编号										指标（氧化铅）含量
分类 1	2	10	8	4	6	3	11				26.3443
分类 2	1	19	23	14	15	21	5	49	40	36	47.3382
	26	38	42	47	48	24	44	25	29	30	
					41	45					
分类 3	7	27	28	34	33	37	9	46	39	12	20.4130
	17	20	16	18	13	31	35	22	32	43	

而对于高钾玻璃，二氧化硅的含量占比较大且各块玻璃之间二氧化硅含量差异较为明显，所以使用聚类后二氧化硅在各类中均值做指标。具体方法与铅钡玻璃同理。亚分类如下表 7。

表 7 高钾玻璃亚分类表

	高钾玻璃编号									指标（二氧化硅）含量
分类 1	9	11	10	8	17	18	2	15	16	89.6633
分类 2	1	4	3	5	12	13	14	7	6	63.6244

6.2.3 共性分类相关系数——亚分类合理性

此外，在进行层次聚类分析的过程中，我们针对两次聚类利用 matlab 求解出 C ——共性分类相关系数。共性分类相关系数描述了由共性分类距离组成的矩阵与原相异度矩阵之间的相似性，相似性越高，说明层次聚类的性能越好（即： C 越接近 1，分类越合理）。由此我们以共性分类相关技术来度量此分类的合理性。

对于我们上述分类方式，得出高钾玻璃亚分类 $C1 = 0.8877$ ，铅钡玻璃亚分类 $C2=0.7862$ ，接近于 1，说明此层次聚类合理。

6.2.4 均值差的线性回归——亚分类敏感性

为了对亚分类的敏感性进行分析，我们希望通过考察随着上述的不同分类的类数（即 n ）的波动，各个亚类别之间的化学成分有显著差异的程度来进行敏感性的分析。

首先，对数据划分类，而后对每一大类的成分含量的所有均值，并对每两大类之间的均值做差。其次，对所得差值作线性回归得到 F 检验的 p 值，进而通过分析 p 值来判断均值差是否显著为 0（判断差值的平均大小）如果得到的 p 值小于 0.05（即无限接近于 0），我们就认为这两大类分类的化学成分的均值差显著为 0，即这两类是同一类，这边说明这种分类方法不够合理准确，若 p 值越大，说明两类别之间化学成分的差异就越大，也就是说这种分法相对更加准确。对于每种分类方式，类别中两两类的 p 值如表 8。

表 8 类型数量及 p 值

	类型数量				max
	n=6	n=5	n=4	n=3	
两两间 p 值	0.9386	0.9386	0.9697	0.9778	
	0.9891	0.9891	0.9909	0.9838	
	0.9956	0.9956	0.9838	0.9674	
	0.9762	0.9658	0.9757		
	0.9604	0.9459	0.9587		
	0.9459	0.9617			
	0.9617	0.9899			
	0.9790	0.9757			
	0.9971	0.9587			
	0.9757	0.9803			
	0.9652				
	0.9554				
	0.9885				
	0.9753				
	0.9401				
均值	0.9696	0.9701	0.9758	0.9763	0.9763

由表可知，在 n 从 6 逐渐减少至 3 中，p 均值增大，意味着分类间相关性越来越低，即分类的敏感性较高。

七、问题三模型的建立与求解

7.1 问题三模型的建立

7.1.1 最小欧式距离分析

在问题一、问题二中，我们已经对各类型玻璃的成分规律做出分析。我们要对待检测文物进行分类可以将高钾已风化、高钾未风化、铅钡已风化、铅钡未风化的样品的各类成分均值作 14 元素，以此作 14 维欧氏空间。

输入编号为 A1,A2,A3,...,A8 的 8 件文物的 14 种化学成分作数据样本，记为 $\omega_m (i=1,2,...,8,m=1,2,...,14)$ ，问题二得出不同类型文物的 14 种化学成分的特征指标作聚类中心，设高钾未风化=1，高钾风化=2，铅钡未风化=3，铅钡风化=4，记为 $\omega_{jm} (j=1,2,3,4,m=1,2,...,14)$

欧式距离函数：

$$D_{ed}(i,j)=\sqrt{(\omega_{i1}-\omega_{j1})^2+(\omega_{i2}-\omega_{j2})^2+...+(\omega_{im}-\omega_{jm})^2},\ s.t.\min(SSE)$$

特征 $\omega_i=(\omega_{i1},\omega_{i2},...,\omega_{im})$, 特征 $\omega_j=(\omega_{j1}+\omega_{j2},...,\omega_{jm})$

误差平方和;
$$SSE=\sum_{x\in E_1}dist(e_1,\varpi)^2+\sum_{x\in E_2}dist(e_2,\varpi)^2+...+\sum_{x\in E_j}dist(e_j,\varpi)^2$$

E_j 为第 j 个聚类中心群, e_j 为第 j 个聚类中心群的中心, ϖ 为聚类的数据样本 (j=1,2,3,4)

用欧式距离去度量样本到聚类中心的距离, 并把误差平方和 SSE 作为度量聚类效果的目标函数, 选取误差平方和最小的分类结果作为最终的聚类结果。

作为欧式空间上的四个点分别表示着分别代表着这四大类的中心点。而我们对于待检测文物可以用这 14 个元素作为他的坐标, 去欧式空间上做相应点与已知四大类的中心点的距离。

其中最小距离的中心点对应分类意味着这个待检测文物更加符合这一分类。

7.2 问题三模型的求解

7.2.1 预测玻璃类型

根据前面问题的分析, 为了对待测文物更准确的归类, 我们将已知样本分为高钾风化、高钾为风化、铅钡风化、铅钡未风化四大类, 并且将这四大类的各化学成分均值组成的点作为中心点进行分析。我们利用 matlab 将每一个待测文物到每一个类中心都进行了距离的计算, 每一个待测文物到各个类中心的距离见表 9。

表 9 待测物到类中心距离表

待检测 文物编号	类型（是否风化）				所属类别
	高钾已风化	高钾未风化	铅钡已风化	铅钡未风化	
A1	17.4265	14.1293	70.2426	34.2361	高钾已风化
A2	67.6820	48.4835	22.3415	27.2264	铅钡已风化
A3	74.3656	54.4404	13.8104	30.4232	铅钡已风化
A4	64.9367	42.8760	23.4851	21.7675	铅钡未风化
A5	34.0507	17.1538	52.3478	17.6045	铅钡未风化
A6	1.2388	27.4330	81.9190	45.4567	高钾已风化
A7	4.4828	24.9036	79.9889	43.3881	高钾已风化
A8	49.7755	31.7737	35.2113	9.4949	铅钡未风化

通过分析, 我们将距离每个待测文物最近的类别作为该待测文物的分类。为了检验分类的准确性, 我们又利用 Spearman 相关性分析法对待测文物和其所归到的类别之间进行了相关性分析与检验, 得到的相关系数和 p 值详见表 10。

表 10 相关性分析表

待检测文物 编号	所属类别	相关系数	p 值
-------------	------	------	-----

A1	高钾已风化	0.6091**	0.0208
A2	铅钡已风化	0.7556***	0.0018
A3	铅钡已风化	0.7174***	0.0039
A4	铅钡未风化	0.7395***	0.0025
A5	铅钡未风化	0.7846*** (0.4330)	0.0014 (0.1239)
A6	高钾已风化	0.9821***	0.0000
A7	高钾已风化	0.9369***	0.0000
A8	铅钡未风化	0.6467**	0.0124

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。

通过分析我们注意到，在待检测文物 A5 中，它所代表的欧式空间上的点与高钾未风化和铅钡未风化中心点的距离分别是 17.1538、17.6045，差距非常小。取最小值应将 A5 归类为高钾未风化。但是分别对 A5 与铅钡未风化和高钾未风化进行相关性分析，高钾未风化中与 A5 的 p 值是 0.1239，相关系数是 0.4330（如图表中括号内记为 A5 与高钾未风化之间相关性检验的结果）。无论是相关系数的值还是 p 值，都无法通过相关性检验。而 A5 与铅钡未风化中心点的 p 值为 0.0014 相关系数为 0.7846，证明了 A5 样本与铅钡未风化具有很强的相关性。由此我们得出 **A5 是属于铅钡未风化而非高钾未风化**。

7.2.2 敏感性分析

关于对未检测文物分类的敏感性分析，我们通过求出待测样本与其余四个类中心点间的距离、求均值与标准差，并求与所属类别中心点间距离，详见表 11。

表 11 待测文物与类中心点距离均值、标准差及与所属类中心点距离

待检测文物编号	与所属类别中心距离	与四中心点距离均值	与四中心点距离方差
A1	14.1293	34.0086	25.7106
A2	22.3415	41.4333	20.8569
A3	13.8104	43.2599	26.6122
A4	21.7675	38.2663	20.1928
A5	17.6045	30.2892	16.6751
A6	1.2388	39.0119	33.8794
A7	4.4828	38.1908	32.0774
A8	9.4949	31.5638	16.6539

通过以上数据分析，易知检测点与所属类别中心点距离远小于与四中心点间距离均值且与四点间距离的标准差较大。这意味着测试点距离不同分类中心的距离具有较大波动性，即这样的归类具有较高的敏感性。

八、问题四模型的建立与求解

8.1 问题四模型的建立

8.1.1 斯皮尔曼相关性分析模型和卡方检验模型

问题四要求分析对于不同类别的文物样品其化学成分之间的相关关系，以及不同类别之间的化学成分的差异关系。故在问题四中我们采用了和问题一中相同的斯皮尔曼相关性分析模型以及卡方检验模型来分析它们之间的相关性和差异性。（详见 5.1.2 和 5.1.3）

8.2 问题四的求解

8.2.1 斯皮尔曼相关性分析模型

基于前面对问题的求解，为了方便后面的相关性分析，我们将已知样本数据分成了四个类别：高钾风化、高钾未风化、铅钡风化、铅钡未风化。

利用斯皮尔曼相关性分析的方法，我们对于四种类别的文物样品分别进行了 14 种化学成分之间的相关性分析，并且分别得到了相关性系数矩阵以及 p 值矩阵。（对于高钾风化类的文物样品的化学成分，我们发现它们的氧化钠、氧化铅、氧化钡、氧化锶、氧化锡、二氧化硫的含量极少（均值为 0），故在此类别的分析中忽略上述六种元素的相关性分析。）

对于具有的显著性的相关系数进行分析，通过显著性检验的相关系数及其 p 值详见表 12（由于数据量较大，仅显示部分具有相关性的数据，详见附录-问题四相关性分析）。

表 12 文物类别中化学成分相关性分析表

类别	化学成分 1	化学成分 2	Spearman 相关系数	p 值
高钾风化	二氧化硅	氧化钙	-1***	0.0028
高钾风化	二氧化硅	氧化铝	-0.9429**	0.0167
高钾风化	氧化钾	五氧化二磷	-0.9276**	0.0222
高钾风化	氧化钙	氧化铝	0.9429**	0.0167
高钾未风化	二氧化硅	氧化钾	-0.6224**	0.0307
			
铅钡未风化	氧化钾	氧化铝	0.6182***	0.0017
铅钡未风化	氧化镁	氧化铝	0.5431***	0.0074
铅钡未风化	氧化铜	氧化钡	0.5515***	0.0064
铅钡未风化	氧化钙	氧化锡	0.4958**	0.0161
铅钡未风化	氧化铁	氧化锡	0.4544**	0.0294

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。

经过分析，我们可以得出结论：

在高钾风化样本中，二氧化硅与氧化钙、二氧化硅与氧化铝以及氧化钾与五氧化二磷的含量成负相关关系；氧化钙与氧化铝的含量成正相关关系。

在高钾未风化样本中，二氧化硅与氧化钾、二氧化硅与氧化钙、氧化钠与氧化镁、二氧化硅与氧化铁、氧化铝与二氧化硫的含量成负相关关系；氧化钠与氧化钾、氧化钠与氧化钙、氧化钾与氧化钙、氧化铝与氧化铁、氧化铁与氧化铜、氧化铝与氧化钡、氧化镁与氧化锶的含量成正相关关系。

在铅钡风化样本中，氧化钠与氧化铜、氧化镁与氧化钡、氧化铝与氧化钡、

氧化铁与氧化钡、氧化铅与氧化钡、二氧化硅与氧化锶、二氧化硅与二氧化硫、氧化镁与二氧化硫、氧化铝与二氧化硫、氧化铁与二氧化硫、氧化铅与二氧化硫的含量成**负相关关系**；二氧化硅与氧化钠、氧化钙与氧化镁、二氧化硅与氧化铝、氧化镁与氧化铝、氧化钙与氧化铁、氧化镁与氧化铁、氧化铝与氧化铁、氧化铜与氧化钡、氧化钙与五氧化二磷、氧化镁与五氧化二磷、氧化铝与五氧化二磷、氧化铝与氧化锡、氧化铜与二氧化硫、氧化钡与二氧化硫的含量成**正相关关系**。

在**铅钡未风化**样本中，氧化钠与氧化铁、氧化钙与氧化铜、氧化铁与氧化铜、二氧化硅与氧化铅、氧化铝与氧化铅、二氧化硅与氧化钡、氧化钠与五氧化二磷、氧化铜与氧化锡的含量成**负相关关系**；氧化钾与氧化镁、氧化钾与氧化铝、氧化镁与氧化铝、氧化铜与氯化钡、氧化钙与氧化锡、氧化铁与氧化锡的含量成**正相关关系**。

8.2.2 分析差异性——卡方检验

在差异性分析的过程中，我们针对高钾风化与铅钡风化样本数据的均值、高钾无风化与铅钡无风化样本数据的均值、高钾风化与高钾无风化样本数据的均值、铅钡风化与铅钡无风化样本数据的均值分别进行了不同类别间元素的卡方检验。得到的 p 值以及卡方值详见表 13。

表 13 文物类别间化学成分差异性卡方检验

样本类别 1	样本类别 2	p 值	χ^2
高钾风化	铅钡风化	3.3307e-16***	103.6719
高钾无风化	铅钡无风化	4.7202e-05***	42.8675
高钾风化	高钾无风化	0.0464**	22.6286
铅钡风化	铅钡无风化	0.0282**	24.3338

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。

我们可以看出，以上四组卡方检验的 p 值均在 5%显著水平下显著，故可以得出结论：高钾风化样本与铅钡风化样本之间、高钾无风化样本与铅钡无风化样本之间、高钾风化样本与高钾无风化样本之间、铅钡风化样本与铅钡无风化样本之间的化学成分均存在着显著性差异。

九、模型的评价与推广

9.1 模型的评价

首先处理了附件数据，将文本数据量化，使数据更直观，易于分析。分析数据相关性，采用斯皮尔曼相关系数，极大程度的排除异常值干扰。对比数据差异性，采用卡方检验对 p 值作分析，简便清晰地呈现差异性情况。

对于亚分类，采取层次聚类分析法——离差平方和方法（Ward 方法），不需要预先制定聚类数的优点，同时可以发现类的层次关系，可视化强。

以各大类均值差作线性回归，根据所得 p 值再求均值。在分类数 n 的改变下， p 值均值随之改变，以 p 均值量化分类敏感性，使得敏感性分析更直观。

9.2 模型的改进

在问题求解中，对四大类型玻璃的各成分含量取期望，我们采取的是取算术

平均。但其实在每一类内部，数据间存在差距，且数据量较小，得出期望与实际情况中可能存在误差。

相较于实际情况，如若结合更多因素，采用某一范围，而非某一定值作为期望，会使模型更贴近于现实。

十、参考文献

- [1] 天工在线 . MATLAB 2020 从入门到精通 实战案例版[M].中国水利水电出版社, 2020.
- [2] 谢金星 , 薛毅 . 优化建模与 LINDO/LINGO 软件 [M]. 北京 : 清华大学出版社 ,2005.
- [3] 李德新.聚类分析模型算法研究及其有效性评价[J].电脑知识与技术 ,2021,17(30):26-27.DOI:10.14004/j.cnki.ckt.2021.2864.
- [4] 瞿俊,姜青山,董槐林. 基于高斯混合模型的层次聚类算法[C]//.第二十三届中国数据库学术会议论文集（研究报告篇） .,2006:332-338.
- [5]方玺,葛权耕,何朗.基于叶脉聚类分析技术的叶片快速分类[J].湖北农业科学,2013,52(14):3423-3427.DOI:10.14088/j.cnki.issn0439-8114.2013.14.002.
- [6]丛占凤,苏利英,彭永宏,王建兵.惠州西湖风景名胜区红花湖绿道植被景观特征[J].福建林业科技,2015.

附录

支撑材料文件列表

文件名称	文件类型	作用
chi2test	MATLAB Code	函数
data1	MATLAB Data	数据
solution	MATLAB Code	主程序
附件	XLSX 工作表	数据

MATLAB 主程序

(一) 第一题代码

```
clear,clc
load data1%其中 ifm 为附件表单 1 中数据 ifm2 为附件表单 2 中数据 ifm3 为附件表单 3 中数据 均
为矩阵
num=numel(ifm(:,1));
rows=[];
for i=1:num
    if ifm(i,4)==0
        rows=[rows,i];
    end
end
ifm(rows,:)=[];%将颜色为空的样本数据剔除
num1=numel(ifm(:,1));

%1.1
a=[];
for i=1:num1
    if ifm(i,5)==0
        a=[a;[ifm(i,2) ifm(i,5)]]; %无风化纹饰
    end
end
anum=numel(a(:,1));
a1=[];a2=[];a3=[];
for i=1:anum
    if a(i,1)==1
        a1=[a1;a(i,:)]; %无风化纹饰 A
    else
        a3=[a3;a(i,:)]; %无风化纹饰 C
    end
end
alnum=numel(a1(:,1));
```

```

a3num=numel(a3(:,1));

b=[];
for i=1:num1
    if ifm(i,5)==1
        b=[b;[ifm(i,2) ifm(i,5)]]; %风化纹饰
    end
end
bnum=numel(b(:,1));
b1=[];b2=[];b3=[];
for i=1:bnum
    if b(i,1)==1
        b1=[b1;b(i,:)]; %风化纹饰 A
    elseif b(i,1)==2
        b2=[b2;b(i,:)]; %风化纹饰 B
    else
        b3=[b3;b(i,:)]; %风化纹饰 C
    end
end
b1num=numel(b1(:,1));
b2num=numel(b2(:,1));
b3num=numel(b3(:,1));

c=[];
for i=1:num1
    if ifm(i,5)==0
        c=[c;[ifm(i,3) ifm(i,5)]]; %无风化类型
    end
end
cnum=numel(c(:,1));
c1=[];c2=[];
for i=1:cnum
    if c(i,1)==1
        c1=[c1;c(i,:)]; %无风化高钾
    else
        c2=[c2;c(i,:)]; %无风化铅钡
    end
end
c1num=numel(c1(:,1));
c2num=numel(c2(:,1));

d=[];
for i=1:num1
    if ifm(i,5)==1

```

```

        d=[d;[ifm(i,3) ifm(i,5)]]; %风化类型
    end
end
dnum=numel(d(:,1));
d1=[];d2=[];
for i=1:dnum
    if d(i,1)==1
        d1=[d1;d(i,:)]; %风化高钾
    else
        d2=[d2;d(i,:)]; %风化铅钨
    end
end
d1num=numel(d1(:,1));
d2num=numel(d2(:,1));

d=[];
for i=1:num1
    if ifm(i,5)==1
        d=[d;[ifm(i,3) ifm(i,5)]]; %风化类型
    end
end
dnum=numel(d(:,1));
d1=[];d2=[];
for i=1:dnum
    if d(i,1)==1
        d1=[d1;d(i,:)]; %风化高钾
    else
        d2=[d2;d(i,:)]; %风化铅钨
    end
end
d1num=numel(d1(:,1));
d2num=numel(d2(:,1));

e=[];
for i=1:num1
    if ifm(i,5)==0
        e=[e;[ifm(i,4) ifm(i,5)]]; %无风化类型
    end
end
enum=numel(e(:,1));

e1=[];e2=[];e3=[];e4=[];e5=[];e6=[];e7=[];e8=[];
for i=1:enum
    if e(i,1)==1

```

```

        e1=[e1;e(i,:)];%无风化颜色 1 蓝绿
    elseif e(i,1)==2
        e2=[e2;e(i,:)];%无风化颜色 2 浅蓝
    elseif e(i,1)==3
        e3=[e3;e(i,:)];%无风化颜色 3 紫
    elseif e(i,1)==4
        e4=[e4;e(i,:)];%无风化颜色 4 深绿
    elseif e(i,1)==5
        e5=[e5;e(i,:)];%无风化颜色 5 深蓝
    elseif e(i,1)==6
        e6=[e6;e(i,:)];%无风化颜色 6 浅绿
    else
        e8=[e8;e(i,:)];%无风化颜色 8 绿
    end
end
e1num=numel(e1(:,1));
e2num=numel(e2(:,1));
e3num=numel(e3(:,1));
e4num=numel(e4(:,1));
e5num=numel(e5(:,1));
e6num=numel(e6(:,1));
e8num=numel(e8(:,1));

f=[];
for i=1:num1
    if ifm(i,5)==1
        f=[f;[ifm(i,4) ifm(i,5)]]; %风化类型
    end
end
fnum=numel(f(:,1));

f1=[];f2=[];f3=[];f4=[];f5=[];f6=[];f7=[];f8=[];
for i=1:fnum
    if f(i,1)==1
        f1=[f1;f(i,:)];%风化颜色 1
    elseif f(i,1)==2
        f2=[f2;f(i,:)];%风化颜色 2
    elseif f(i,1)==3
        f3=[f3;f(i,:)];%风化颜色 3
    elseif f(i,1)==4
        f4=[f4;f(i,:)];%风化颜色 4
    elseif f(i,1)==6
        f6=[f6;f(i,:)];%风化颜色 6
    else

```

```

        f7=[f7;f(i,:)];%风化颜色 7 黑
    end
end
f1num=numel(f1(:,1));
f2num=numel(f2(:,1));
f3num=numel(f3(:,1));
f4num=numel(f4(:,1));
f6num=numel(f6(:,1));
f7num=numel(f7(:,1));

y=ifm(:,5);
Rs=[];%保存 spearman 相关系数
Pvals=[];%相应的 p 值
for i=1:3
    [r,p]=corr(ifm(:,i+1),y,'type','Spearman');%spearman 相关性分析
    Rs=[Rs;r];
    Pvals=[Pvals;p];
end
Rs,Pvals

C1=[alnum,b1num;0,b2num;a3num,b3num];
C2=[c1num,d1num;c2num,d2num];
C3=[e1num,f1num;e2num,f2num;e3num,f3num;e4num,f4num;e5num,0;e6num,f6num;0,f7
num;e8num,0];

[p1,q1]=chi2test(C1);%风化与纹饰 卡方检验
[p2,q2]=chi2test(C2);%风化与类型 卡方检验
[p3,q3]=chi2test(C3);%风化与颜色 卡方检验

Ps=[p1;p2;p3]%卡方检验 p 值
Qs=[q1;q2;q3]%卡方检验 卡方值
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%1.2

%数据筛选 每一个样本所有的元素比例加在一起 保存 85-105 之间的数据
num2=numel(ifm2(:,1));
rows1=[];
for i=1:num2
    total=sum(ifm2(i,7:20));
    if total<85||total>105
        rows1=[rows1;i];
    end
end

```

```

end
ifm2(rows1,:)=[];

num3=numel(ifm2(:,1));

gj=[];%高钾
qb=[];%铅钡
for i=1:num3
    if ifm2(i,3)==1
        gj=[gj;ifm2(i,:)];
    else
        qb=[qb;ifm2(i,:)];
    end
end

num4=numel(gj(:,1));
gjfh=[];%高钾 风化
gjwfh=[];%高钾 无风化
for i=1:num4
    if gj(i,6)==1
        gjfh=[gjfh;gj(i,:)];
    else
        gjwfh=[gjwfh;gj(i,:)];
    end
end

means1=mean(gjfh(:,7:20));%所有高钾 风化的样本的化学成分均值
means2=mean(gjwfh(:,7:20));%所有高钾 无风化的样本的化学成分均值
stds1=std(gjfh(:,7:20));%所有高钾 风化的样本的化学成分标准差
stds2=std(gjwfh(:,7:20));%所有高钾 无风化的样本的化学成分标准差

num5=numel(qb(:,1));
qbfbh=[];%铅钡 风化
qbwfh=[];%铅钡 无风化
for i=1:num5
    if qb(i,6)==1
        qfbfh=[qfbfh;qb(i,:)];
    else
        qbwfh=[qbwfh;qb(i,:)];
    end
end

means3=mean(qfbfh(:,7:20));%所有铅钡 风化的样本的化学成分均值
means4=mean(qbwfh(:,7:20));%所有铅钡 无风化的样本的化学成分均值
stds3=std(qfbfh(:,7:20));%所有铅钡 风化的样本的化学成分标准差

```



```
stds4=std(qbwfh(:,7:20));%所有铅钡 无风化的样本的化学成分标准差
```

```
num6=numel(gjfh(:,1));
gjfhA=[];gjfhB=[];gjfhC=[];
for i=1:num6
    if gjfh(i,4)==1
        gjfhA=[gjfhA;gjfh(i,:)];%高钾 风化 纹饰 A
    elseif gjfh(i,4)==2
        gjfhB=[gjfhB;gjfh(i,:)];%高钾 风化 纹饰 B(空矩阵)
    else
        gjfhC=[gjfhC;gjfh(i,:)];%高钾 风化 纹饰 C(空矩阵)
    end
end
means5=mean(gjfhA(:,7:20));%高钾风化纹饰 A 化学成分均值
```

```
num7=numel(gjwfh(:,1));
gjwfhA=[];gjwfhB=[];gjwfhC=[];
for i=1:num7
    if gjwfh(i,4)==1
        gjwfhA=[gjwfhA;gjwfh(i,:)];%高钾无风化纹饰 A
    elseif gjwfh(i,4)==2
        gjwfhB=[gjwfhB;gjwfh(i,:)];%高钾无风化纹饰 B
    else
        gjwfhC=[gjwfhC;gjwfh(i,:)];%高钾无风化纹饰 C
    end
end
means8=mean(gjwfhA(:,7:20));%高钾无风化纹饰 A 化学成分的均值
means9=mean(gjwfhB(:,7:20));%高钾无风化纹饰 B 化学成分的均值
means10=mean(gjwfhC(:,7:20));%高钾无风化纹饰 C 化学成分的均值
```

```
num8=numel(qbfh(:,1));
qbfhA=[];qbfhB=[];qbfhC=[];
for i=1:num8
    if qbfh(i,4)==1
        qbfhA=[qbfhA;qbfh(i,:)];%铅钡风化纹饰 A
    elseif qbfh(i,4)==2
        qbfhB=[qbfhB;qbfh(i,:)];%铅钡风化纹饰 B
    else
        qbfhC=[qbfhC;qbfh(i,:)];%铅钡风化纹饰 C
    end
end
means11=mean(qbfhA(:,7:20));%铅钡风化纹饰 A 化学成分均值
```

```

means12=mean(qbfhB(:,7:20));%铅钡风化纹饰 B 化学成分均值
means13=mean(qbfhC(:,7:20));%铅钡风化纹饰 C 化学成分均值

num9=numel(qbwfh(:,1));
qbwfhA=[];qbwfhB=[];qbwfhC=[];
for i=1:num9
    if qbwfh(i,4)==1
        qbwfhA=[qbwfhA;qbwfh(i,:)];%铅钡无风化纹饰 A
    elseif qbwfh(i,4)==2
        qbwfhB=[qbwfhB;qbwfh(i,:)];%铅钡无风化纹饰 B
    else
        qbwfhC=[qbwfhC;qbwfh(i,:)];%铅钡无风化纹饰 C
    end
end
means14=mean(qbwfhA(:,7:20));%铅钡无风化纹饰 A 化学成分均值
means15=mean(qbwfhB(:,7:20));%铅钡无风化纹饰 B 化学成分均值
means16=mean(qbwfhC(:,7:20));%铅钡无风化纹饰 C 化学成分均值

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%1.3
alphas1=[];%高钾类 无风化与风化均值的比值
for i=7:20
    alpha=mean(gjwfh(:,i))/mean(gjfh(:,i));
    alphas1=[alphas1;alpha];
end

Predicts1=[];
for i=1:num6
    predict=gjfh(i,7:20).*alphas1';
    Predicts1=[Predicts1;[gjfh(i,1) predict]];
end
M=means2-means1;%高钾无风化与风化均值差
id=[2 9 10 12 13 14];%高钾类不适用比例预测的化学成分序号
Predicts1(:,id+1)=gjfh(:,id+6)+M(:,id);%利用均值差预测含量，并得到最终的高钾风化样本
未风化前的数据

Num1=numel(Predicts1(:,1));
Rows1=[];
for i=1:Num1
    total=sum(Predicts1(i,2:15));
    if total<85||total>105
        Rows1=[Rows1;i];
    end
end
end

```

```

Predicts1(Rows1,:)=[];%对于预测后总量在 85-105 以外的数据进行剔除

alphas2=[];%铅钡类 无风化与风化均值的比值
for i=7:20
    alpha=mean(qbwfh(:,i))/mean(qbfh(:,i));
    alphas2=[alphas2;alpha];
end

Predicts2=[];%预测值
for i=1:num8
    predict=qbfh(i,7:20).*alphas2';
    Predicts2=[Predicts2;[qbfh(i,1) predict]];
end

Num2=numel(Predicts2(:,1));
Rows2=[];
for i=1:Num2
    total=sum(Predicts2(i,2:15));
    if total<85||total>105
        Rows2=[Rows2;i];
    end
end

Predicts2(Rows2,:)=[];%对于预测后总量在 85-105 以外的数据进行剔除

```

(二) 第二题代码

```

%2.1
Mgj=means2-means1;%高钾无风化-高钾风化
Mqb=means4-means3;%铅钡无风化-铅钡风化
Mwfh=means2-means4;%高钾无风化-铅钡无风化
Mfh=means1-means3;%高钾风化-铅钡风化

%2.2 层次聚类算法
% 高钾类 层次聚类
Y1=pdist(gj(:,7:20));
Z1=linkage(Y1,'ward');%离差平方和方法(Ward方法)
F1=dendrogram(Z1,0); %显示系统聚类树
title('高钾类层次聚类法聚类结果');
set(F1,'linewidth',2);
ylabel('标准距离');
[c1,D1] = cophenet(Z1,Y1);
C1=corr(Y1',D1','type','spearman')

```

```

%n=3
p1=[9 11 10 8 17 18];
p2=[2 15 16];
p3=[1 4 3 5 12 13 14 7 6];

P1=gj(p1,:);P1means=mean(P1(:,7:20));
P2=gj(p2,:);P2means=mean(P2(:,7:20));
P3=gj(p3,:);P3means=mean(P3(:,7:20));

Pmeans=[P1means'-P2means' P1means'-P3means' P2means'-P3means'];%将均值做差
pvalues1=[];%线性回归 F 检验 p 值
for i=1:3
    X=ones(size(Pmeans(:,i)));
    [b,bint,r,rint,stats]=regress(Pmeans(:,i),X);%对均值的差做回归,判断其是否显著为
0
    Y_pre=X*b;%改进 p 值算法 %求均值的差的 p 值
    A=X'*X;
    C=inv(A);
    SSy=var(Pmeans(:,i))*(size(Pmeans(:,i),1)-1);
    yhat=Y_pre;
    RSS=(Pmeans(:,i)-yhat)'*(Pmeans(:,i)-yhat);
    MSE=RSS/(size(Pmeans(:,i),1)-size(b,1)-1);
    rsquare=(SSy-RSS)/SSy;
    sp=sqrt(MSE*diag(C));
    t=b./sp;
    n=size(X, 1);
    v=size(X, 2);
    pvalue=2*(1-tcdf(abs(t),n-v));
    pvalues1=[pvalues1;pvalue];
end

%n=2
q1=[p1 p2];
q2=p3;

Q1=gj(q1,:);Q1means=mean(Q1(:,7:20));
Q2=gj(q2,:);Q2means=mean(Q2(:,7:20));

Qmeans=Q1means-Q2means;

X1=ones(size(Qmeans'));
[b,bint,r,rint,stats]=regress(Qmeans',X1);
Y_pre=X1*b;
A=X1'*X1;

```

```

C=inv(A);
SSy=var(Qmeans')*(size(Qmeans',1)-1);
yhat=Y_pre;
RSS=(Qmeans'-yhat)'*(Qmeans'-yhat);
MSe=RSS/(size(Qmeans',1)-size(b,1)-1);
rsquare=(SSy-RSS)/SSy;
sp=sqrt(MSe*diag(C));
t=b./sp;
n=size(X1, 1);
v=size(X1, 2);
pvalue2=2*(1-tcdf(abs(t),n-v));%线性回归 F 检验 p 值

%铅钡类 层次聚类
Y2=pdist(qb(:,7:20));
Z2=linkage(Y2,'ward');%离差平方和方法(Ward方法)
F2=dendrogram(Z2,0); %显示系统聚类树
title('铅钡类层次聚类法聚类结果');
set(F2,'linewidth',2);
ylabel('标准距离');
[c2,D2] = cophenet(Z2,Y2);
C2=corr(Y2',D2','type','spearman')

%n=6
w1=[2 10 8 4 6];
w2=[3 11];
w3=[1 19 23 14 15 21 5 49 40 36 26 38 42 47 48];
w4=[24 44 25 29 30 41 45];
w5=[7 27 28 34 33 37 9 46 39];
w6=[12 17 20 16 18 13 31 35 22 32 43];

W1=qb(w1,:);W1means=mean(W1(:,7:20));
W2=qb(w2,:);W2means=mean(W2(:,7:20));
W3=qb(w3,:);W3means=mean(W3(:,7:20));
W4=qb(w4,:);W4means=mean(W4(:,7:20));
W5=qb(w5,:);W5means=mean(W5(:,7:20));
W6=qb(w6,:);W6means=mean(W6(:,7:20));

Wmeans=[W1means'-W2means' W1means'-W3means' W1means'-W4means'
W1means'-W5means' W1means'-W6means' W2means'-W3means' W2means'-W4means'
W2means'-W5means' W2means'-W6means' W3means'-W4means' W3means'-W5means'
W3means'-W6means' W4means'-W5means' W4means'-W6means' W5means'-W6means'];
pvalues3=[];%线性回归 F 检验 p 值
for i=1:15
    X=ones(size(Wmeans(:,i)));

```

```

[b,bint,r,rint,stats]=regress(Wmeans(:,i),X);
Y_pre=X*b;
A=X'*X;
C=inv(A);
SSy=var(Wmeans(:,i))*(size(Wmeans(:,i),1)-1);
yhat=Y_pre;
RSS=(Wmeans(:,i)-yhat)'*(Wmeans(:,i)-yhat);
MSe=RSS/(size(Wmeans(:,i),1)-size(b,1)-1);
rsquare=(SSy-RSS)/SSy;
sp=sqrt(MSe*diag(C));
t=b./sp;
n=size(X, 1);
v=size(X, 2);
pvalue=2*(1-tcdf(abs(t),n-v));
pvalues3=[pvalues3;pvalue];
end

%n=5
t1=w1;
t2=w2;
t3=w3;
t4=w4;
t5=[w5 w6];

T1=qb(t1,:);T1means=mean(T1(:,7:20));
T2=qb(t2,:);T2means=mean(T2(:,7:20));
T3=qb(t3,:);T3means=mean(T3(:,7:20));
T4=qb(t4,:);T4means=mean(T4(:,7:20));
T5=qb(t5,:);T5means=mean(T5(:,7:20));

Tmeans=[T1means'-T2means' T1means'-T3means' T1means'-T4means'
T1means'-T5means' T2means'-T3means' T2means'-T4means' T2means'-T5means'
T3means'-T4means' T3means'-T5means' T4means'-T5means'];
pvalues4=[];%线性回归 F 检验 p 值
for i=1:10
    X=ones(size(Tmeans(:,i)));
    [b,bint,r,rint,stats]=regress(Tmeans(:,i),X);
    Y_pre=X*b;
    A=X'*X;
    C=inv(A);
    SSy=var(Tmeans(:,i))*(size(Tmeans(:,i),1)-1);
    yhat=Y_pre;
    RSS=(Tmeans(:,i)-yhat)'*(Tmeans(:,i)-yhat);
    MSe=RSS/(size(Tmeans(:,i),1)-size(b,1)-1);

```

```

rsquare=(SSy-RSS)/SSy;
sp=sqrt(MSe*diag(C));
t=b./sp;
n=size(X, 1);
v=size(X, 2);
pvalue=2*(1-tcdf(abs(t),n-v));
pvalues4=[pvalues4;pvalue];
end

%n=4
v1=[t1 t2];
v2=t3;
v3=t4;
v4=t5;

V1=qb(v1,:);V1means=mean(V1(:,7:20));
V2=qb(v2,:);V2means=mean(V2(:,7:20));
V3=qb(v3,:);V3means=mean(V3(:,7:20));
V4=qb(v4,:);V4means=mean(V4(:,7:20));

Vmeans=[V1means'-V2means' V1means'-V3means' V1means'-V4means'
V2means'-V3means' V2means'-V4means' V3means'-V4means'];
pvalues5=[];%线性回归 F 检验 p 值
for i=1:6
    X=ones(size(Vmeans(:,i)));
    [b,bint,r,rint,stats]=regress(Vmeans(:,i),X);
    Y_pre=X*b;
    A=X'*X;
    C=inv(A);
    SSy=var(Vmeans(:,i))*(size(Vmeans(:,i),1)-1);
    yhat=Y_pre;
    RSS=(Vmeans(:,i)-yhat) *(Vmeans(:,i)-yhat);
    MSE=RSS/(size(Vmeans(:,i),1)-size(b,1)-1);
    rsquare=(SSy-RSS)/SSy;
    sp=sqrt(MSe*diag(C));
    t=b./sp;
    n=size(X, 1);
    v=size(X, 2);
    pvalue=2*(1-tcdf(abs(t),n-v));
    pvalues5=[pvalues5;pvalue];
end

%n=3

```

```

u1=v1;
u2=[v2 v3];
u3=v4;

U1=qb(u1,:);U1means=mean(U1(:,7:20));
U2=qb(u2,:);U2means=mean(U2(:,7:20));
U3=qb(u3,:);U3means=mean(U3(:,7:20));

Umeans=[U1means'-U2means' U1means'-U3means' U2means'-U3means'];
pvalues6=[];%线性回归 F 检验 p 值
for i=1:3
    X=ones(size(Umeans(:,i)));
    [b,bint,r,rint,stats]=regress(Umeans(:,i),X);
    Y_pre=X*b;
    A=X'*X;
    C=inv(A);
    SSy=var(Umeans(:,i))*(size(Umeans(:,i),1)-1);
    yhat=Y_pre;
    RSS=(Umeans(:,i)-yhat) *(Umeans(:,i)-yhat);
    MSe=RSS/(size(Umeans(:,i),1)-size(b,1)-1);
    rsquare=(SSy-RSS)/SSy;
    sp=sqrt(MSe*diag(C));
    t=b./sp;
    n=size(X, 1);
    v=size(X, 2);
    pvalue=2*(1-tcdf(abs(t),n-v));
    pvalues6=[pvalues6;pvalue];
end

%n=2
h1=[u1 u2];
h2=u3;

H1=qb(h1,:);H1means=mean(H1(:,7:20));
H2=qb(h2,:);H2means=mean(H2(:,7:20));

Hmeans=H1means-H2means;
X2=ones(size(Hmeans'));
[b,bint,r,rint,stats]=regress(Hmeans',X2);
Y_pre=X2*b;
A=X2'*X2;
C=inv(A);
SSy=var(Hmeans')*(size(Hmeans',1)-1);

```



```

yhat=Y_pre;
RSS=(Hmeans'-yhat) *(Hmeans'-yhat);
MSe=RSS/(size(Hmeans',1)-size(b,1)-1);
rsquare=(SSy-RSS)/SSy;
sp=sqrt(MSe*diag(C));
t=b./sp;
n=size(X2, 1);
v=size(X2, 2);
pvalue7=2*(1-tcdf(abs(t),n-v));%线性回归 F 检验 p 值

```

(三) 第三题代码

```

%3
distances=[];
for i=1:8
    g1=(ifm3(i,3:16)-means1).^2;
    distance1=sqrt(sum(g1(:)));
    distances=[distances;distance1];
    g2=(ifm3(i,3:16)-means2).^2;
    distance2=sqrt(sum(g2(:)));
    distances=[distances;distance2];
    g3=(ifm3(i,3:16)-means3).^2;
    distance3=sqrt(sum(g3(:)));
    distances=[distances;distance3];
    g4=(ifm3(i,3:16)-means4).^2;
    distance4=sqrt(sum(g4(:)));
    distances=[distances;distance4];
end
distances1=distances(1:4,:);%1 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances2=distances(5:8,:);%2 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances3=distances(9:12,:);%3 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances4=distances(13:16,:);%4 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances5=distances(17:20,:);%5 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances6=distances(21:24,:);%6 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances7=distances(25:28,:);%7 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离
distances8=distances(29:32,:);%8 号样本到高钾风化样本、高钾无风化样本、铅钡风化样本、铅钡无风化样本均值的距离

```

```
[r1,pval1]=corr(ifm3(1,3:16)',means2','type','Spearman');%检验
[r2,pval2]=corr(ifm3(2,3:16)',means3','type','Spearman');
[r3,pval3]=corr(ifm3(3,3:16)',means3','type','Spearman');
[r4,pval4]=corr(ifm3(4,3:16)',means4','type','Spearman');
[r5,pval5]=corr(ifm3(5,3:16)',means2','type','Spearman');
[r6,pval6]=corr(ifm3(6,3:16)',means1','type','Spearman');
[r7,pval7]=corr(ifm3(7,3:16)',means1','type','Spearman');
[r8,pval8]=corr(ifm3(8,3:16)',means4','type','Spearman');
```

```
Dstmeans1=mean(distances1);Dststds1=std(distances1);%均值 标准差
Dstmeans2=mean(distances2);Dststds2=std(distances2);
Dstmeans3=mean(distances3);Dststds1=std(distances3);
Dstmeans4=mean(distances4);Dststds1=std(distances4);
Dstmeans5=mean(distances5);Dststds1=std(distances5);
Dstmeans6=mean(distances6);Dststds1=std(distances6);
Dstmeans7=mean(distances7);Dststds1=std(distances7);
Dstmeans8=mean(distances8);Dststds1=std(distances8);
```

(四) 第四题代码

%4.1 每一类别下 各个化学成分间的相关性分析

```
[Rgjfh,Pgjfh]=corr(gjfh(:,7:20),'type','Spearman');%高钾风化内 各元素间的相关系数
及 p 值
[Rgjwfh,Pgjwfh]=corr(gjwfh(:,7:20),'type','Spearman');%高钾无风化内 各元素间的相
关系数及 p 值
[Rqbfh,Pqbfh]=corr(qbfh(:,7:20),'type','Spearman');%铅钡风化内 各元素间的相关系数
及 p 值
[Rqbwfh,Pqbwfh]=corr(qbwfh(:,7:20),'type','Spearman');%铅钡无风化内 各元素间的相
关系数及 p 值
JudgePgjfh=Pgjfh<0.05;
JudgePgjwfh=Pgjwfh<0.05;
JudgePqbfh=Pqbfh<0.05;
JudgePqbwfh=Pqbwfh<0.05;
```

%4.2 每一化学成分在各个类别之间的差异性分析

```
[Pgjqbfh,Qgjqbfh]=chi2test([means1' means3']);%高钾风化 铅钡风化之间成分差异
[Pgjqbwfh,Qgjqbwfh]=chi2test([means2' means4']);%高钾无风化 铅钡无风化之间成分差异
[Pgjfhwh,Qgjfhwh]=chi2test([means1' means2']);%高钾风化 高钾无风化之间成分差异
[Pqbfhwh,Qqbfhwh]=chi2test([means3' means4']);%铅钡风化 铅钡无风化之间成分差异
```

全部预测数据

样品编号	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)	氧化铝 (Al ₂ O ₃)	氧化铁 (Fe ₂ O ₃)
7	67.0195	0.6950	0.0000	6.5584	0.0000	6.7915	1.2392
9	68.7487	0.6950	10.1323	3.8002	0.0000	4.5277	2.3326
10	70.0148	0.6950	15.7994	1.2872	0.0000	2.7783	1.8952
12	68.2205	0.6950	17.3450	4.4131	0.0000	5.0079	2.1139
27	67.0846	0.6950	0.0000	5.7616	2.9631	8.6094	1.4579
8	44.1881	0.0000	0.0000	0.7250	0.0000	2.0105	0.0000
19	65.0315	0.0000	0.0000	1.4354	0.5813	5.3563	1.6756
26	43.4201	0.0000	0.0000	0.7054	0.0000	1.0503	0.0000
39	57.5937	0.0000	0.0000	0.5438	0.0000	0.7502	0.0000
41	40.5021	0.0000	0.7210	2.4298	2.6898	4.9962	2.2551
43	47.6108	0.0000	0.0000	3.1353	0.9360	5.1162	1.7512
49	63.1666	0.0000	0.0000	2.2437	1.4484	8.0720	3.4520
51	53.9954	0.0000	0.0000	1.7538	1.1725	7.8769	1.4992
52	56.4747	9.4969	0.0000	1.1120	0.5419	1.7404	0.2898
54	48.8833	0.0000	0.5244	1.5627	1.2612	6.2265	0.0000
56	63.9564	0.0000	0.0000	0.5928	0.0000	2.7757	0.0000
57	55.7726	0.0000	0.0000	0.6418	0.0000	3.2708	0.0000

样品编号	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷 (P ₂ O ₅)	氧化锶 (SrO)	氧化锡 (SnO ₂)	二氧化硫 (SO ₂)
7	5.0882	0.4117	0.5983	3.0554	0.0417	0.1967	0.1017
9	2.4342	0.4117	0.5983	1.7531	0.0417	0.1967	0.1017
10	1.3192	0.4117	0.5983	0.0000	0.0417	0.1967	0.1017
12	2.5912	0.4117	0.5983	0.7513	0.0417	0.1967	0.1017
27	2.4185	0.4117	0.5983	1.8032	0.0417	0.1967	0.1017
8	6.5492	14.6233	23.8093	0.7137	0.2372	0.0000	0.3005
19	2.2082	21.8330	4.0788	1.7554	0.1218	0.0000	0.0000
26	6.6498	15.0567	24.5870	0.6222	0.2885	0.0000	0.2283
39	0.5536	31.1179	5.5044	0.2306	0.3910	0.0000	0.0000
41	0.1195	22.4958	7.4409	1.4831	0.3013	0.0000	0.0000
43	0.9500	22.8170	2.4854	2.5506	0.3013	0.0000	0.0000
49	0.4404	17.4276	4.6506	2.2067	0.2949	0.0000	0.0000
51	0.8619	20.5175	6.8157	1.6103	0.2500	0.3194	0.0000
52	0.4404	24.1784	6.5870	1.1351	0.2821	0.0000	0.0000
54	0.5222	28.2778	5.3672	0.8429	0.5641	0.0000	0.0000
56	0.4970	21.0325	11.7789	0.5050	0.0000	0.0000	0.0000
57	0.7298	22.9955	13.1893	0.0000	0.0000	0.0000	0.0000

问题四相关性分析

类别	化学成分 1	化学成分 2	Spearman 相关系数	p 值
高钾风化	二氧化硅	氧化钙	-1***	0.0028
高钾风化	二氧化硅	氧化铝	-0.9429**	0.0167
高钾风化	氧化钾	五氧化二磷	-0.9276**	0.0222
高钾风化	氧化钙	氧化铝	0.9429**	0.0167
高钾未风化	二氧化硅	氧化钾	-0.6224**	0.0307
高钾未风化	二氧化硅	氧化钙	-0.6620**	0.0190
高钾未风化	氧化钠	氧化镁	-0.5911**	0.0401
高钾未风化	二氧化硅	氧化铁	-0.7180***	0.0085
高钾未风化	氧化铝	二氧化硫	-0.5996**	0.0393
高钾未风化	氧化钠	氧化钾	0.6609**	0.0193
高钾未风化	氧化钠	氧化钙	0.7356***	0.0064
高钾未风化	氧化钾	氧化钙	0.8266***	9.1501e-04
高钾未风化	氧化铝	氧化铁	0.5884**	0.0441
高钾未风化	氧化铁	氧化铜	0.6175**	0.0324
高钾未风化	氧化铅	氧化钡	0.6083**	0.0358
高钾未风化	氧化镁	氧化锶	0.6281**	0.0287
铅钡风化	氧化钠	氧化铜	-0.4607**	0.0179
铅钡风化	氧化镁	氧化钡	-0.6308***	5.5128e-04
铅钡风化	氧化铝	氧化钡	-0.5368***	0.0047
铅钡风化	氧化铁	氧化钡	-0.5350***	0.0049
铅钡风化	氧化铅	氧化钡	-0.5847***	0.0017
铅钡风化	二氧化硅	氧化锶	-0.6672***	1.9666e-04
铅钡风化	二氧化硅	二氧化硫	-0.4964***	0.0099
铅钡风化	氧化镁	二氧化硫	-0.4415**	0.0240
铅钡风化	氧化铝	二氧化硫	-0.4779**	0.0135
铅钡风化	氧化铁	二氧化硫	-0.4662**	0.0164
铅钡风化	氧化铅	二氧化硫	-0.5062***	0.0083
铅钡风化	二氧化硅	氧化钠	0.4686**	0.0157
铅钡风化	氧化钙	氧化镁	0.6364***	4.7344e-04
铅钡风化	二氧化硅	氧化铝	0.3894**	0.0493
铅钡风化	氧化镁	氧化铝	0.8145***	4.1126e-04
铅钡风化	氧化钙	氧化铁	0.4910**	0.0109
铅钡风化	氧化镁	氧化铁	0.6262***	6.2136e-04
铅钡风化	氧化铝	氧化铁	0.6032***	0.0011
铅钡风化	氧化铜	氧化钡	0.4097**	0.0376
铅钡风化	氧化钙	五氧化二磷	0.5047***	0.0086
铅钡风化	氧化镁	五氧化二磷	0.4864**	0.0117
铅钡风化	氧化铝	五氧化二磷	0.3891**	0.0494
铅钡风化	氧化铝	氧化锡	0.4253**	0.0303
铅钡风化	氧化铜	二氧化硫	0.542 ⁹ ***	0.0042
铅钡风化	氧化钡	二氧化硫	0.6242***	6.5466e-04

铅钡未风化	氧化钠	氧化铁	-0.4192**	0.0465
铅钡未风化	氧化钙	氧化铜	-0.4734**	0.0225
铅钡未风化	氧化铁	氧化铜	-0.5141**	0.0121
铅钡未风化	二氧化硅	氧化铅	-0.6024***	0.0024
铅钡未风化	氧化铝	氧化铅	-0.4150**	0.0489
铅钡未风化	二氧化硅	氧化钡	-0.6513***	7.6111e-04
铅钡未风化	氧化钠	五氧化二磷	-0.6747***	1.1336e-04
铅钡未风化	氧化铜	氧化锡	-0.5494***	0.0066
铅钡未风化	氧化钾	氧化镁	0.4234**	0.0441
铅钡未风化	氧化钾	氧化铝	0.6182***	0.0017
铅钡未风化	氧化镁	氧化铝	0.5431***	0.0074
铅钡未风化	氧化铜	氧化钡	0.5515***	0.0064
铅钡未风化	氧化钙	氧化锡	0.4958**	0.0161
铅钡未风化	氧化铁	氧化锡	0.4544**	0.0294

注：*表示在 10%显著水平下显著，**表示在 5%显著水平下显著，***表示在 1%显著水平下显著。