**针对问题一**，本文首先对数据进行预处理，对标单2中的未检测数据填充1e-5，再采用**中心对数比变换方法**对成分数据进行转换用于**消除定和限制对后续分析的影响**，并对表单1和表单2中定类数据量化，运用梁化后的数据来分析和聚类。对于第一小问，题目要求分别研究文物样品表面的风化与颜色、玻璃类型以及纹饰间的关系，本文对四类别进行**pearson卡方检验法**，并绘制热力图分析。在第二小问中，题目要求描述玻璃化学成分的统计规律，本文引入均值、最大值、最小值、标准差、偏度系数、峰度系数描述化学成分的统计量。

**针对问题二**，在第一小问分析铅钡玻璃和高钾玻璃的分类规律时，本文引入了监督学习方法，采用**决策树模型**对数据进行分类。通过模型的求解，发现该模型在精确率、召回率、准确率以及 F1 系数上均为 1，这表明模型具有优异的分类性能。第二小问要求对不同的玻璃类型选择适合的化学成分进行亚类分类，本文利用**k-medoids聚类法**，通过**肘部法则和Gap Statistic 值**确定四种聚类结果，分类结果通过计算的WCSS和DB系数反应具有合理性。同时在增加一个0.02的噪声强度范围内，**Gap Statistic**峰值始终为k=4，聚类中心没有发生改变，由此说明模型敏感性。

**针对问题三**，题目要求通过分析表单3中位置玻璃类别的文物样品的化学成分，鉴别未知玻璃类型的文物编号所属类型。本文基于问题二的建模与求解，使用问题二中训练好的决策树进行分类，分别对高钾和铅钡玻璃分类进行敏感性分析，结果表明本文使用方法在准确性和敏感性上皆有不错表现。

**针对问题四**，题目要求分析不同玻璃类别的化学成分间的相关性，即分别研究高钾玻璃与铅钡玻璃的化学成分指标间的关联，本文使用**pearson卡方检验法**，根据相关性大小衡量各化学成分之间的关联关系；其次比较不同玻璃类别见的关联度值得差异程度，最后通过观测热力图发现，高钾玻璃关联程度较为集中，而铅钡玻璃关联程度较为分散，并根据关联矩阵分析差异性。

关键词：中心对数比变换 决策树模型 k-medoids聚类分析 卡方检验