

¹ Ch5 聚类分析

1.1 简答题

K-Means聚类分析的基本步骤

1. 指定聚类数目K
2. 确定K个初始类中心点：用户指定和系统指定两种方式
3. 根据距离最近原则进行分类：依次计算每个数据点到K个类中心点的欧氏距离，并按照与K个类中心点距离最短的原则将所有观测分派到K个分类中
4. 重新确定K个类中心点：依次计算各类中各个变量的均值，并以均值点作为K个类的中心点，完成依次迭代
5. 判断是否已经满足终止聚类分析的条件
 1. 目前的迭代次数等于指定的迭代次数时终止聚类
 2. 新确定的类中心点距上次迭代形成的类中心点的最大偏移量小于指定的量时终止聚类（默认0.02）

相对于层次聚类，K-Means聚类有怎样的特点

- K-Means快速聚类是一个反复迭代的分类过程，在聚类过程中，观测所属的类会不断调整，直至最终达到稳定为止
- K-Means聚类是一种划分聚类方法，它将数据集划分为K个不重叠的聚类簇。每个数据点只属于一个聚类簇，而不是属于多个层次

变量之间的高度相关性是否会对层次聚类分析结果造成影响以及原因

- 如果所选变量之间存在较强的线性关系，能够互相替代，那么计算距离时同类变量将重复“贡献”，在距离中占较高权重
- 层次聚类算法通常基于距离或相似度度量来衡量数据点之间的相似程度。如果变量之间高度相关，它们可能会在距离度量上显示出较小的差异，被视为相似的数据点。这可能导致算法将它们划分到同一簇中，即使它们在其他未考虑的方面可能存在明显的差异

1.2 实际运用

☑ 聚类分析

² Ch6 因子分析

2.1 简答题

简述因子分析的主要步骤

- 判断因子分析的前提条件是否满足
- 因子提取
- 使因子具有命名解释性
- 计算各观测的因子得分

2.2 实际运用

☑ 因子分析方法