### 分析流程 数据源： 2021江苏指标\_运行阶段.xlsx 算法配置： 算法： 分层聚类 分析结果： 暂无数据

### 分析步骤 1. 检查数据的缺失值情况，可以初步判断聚类的可信率。 2. 根据聚类表了解样本之间的距离和聚类情况。 3. 根据聚类树状图进行分析，可以观测聚类情况和对数据进行探测，发现类之间的层次关系。

### 详细结论

**输出结果1：数据情况汇总**

**图表说明：**

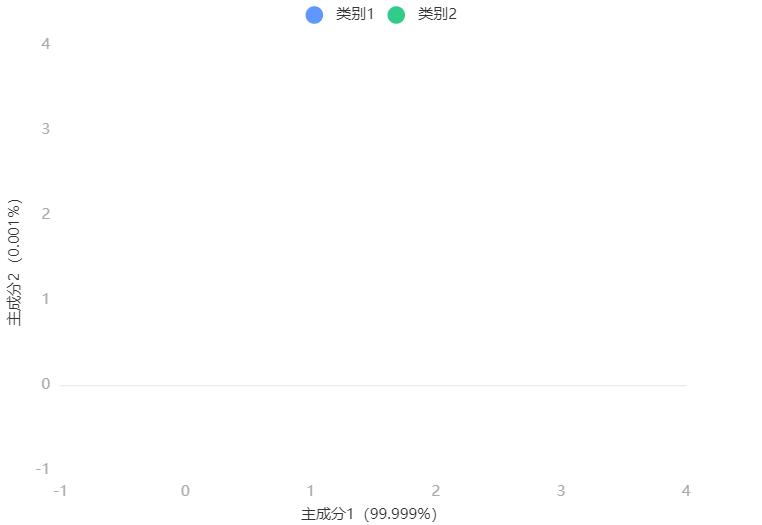
上表展示了数据的有效和缺失情况，当某行数据存在缺失时，则计入缺失数据。

**输出结果2：聚类表**

**图表说明：**

上图是聚类表，在聚类表中列出了逐步聚类的过程。

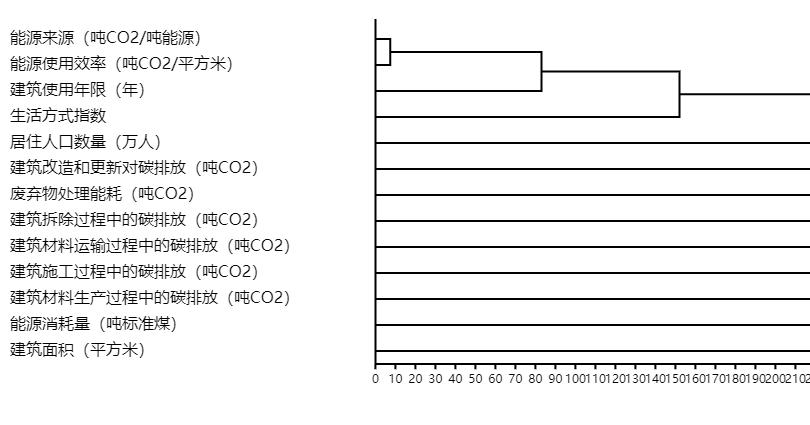
**输出结果3：聚类散点图**

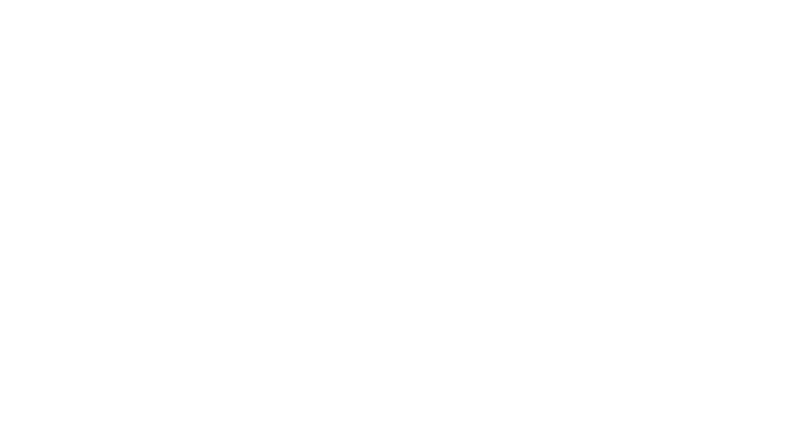


**图表说明：**

若对不同样本量（行）进行聚类分群，此时变量数等于两个，上图是根据两变量的数据绘制出来的散点图；或者此时变量数大于两个，上图是提取主成分分析(PCA)降维后前两个主成分来绘制散点图，在一定程度上可查看聚类效果（若是前两个主成分的方差解释率较低，该图的意义不大）。

**输出结果4：聚类树状图**





**图表说明：**

上图以可视化的形式展示了分层聚类的聚类情况，如果聚类的类别超过50个，此图将仅能下载查看。

**输出结果5：评价指标**

**图表说明：**

● 轮廓系数：对于一个样本集合，它的轮廓系数是所有样本轮廓系数的平均值。轮廓系数的取值范围是[-1,1]，同类别样本距离越相近不同类别样本距离越远，分数越高，聚类效果越好。  
● DBI（Davies-bouldin)：该指标用来衡量任意两个簇的簇内距离之后与簇间距离之比。该指标越小表示聚类效果越好。  
● CH(Calinski-Harbasz Score)：通过计算类内各点与类中心的距离平方和来度量类内的紧密度（分母），通过计算类间中心点与数据集中心点距离平方和来度量数据集的分离度（分子），CH指标由分离度与紧密度的比值得到，CH越大表示聚类效果越好。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] Jain, A. K. , and R. C. Dubes . "Algorithms for clustering data." Technometrics 32.2(1988):227-229.