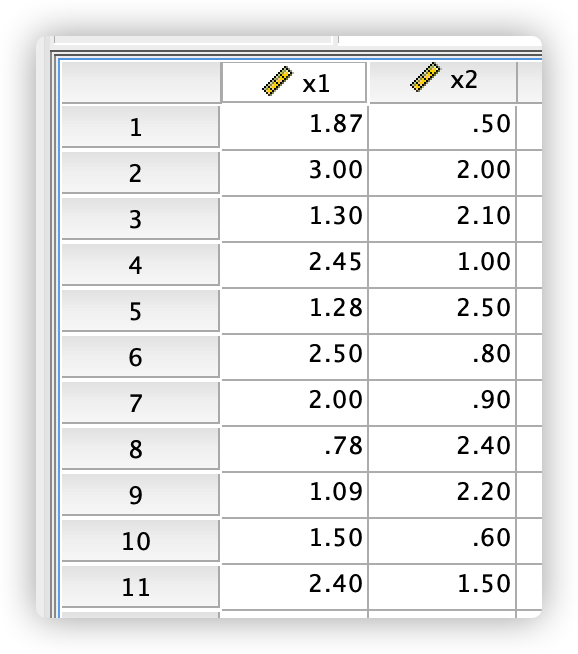
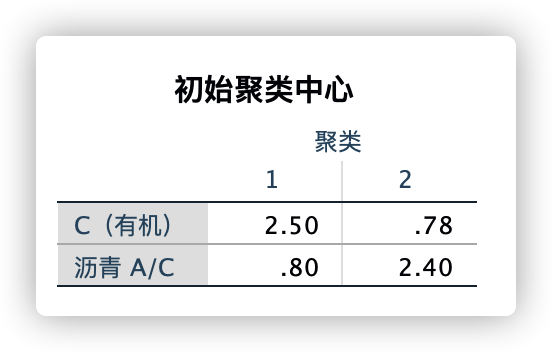
**多元统计分析聚类分析**

**1202110887 徐鸿飞**

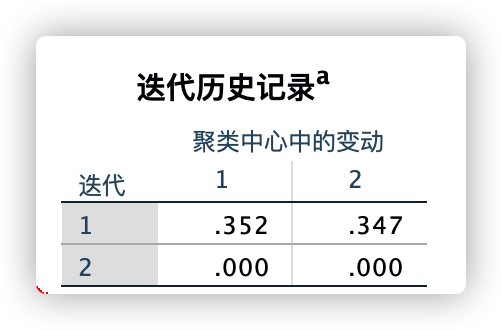
**题目**（教材P129第6题）：已测得可反映生油地质特征的两个变量如下表所示。已知底层3号、5号为非生油层，地层4号、10号为生油层，试用聚类分析评议哪些地层可能是生油层？哪些不是？



因为题目要判断是否生油层，先使用k均值聚类，指定k=2



快速聚类初始聚类中心分别选择了极大值和极小值。



由于聚类中心中不存在变动或者仅有小幅变动，因此实现了收敛。任何中心的最大绝对坐标变动为 .000。当前迭代为 2。初始中心之间的最小距离为 2.349。



分别指出了每一项属于哪一类，题目中指出3、5属于非生油层，此处类别都是2，4、10位生油层，此处类别都是1，所以聚类为1的是生油层，2是非生油层。则1、2、4、6、7、10、11是生油层，3、5、8、9是非生油层。



最终迭代完的聚类中心和中心之间的距离。



显著性均为0.002，说明聚类效果是比较好的。由于已选择聚类以使不同聚类中个案之间的差异最大化，因此 F 检验只应该用于描述目的。实测显著性水平并未因此进行修正，所以无法解释为针对“聚类平均值相等”这一假设的检验。



仅作为统计，有7个属于1类，4个属于2类。