**一．这是环保监测站某年测得的二氧化硫、氮氧化物与飘尘的含量。**

**序号 二氧化硫(X1) 氮氧化物(X2) 飘尘(X3)**

**1 0.045 0.043 0.265**

**2 0.066 0.039 0.264**

**3 0.094 0.061 0.194**

**4 0.003 0.003 0.102**

**5 0.048 0.015 0.106**

**6 0.210 0.066 0.263**

**7 0.086 0.072 0.274**

**8 0.196 0.072 0.211**

**9 0.187 0.082 0.301**

**10 0.053 0.060 0.209**

**11 0.020 0.008 0.112**

**12 0.035 0.015 0.170**

**13 0.205 0.068 0.284**

**14 0.088 0.058 0.215**

**15 0.101 0.052 0.181**

**16 0.045 0.005 0.122**

**要求：**

1. **采用三种聚类分析法以上将1-14组数据分为三类；**
2. **根据每种分类情况，建立二种以上判别分析，寻找最佳分类情况；**
3. **根据最佳分类，给第15组与16组数据归类。**

**二．USJudgeRatings(R语言自带数据集):**

1. **做主成分分析，选择主成分，使得其累积方差贡献率在80%以上；**
2. **做碎石图；**
3. **做因子分析，使用最大似然法的估计法，选择因子，使得累积因子方差大于70%；**
4. **做相应的因子旋转，比较旋转前与旋转后的载荷阵，选择合适的载荷阵做后续分析；**
5. **用回归法计算因子得分，并且基于系数阵与因子得分画出biplot()。**

**需要对得到的主成分与因子做相应的解释。**

**三. nba数据(部分数据来自于www.espn.com/nba/statistics)**

**预测今年的总冠军：**

**寻找从前总冠军的特征，预测今年的总冠军归属.**

**可能的变量：**

**常规赛成绩(什么样的成绩几乎得冠军，什么样的成绩几乎没有得过冠军)**

**常规赛的主场战绩**

**比赛平均分差**

**单打次数**

**投篮次数**

**平均得分**

**平均篮板**

**平均助攻**

**人员的年龄**

**教练的执教年龄**

**常规赛的进攻排名**

**常规赛的防守排名等。**

**可能的思路：**

**采用判别分析法建立模型；**

**采用logistic模型建立相应的回归方程；**

**四．采用聚类分析的方法对不同的球队进行归类，比如说进攻型球队，防守型球队，攻守平衡型球队等；**

**(something beyond:分析球队的不足，用数据说话，给出相应的引援建议；)**

1. **利用NCI60的数据，寻找一个合理的分类方法，使得分类的错误率尽可能地低。(ISLR包)**

**可能的思路：如何挑选一些与分类相关的变量进行相应的分类。**

**可以对变量作一定的变换后，进行相应的分类。**

**是否可以提供一些新的分类方法。**

1. **通过影片的简介(也就是一段话)，寻找同类型影片的特征（关键字等）(Crime,Drama,Action,Thriller,Biography,History,Adventure,Fantasy...)并建立相应的判别模型。(www.imdb.com/)**

**注意：这组数据需要自己整理数据，用相应的R/Matlab/Python/Java/C将数据从网络整理到EXCEL表格中做统计分析。**

1. **利用NBA数据，分析NBA比赛中的变化趋势。**

**比如：**

**三分球所占的比重，三分命中率的变化，比赛回合的变化，**

**比赛分差的变化，比赛犯规的变化，助攻的变化，篮板的变化等。**