

# 1. 全国2011年31个省、市、自治区8项消费性支出的聚类分析

## 1.1 读取数据

In [ ]:

```
1 #####读入数据#####
2 #打开数据文件consumption.xls, 选取A1:I32区域, 然后复制
3 #data1 <- read.table("clipboard", header=TRUE)
4 data1<-read.table("C:/Users/bff/Desktop/me-ppt/consumption.txt", header=T, fileEncoding = "GB18030")
5 head(data1)
```

In [ ]:

```
1 summary(data1) #描述性统计量
2 rownames(data1) # 显示前几行看一下数据
```

In [ ]:

```
1 class(data1$x1) #看一下数据的类型
2 boxplot(data1)
```

## ## 1.2 k-mean聚类

In [ ]:

```
1 KM <- kmeans(data1, 4, iter.max = 20, nstart = 1, algorithm = "Hartigan-Wong")
2 #聚类的个数为4, 随机集合的个数为20, 算法为 "Hartigan-Wong", 其他备选算法为 "Lloyd", "Forgy", "MacQueen"
3 #sort(KM$cluster) #对分类结果进行排序并查看分类情况
4 KM
```

In [ ]:

```
1 #通过图示的方式看一下聚类
2 d1<-data1[, 1:2]
3 dim(d1)
4 #head(d1)
5 library(factoextra)
6 library(ggplot2)
7 fviz_cluster(KM, d1, geom = "point",
8               ellipse= FALSE, show.clust.cent = T,
9               palette = "jco", ggtheme = theme_classic())
```

## 1.3 k-medoids聚类

In [ ]:

```
1 library(cluster) # 加载包
2 pam3=pam(data1,3) #k-mediod聚类方法, pam算法
3 pam3
```

In [ ]:

```
1 pam3$silinfo$avg.width#查看silhouette width均值
```

In [ ]:

```
1 si3=silhouette(pam3) #查看每个silhouette统计量
2 plot(pam3) #Silhouette Plot
```

## 2. 红酒的聚类问题

### 10种红葡萄酒颜色、香味、酸度、甜度、纯度、果味6个维度的得分数据

In [ ]:

```
1 # 系统聚类
2 data2<-read.table("C:/Users/bff/Desktop/me-ppt/wine.txt",header=T) #读取数据
3 summary(data2)
4 boxplot(data2)
```

In [ ]:

```
1 d<-dist(data2,method="minkowski")
```

In [ ]:

```
1 HC<-hclust(d,method="single") #采用最小距离法(single) 聚类
2 #method为系统聚类方法, 包括"single"(最小距离法), "complete"(最大#距离法),
3 #"average"(类平均法), "median"(中间距离法), "centroid"(#重心法), "ward"(Ward法) 等
4
```

In [ ]:

```
1 plot(HC) ##绘制聚类树状图
```

In [ ]:

```
1 HC<-hclust(d,method="complete") #采用最大距离(complete) 聚类
2 plot(HC) #绘制聚类树状图
```

## 3. 世界146个国家和地区人文发展情况的聚类分析

In [ ]:

```
1 data3<-read.table("C:/Users/bff/Desktop/me-ppt/world.txt",header=T,fileEncoding = "GB18030")
2 head(data3)
```