

# 使用R语言进行聚类与判别分析

## 引言

随着

## 数据采集

本文中提到的数据来源于《中国统计年鉴》中2014年全国各大重要城市的年度数据，本文提取了


## 主成分聚类

由于所取数据中涉及的特征较多，也为了避免多个特征之间存在较强的相关关系使得距离度量失衡，聚类效果变差，因此本文选择使用主成分聚类的方法进行聚类分析。

主成分分析法由 K. Pearson 在1901年提出，是一种可以将多项相关指标化为少数几个不相关综合指标的降维方法。聚类分析则是数理统计中的一种多元分析方法，它使用数学方法定量地确定样本的亲疏关系，从而对样本进行合理地划分。所谓主成分聚类分析就是将主成分分析与聚类分析相结合的一种综合评价方法，即先作主成分分析，再进行聚类分析。

## 主成分分析

主成分分析算法可以分为以下几个步骤：

- 1) 设有  $n$  个观察样本，每个样本有  $p$  项指标，

使用R语言进行主成分分析的步骤如下

## 使用 k-means 方法对进行聚类

K 均值方法是

## 判别分析

## Fisher判别分析

