

**聚类分析实验报告**

院 系 数学与统计学院

专 业 应用统计学

姓名-学号 王思璐201813870036

丁浩洋201813230001

孙咏晟201813870013

刘治宇201813870011

指导教师 来 鹏

**二Ｏ 20 年 4 月 11 日**

## 基于聚类分析的1995全国工业企业普查数据的实证研究

摘要

1995年4月3号, 第三次全国工业普查全国动员电话会召开，为了摸清家底、制定各地区、各行业的长远发展规划和政策, 更有效地进行宏观经济调控, 展开了全国第三次工业普查。为了综合评价我国各省市自治区的工业企业发展情况，本文运用聚类分析法对全部工业企业的主要经济指标进行分析，研究不同工业企业下全国省市之间的分类划分情况、各省市之间形成的分类划分情况及其体现的地域、经济、文化等特征，并分析数据所体现的工业企业现状。

**正文**

## ****一、建立评价指标体系****

在综合考虑指标的系统性、全面性以及数据的可获得性之后，本文选取工业总产值(当年价格新规定)(亿元)作为评价指标。相关数据来源于1995年全国工业企业普查数据。

。

## 评价方法

本文研究方法为K-means聚类分析法。

聚类分析是指基于收集到的数据按照一定的原则进行分类，分类的结果是同类别数据间的差异性最小，不同类别数据间的差异性则较大，便于后期对数据进行分类研究。

## 评价过程

本文主要通过R软件实现对数据的聚类分析，过程如下：

1. 从本地导入全部工业企业、国有工业企业和私营工业企业的省市与工业总产值数据，并存储在数据框中。
2. 通过调用NbClust工具包，确定合适的分类数目。
3. 利用K-means函数实现对各省市的聚类，得出聚类结果。

程序实现及其结果如下：

library(xlsx)

guoyoudata=read.xlsx('industry\_data.xlsx',sheetIndex = 1,encoding = 'UTF-8')

quanbudata=read.xlsx('industry\_data.xlsx',sheetIndex = 2,encoding = 'UTF-8')

siyindata=read.xlsx('industry\_data.xlsx',sheetIndex = 3,encoding = 'UTF-8')

library(NbClust)

geshu=function(mydata){

df=scale(mydata[-1])

set.seed(1234)

devAskNewPage(ask=TRUE)

nc=NbClust(df,distance="euclidean", min.nc=3, max.nc=7,method = 'kmeans')

print(table(nc$Best.n[1,]))

barplot(table(nc$Best.n[1,]),xlab='Clusters',ylab='Criteria',main='Numbers of Clusters Chosen by 26 Criteria')}

fenxi=function(mydata,n){

df=scale(mydata[-1])

km<-kmeans(df, n) #k均值法

km

}

geshu(quanbudata)

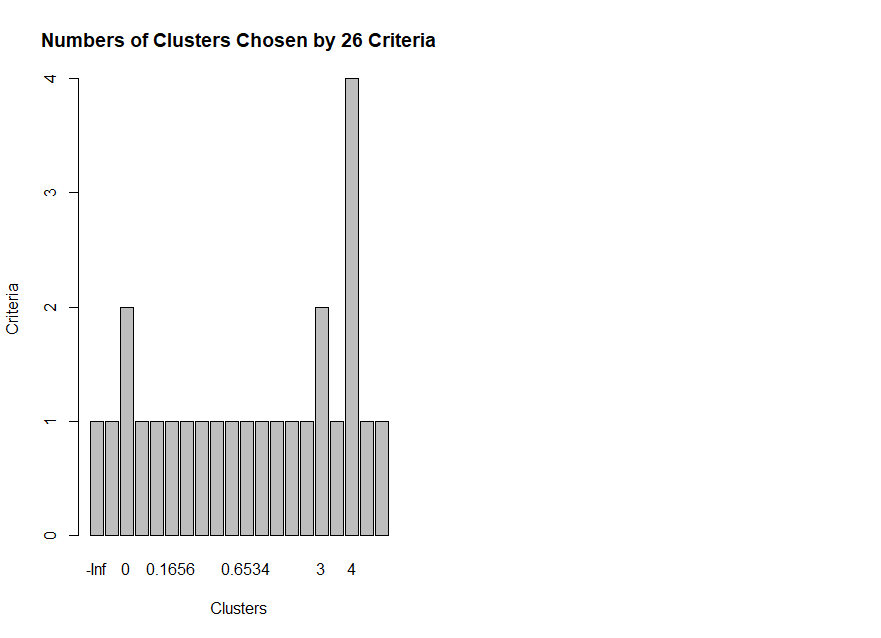
geshu(guoyoudata)

geshu(siyindata)

fenxi(quanbudata,4)

fenxi(guoyoudata,5)

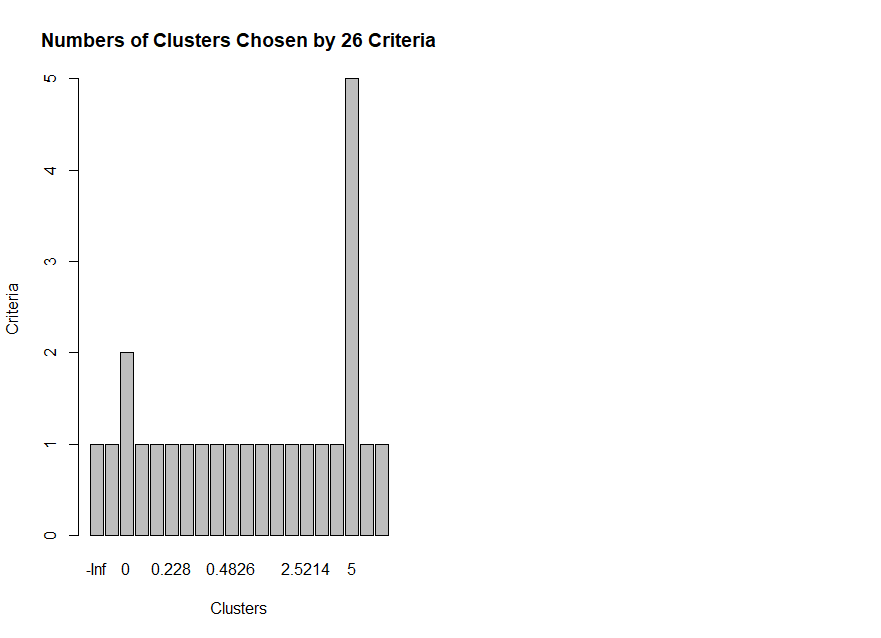
fenxi(siyindata,4)



**对全部工业企业的主要经济指标进行分析**

|  |  |
| --- | --- |
| 组别 | 省份 |
| 1 | 内蒙古、吉林、江西、海南、贵州、云南、陕西、甘肃、青海、宁夏、西藏、新疆 |
| 2 | 江苏、浙江、山东、广州 |
| 3 | 北京、天津、山西、黑龙江、安徽、福建、湖南、广西 |
| 4 | 河北、辽宁、上海、河南、湖北、四川 |

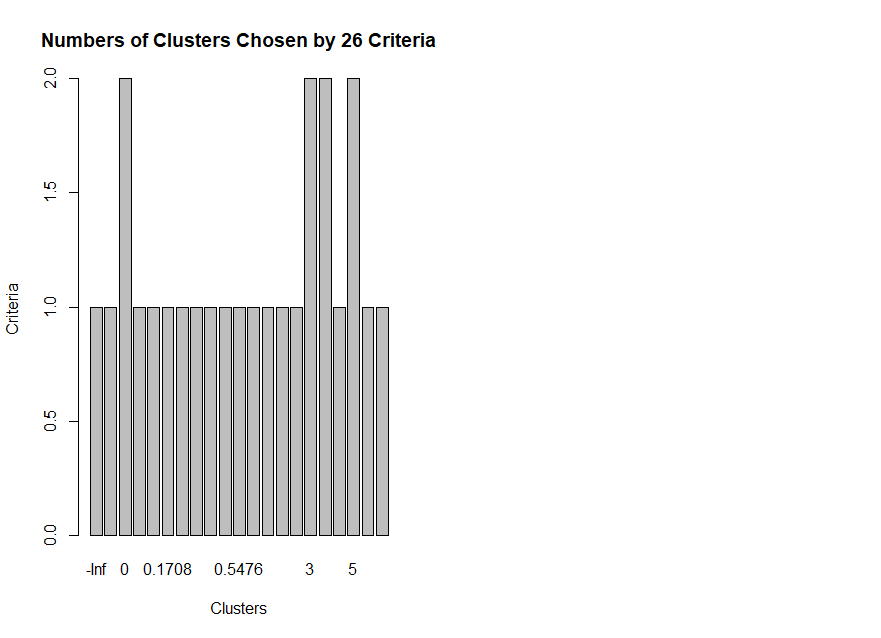
表1



**对国有企业的主要经济指标进行分析**

|  |  |
| --- | --- |
| 组别 | 省份 |
| 1 | 西藏 |
| 2 | 北京、山西、吉林、浙江、安徽、湖南、云南 |
| 3 | 天津、内蒙古、福建、江西、广西、贵州、陕西、甘肃、新疆 |
| 4 | 海南、青海、宁夏 |
| 5 | 河北、辽宁、上海、江苏、山东、河南、湖北、广东、四川 |

表2



**对私营企业的主要经济指标进行分析**

|  |  |
| --- | --- |
| 组别 | 省份 |
| 1 | 河北、山东、河南、四川2 |
| 2 | 山西、上海、江苏、安徽、福建、湖北、湖南、广西  陕西3 |
| 3 | 北京、天津、内蒙古、吉林、黑龙江、江西、海南、贵州、云南、甘肃、新疆、青海、宁夏4 |
| 4 | 浙江、广东1 |

表3

## 四、结论分析

### 4.1对工业企业的经济指标进行分析

从总体的工业企业分类结果来看，我国当时的工业经济水平呈现明显的基于地域经济发展不平衡的特征。其中第一类分类主要为我国西北部及部分中部发展相对落后的地区,由于地域、资源及生产力的限制，这部分地区的生产水平受到较大的限制。其中第三类与第四类为中部力量，工业发展情况相对第一类而言更加发达。第二类则为我国当时最为发达的四个地区，可以发现它们均为东部沿海城市，拥有便捷的交通优势，且相对享受政策倾斜，拥有良好的外贸条件。

### 4.2对国有企业工业经济指标进行分析

其中第一类与第四类均为我国边界地区，交通不便且气候、地形等自然条件极端，在我国当时有限的能力下并无法对这些地区的资源采取充分的利用，故发展最为落后。而第二类与第三类均为我国国有企业发展的中部力量，其中山西、吉林等地拥有丰厚的自然资源，这对以重工业为主的国有企业有极大的裨益，北京、天津等地由于其政治地位，也带动自身及周边地区的经济发展。第五类中的地区即中坚力量，可以发现其分布与总体分类结果基本类似，拥有优良的区位优势、资源优势及政策扶持优势。

### 4.3对私营企业工业经济指标进行分析

在私营企业的划分中我们可以发现，像北京、吉林黑龙江这些总体工业水平比较高的地区与较低的青海、西藏等地划为一类，其中可以看出该地当时发展的政策导向以国有重工业为主。其他的中部及东部地区在私营企业有所发展，浙江和广东凭借其地域优势和文化特色，成为私营企业生态最为良好的两个地区，其中浙江的温州被称为中国民营之乡，曾是商品经济最为发达的地区，“温州精神”也广为流传。

参考文献

[1].第三次全国工业普查[J].现代工业经济和信息化,2012(07):68.

[2]秦杨杨.基于主成分分析和K-means聚类分析的地区经济发展研究[J].中国商论,2020(04):214-215.

[3]刘小玄.中国工业企业的所有制结构对效率差异的影响——1995年全国工业企业普查数据的实证分析.中国社会科学院经济研究所.100836.