

# The Cluster Analysis of Financial Data about the Enterprises above Designated Size in Yunnan Province

Kaipeng Hu, Lan Feng

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan  
Email: hu.kp1993@hotmail.com

Received: Nov. 20<sup>th</sup>, 2018; accepted: Dec. 3<sup>rd</sup>, 2018; published: Dec. 10<sup>th</sup>, 2018

---

## Abstract

In this paper, we investigated the financial data of industrial enterprises of Yunnan province through cluster analysis. Besides, we analyzed the relationship between various industries, the overall financial and economic status, as well as whether there were significant differences in financial status between industrial categories. Based on the data, we find that the economic benefits and financial status of tobacco products industry and chemical fiber manufacturing industry is better. As a whole, the financial status of manufacturing is better than mining, power, thermal production and supply industry.

## Keywords

Enterprises above Designated Size, Financial Data, Cluster Analysis

---

# 云南省规模以上工业企业财务数据聚类分析

胡凯鹏, 冯 兰

云南财经大学, 统计与数学学院, 云南 昆明  
Email: hu.kp1993@hotmail.com

收稿日期: 2018年11月20日; 录用日期: 2018年12月3日; 发布日期: 2018年12月10日

---

## 摘 要

本文通过聚类分析, 对云南省规模以上工业企业财务数据进行研究, 对各行业之间的关系, 和整体的财务经济状况, 以及产业类别之间财务状况是否有明显差异进行分析研究。通过分析了解到云南省规模以

上工业企业中烟草制品业、化学纤维制造业两个工业企业的各种经济效益与财务状况都比较好。整体看来制造业的状况比采矿业、电力、热力生产和供应业的财务状况好。

## 关键词

规模以上企业, 财务数据, 聚类分析

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

企业是国民经济的重要组成部分, 2015年云南省工业生产总值达到3848.26亿元, 占全省生产总值的28.26%。云南省第二产业生产总值有71.05%是由工业企业贡献的。可见工业企业对经济增长起着举足轻重的作用。工业企业的经济财务指标不仅关系到工业企业的生存与发展, 也是影响国民经济可持续发展的重要因素之一。企业管理经营中, 企业的财务指标对企业未来发展至关重要, 因此有必要对云南省规模以上工业企业的财务数据进行研究。对其财务指标进行分析, 反映公司业务活动的变化情况, 从而综合分析企业的财务状况和经营成果。利用上述指标, 很多研究人员做出了丰硕的成果, 金硕元等使用因子分析的方法对规模以上工业的财务情况进行分析[1]。余宏对国有企业的主要财务指标进行了分析[2]。

## 2. 数据来源及说明

本文采用数据来源于中华人民共和国统计局编制的《云南省统计年鉴(2016版)》。本组数据内容是2015年云南省规模以上工业企业主要财务指标, 计量单位: %。其中原始变量为八个基础指标。

### 2.1. 其变量名及含义

Table 1. Variable description

表 1. 变量说明

| 变量名 | 变量含义     |
|-----|----------|
| X1  | 综合经济效益指数 |
| X2  | 总资产贡献率   |
| X3  | 资本保值增值率  |
| X4  | 资产负债率    |
| X5  | 流动资产周转率  |
| X6  | 成本费用利润率  |
| X7  | 全员劳动生产率  |
| X8  | 产品销售率    |

经济效益指数的选取必须可以正确体现出企业的经济效益, 所以以利润为主的指标将被使用以实现上述目标。以利润和实物相关指标挂钩的企业, 应降低实物指标比重, 逐步提高以利润为主的指标的比重。总资产贡献率反映企业全部资产的获利能力, 是企业经营业绩和管理水平的集中体现, 是评价和考

核企业盈利能力的核心指标。资本保值增值率是财政部制定的评价企业经济效益的十大指标之一, 资本保值增值率反映了企业净资本的变动情况, 是企业发展能力的集中体现。资产负债率又称举债经营比率, 它反映债权人发放贷款的安全程度的指标, 通过将企业的负债总额与资产总额相比较得出, 反映在企业全部资产中属于负债比率。流动资产周转率指企业一定时期内流动资产的周转次数, 反映企业经营的情况, 也反映资金利用效果和再生产循环的速度。成本费用利润率是企业全部生产投入与实现利润之间的比率。全员劳动生产率指平均每个职工在单位时间内创造的工业生产的最终结果, 反映企业的生产效率和劳动投入的经济效益。产品销售率: 反映产品已实现销售的程度, 是分析产销链接的情况, 研究产品满足社会需求的重要指标。考虑到以上指标(表 1)是反映企业运行的核心财务数据, 本文着重对以上指标进行分析。

## 2.2. 各行业名称

行业细分见表 2。

**Table 2.** Sub-sectors  
**表 2.** 各细分行业

| 按行业类别分                     | 制造业                  |
|----------------------------|----------------------|
| <b>采矿业(1)</b>              | 造纸和纸制品业              |
| 黑色金属矿采行业                   | 印刷和记录媒介复制业           |
| 有色金属矿采行业                   | 文教、工美、体育和娱乐用品制造业     |
| 非金属矿采行业                    | 石油加工、炼焦和核燃料加工业       |
| 开采辅助活动                     | 化学原料和化学制品制造业         |
| <b>电力、热力、燃气及水生产和供应业(3)</b> | 医药制造业                |
| 电力、热力生产和供应业                | 化学纤维制造业              |
| 燃气生产和供应业                   | 橡胶和塑料制品业             |
| 水的生产和供应业                   | 非金属矿物制品业             |
| <b>制造业(2)</b>              | 黑色金属冶炼和压延加工业         |
| 农副食品加工业                    | 有色金属冶炼和压延加工业         |
| 食品制造业                      | 金属制品业                |
| 酒、饮料和精制茶制造业                | 通用设备制造业              |
| 烟草制品业                      | 专用设备制造业              |
| 纺织业                        | 汽车制造业                |
| 纺织服装、服饰业                   | 铁路、船舶、航空航天和其他运输设备制造业 |
| 皮革、毛皮、羽毛及其制品和制鞋业           | 电气机械和器材制造业           |
| 木材加工和木、竹、藤、棕、草制品业          | 计算机、通信和其他电子设备制造业     |
| 家具制造业                      | 仪器仪表制造业              |

## 3. 理论与实证分析

### 3.1. 聚类分析

聚类就是把未知分类的情况下, 相似的样品聚集成一类。聚类中有一种方法叫系统聚类法(或层次聚类法), 它的基本思想是: 第一步, 将  $n$  个样品各自聚为一类; 第二步, 用规定的距离方法将最小的两

个距离的样品聚为一个新类, 将把新类与其他类的距离计算出来, 进行合并; 第三步, 重复上述合并, 直到合并完全为止[3] [4] [5] [6]。

聚类使用的距离的方法最常见的是一下几种, 即最短距离法, 最长距离法, 类平均法, 重心法, 离差平方和法[7] [8]。

最短距离法也就是类与类之间的距离定义为两类最近的距离, 即

$$D_{KL} = \min_{i \in G_k, j \in G_l} d_{ij} \quad (1)$$

其中:

$d_{ij}$  表示第  $i$  个样品与第  $j$  个样品之间的距离;

$G_i$  表示类别;

$D_{KL}$  表示  $G_k$  与  $G_l$  之间的距离。

最长距离法也就是类与类之间的距离定义为两类最长的距离, 即

$$D_{KL} = \max_{i \in G_k, j \in G_l} d_{ij} \quad (2)$$

其中公式(2)中各字母含义与公式(1)相同。

类平均法也就是把类与类之间的距离定义为所有物之间的距离的平均, 即

$$D_{KL} = \frac{1}{n_k n_l} \sum_{i \in G_k, j \in G_l} d_{ij} \quad (3)$$

其中

$n_k, n_l$  分别表示类  $G_k$ , 类  $G_l$  的个数。

重心法也就是把类别的重心之间的欧式距离, 即

$$D_{ML} = (\bar{x}_k - \bar{x}_l)' (\bar{x}_k - \bar{x}_l) \quad (4)$$

其中:

$\bar{x}_k$  是  $G_k$  重心;

$\bar{x}_l$  是  $G_l$  重心。

在论文中, 我们将采用的是类平均法, 重心法和 ward 法。它的聚类步骤如下:

- 1) 定义样品与样品之间的距离, 计算样品的初始距离矩阵  $D(0)$ ;
- 2) 把  $D(0)$  中最短的两类合并, 即  $G_k$  与  $G_l$  合并为一个新类, 记作  $G_M$ , 即  $G_M = G_k \cup G_l$ ;
- 3) 计算新类  $G_M$  与其他任意类  $G_j$  的距离的递推公式如下:

$$D_{MJ} = \min \left\{ \min_{i \in G_k, j \in G_j} d_{ij}, \min_{i \in G_l, j \in G_j} d_{ij} \right\} = \min \{D_{KJ}, D_{LJ}\} \quad (5)$$

- 4) 将上述合并的距离矩阵算出, 记作  $D(1)$ , 重复上述步骤, 直到所有元素聚为一类为止。

其中:

$d_{ij}$  表示第  $i$  个样品与第  $j$  个样品的距离;

$G_1, G_2$  等表示类;

$D_{KL}$  为  $G_k$  与  $G_l$  之间的距离。

### 3.2. 实证分析

由类平均法聚类结果如图 1, 我们可以把类别整理为五类, 第一类为烟草制造业; 第二类为化学纤维制造; 第三类包含热力电力娱乐; 第四类铁路航空; 剩下为第五类。

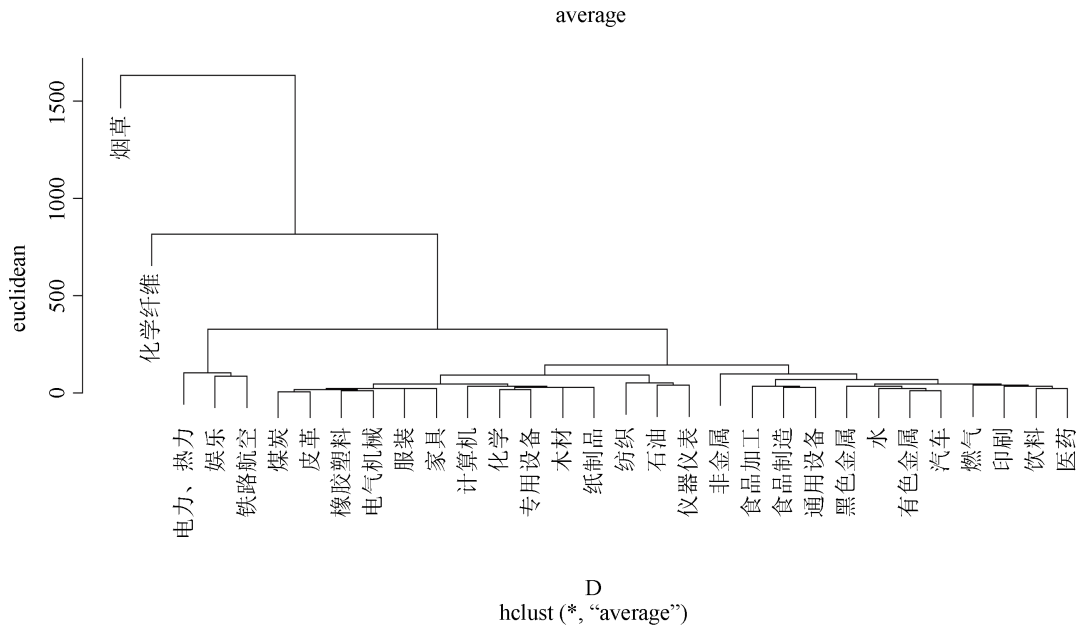


Figure 1. Class average method clustering  
图 1. 类平均法聚类

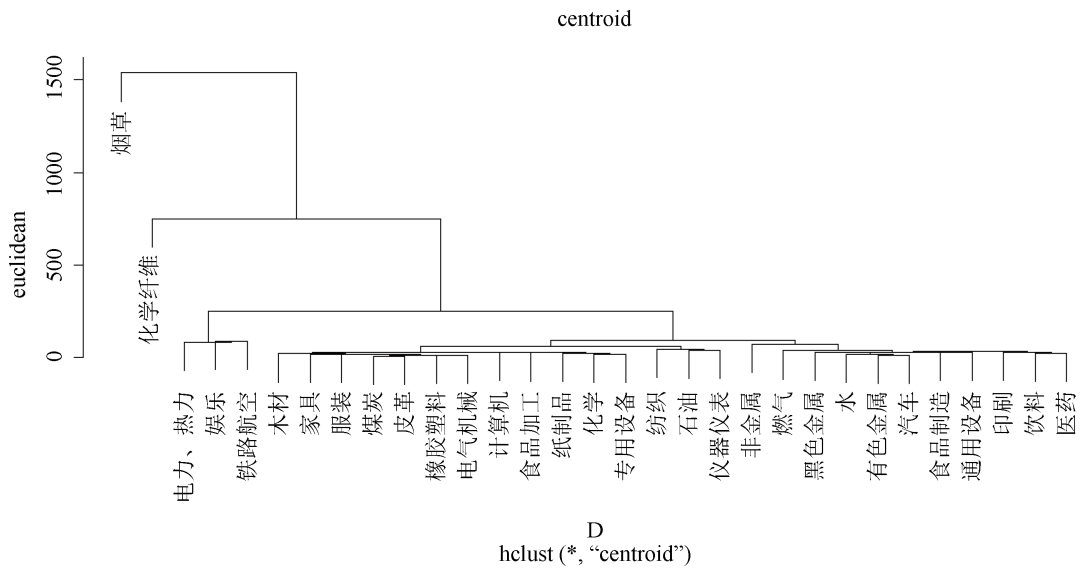


Figure 2. Center of gravity clustering  
图 2. 重心法聚类

由重心法聚类结果如图 2 所示，与类平均法相同，可以把上面结果分为五类。所包含的行业有一些细小差异，但是总体来讲与原有的类别出入都比较大。

Ward 法聚类结果如图 3，用 ward 法聚类与类平均法与重心法聚类结果有很大差异，ward 法聚类，可以明显的把聚类结果分为三类：第一类包含烟草、制品业、化学纤维制造业、电力热力行业、娱乐用品制造业、铁路航空及其他运输设备制造业；第二类包含黑色金属采矿业、水的生产和供应业、有色金属采矿行业，汽车制造业，非金属矿采行业，酒水饮料制造业；剩下的为第三类。与原来的分类也不太相符。

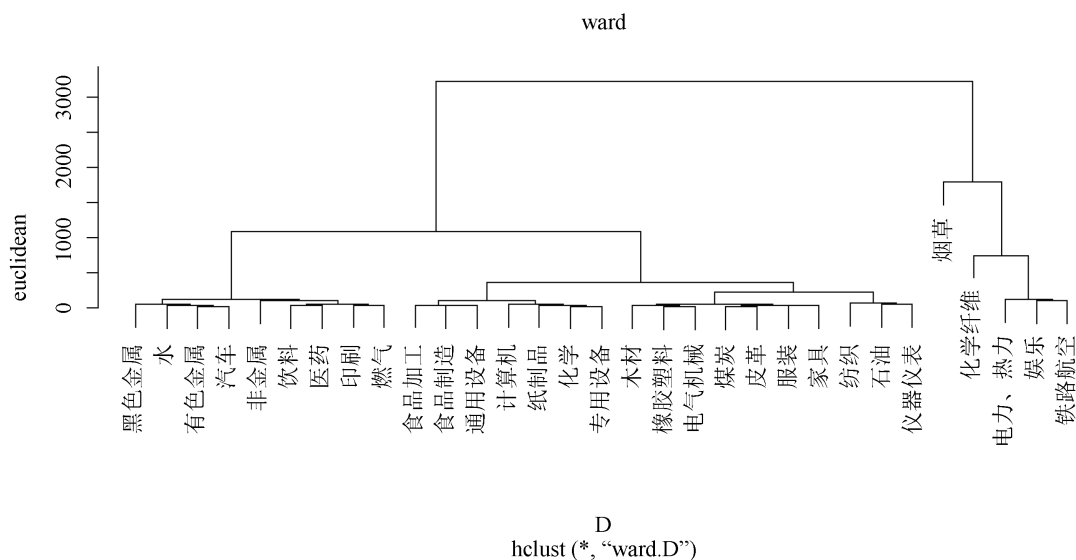


Figure 3. Word method clustering

图 3. Word 法聚类

总结三种方法聚类, 发现聚类的结果与原来的类别都不符合, 原来的类别是按照行业来分, 通过各类财务指标再来聚类, 可以看到烟草制造在各种方法中都单独被分为一类。从经济学的角度看, 云南省的烟草制造行业是云南省的较为核心的产业, 自 1987 年起连续 13 年, 云南烟草上缴的税利一直占云南财政收入的 70% 以上。2016 年全国烟草行业工作会议上指出, “十二五”期间, 全国烟草行业累计实现工商税利 47,680 亿元, 累计上缴国家财政 41,323 亿元; 云南省“十二五”云南省烟草产业创造的增加值约为 8880 亿元, 占全行业的 18.6%。根据财务指标来聚类, 把烟草分为一类是比较合理的。从 2014 年开始云南省电力热力供应行业形式越来越好, 各项财务指标反应的企业的财务状况较好, 被分为一类比较合理。

#### 4. 总结

本文通过使用多元统计的三种不同的聚类分析方法对云南省 2015 年规模以上工业企业的财务数据进行分析。结果显示云南省规模以上工业企业烟草制品与化学纤维行业在经济效应状况以及资本情况较其他行业都好, 可以把他们归为一类。对于铁路、航空航天和其他设备制造业在资产保值增值方面较其他行业要突出, 市场的认可度较高。对于一些生产日常生活用品的工业企业, 它们的流动资产周转情况较好, 这与生产的产品性质有很大关系。综合各项财务指标烟草制品业、化学纤维制造业两个工业企业的各种经济效益与财务状况都比较好。整体看来制造业的状况比采矿业与电力、热力生产和供应业的财务状况好。

#### 参考文献

- [1] 金硕元, 汤勤芳. 基于因子分析的温州市规模以上工业企业财务评价[J]. 温州职业技术学院学报, 2013(13): 70-74.
- [2] 余宏, 郭新磊, 宋世伟. 中国国有工业企业主要财务指标运行特征分析(2003-2014) [J]. 山东社会科学, 2016(S1): 75-80.
- [3] 费宇. 多元统计分析——基于 R [M]. 北京: 中国人民大学出版社, 2014: 14.
- [4] 王斌会. 多元统计分析及 R 语言建模[M]. 第 3 版. 广州: 暨南大学出版社, 2014: 82-87.

- [5] 张金昌. 现代企业财务分析[M]. 北京: 经济管理出版社, 1996: 8.
- [6] Johnson, R.A., Wichern, D.W. 实用多元统计分析[M]. 第 6 版. 陆璇, 叶俊, 译. 北京: 清华大学出版社, 2008: 524-548.
- [7] 刘文卿, 何晓. 群应用回归分析[M]. 北京: 中国人民大学出版社, 2001.
- [8] Kabacoff, R.I. R 语言实战[M]. 第 2 版. 王小宁, 刘擷芯, 黄俊文, 等, 译. 北京: 人民邮电出版社, 2016: 182-188.

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2169-2556, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [ass@hanspub.org](mailto:ass@hanspub.org)