

聚类分析法在企业经济效益分类中的应用

汤 强

(西华师范大学 数学与信息学院,四川 南充 637002)

【摘 要】对同行业、同系统的几个企业按经济指标进行分类需要科学、合理的分类方法。聚类分析法就是一种行之有效的数字聚类方法。根据四川省南充市丝绸企业经济效益指标的数据资料,应用聚类分析方法对南充市五大丝绸企业进行了经济效益分类。

【关键词】经济效益;聚类分析

【中图分类号】F272.5 **【文献标识码】**A **【文章编号】**1673-1891(2007)02-0073-03

在经济工作中,我们常常需要将同行业、同系统的企业按经济指标进行分类。如何对企业进行科学的分类呢?聚类分析就是一种行之有效的数字聚类方法。本文以四川省南充市五个丝绸企业的经济效益分类为例,讨论聚类分析在企业经济效益分类中的应用。

1 聚类分析的基本思想方法

聚类分析的基本思想是先建立衡量事物(类)间接近(相似)程度的统计量(比如常用距离、相关系数等),然后按照接近(相似)程度的大小,把样品逐一归并成类,将关系比较密切的归并到一个小的分类单位,关系疏远的则聚集到一个大的分类单位,直到所有的样品都聚集完毕为止,从而形成一个亲疏关系的谱系图,为实际的分类提供定量的依据。

1.1 数据资料及其标准化

为对南充市五大丝绸企业的经济效益进行分类,我们选择了反映企业经济效益的 5 个指标:每百元固定资产实现利税 (x_1)、每百元销售收入实现利税 (x_2)、产值利税率 (x_3)、资产利税率 (x_4)和全员劳动生产率 (x_5)。

南充市五大丝绸企业同年经济效益指标的原始数据如表 1:

表 1 南充市五大丝绸企业同年经济效益指标

	每百元固定资产 实现利税(元)	每百元销售收入 实现利税(元)	产值利税率 (%)	资产利税率 (%)	全员劳动生产率 (元/人·年)
阆中丝绸电厂(1)	6.58	12.45	11.72	8.59	9858
南部丝绸厂(2)	-7.49	-12.62	-11.77	-5.80	3013
嘉丽华集团(3)	-28.95	-49.55	-28.11	-26.62	2861
南泰集团(4)	-22.71	-104.83	-99.9	-22.22	-1945
美亚集团(5)	5.77	9.37	4.58	2.98	5847

在进行聚类分析之前,为消除各个经济指标变量所用量纲不同造成的影响,应先将原始数据 (x_{ij}) 进行标准化。

$$\text{标准化公式为 } x_{ij} = \frac{X_{ij} - X_i}{S_j} \quad (i = 1, 2, 3, 4, 5 \quad j = 1, 2, 3, 4, 5)$$

收稿日期 2007-03-18

作者简介 汤强(1975-)男,四川南充人,讲师,博士,主要从事数学教育以及应用统计学研究。

$$\text{其中 } \chi_j = \frac{1}{5} \sum_{i=1}^5 \chi_{ij} \quad S_j = \left[\frac{1}{4} \sum_{i=1}^5 (\chi_{ij} - \chi_j)^2 \right]^{1/2}$$

表 2 南充市五大丝绸企业同年经济效益指标的标准化数据

	每百元固定资产 实现利税(元)	每百元销售收入 实现利税(元)	产值利税率 (%)	资产利税率 (%)	全员劳动生产率 (元/人·年)
阆中丝绸电厂(1)	-0.62219	-0.27058	-0.33936	1.552458	1.3339
南部丝绸厂(2)	0.02996	-0.76766	-0.95456	-0.69147	-1.13265
嘉丽华集团(3)	1.770882	1.70166	1.60676	0.50264	-0.75953
南泰集团(4)	-0.74337	0.00129	0.25774	-0.50542	-0.04168
美亚集团(5)	-0.37323	-0.66471	-0.57057	-0.83032	0.60047

1.2 系统聚类方法

1.2.1 计算样点(企业)两两间的距离矩阵

根据标准化指标数据,计算同年度的样点(企业)间的欧氏距离矩阵,

$$D_0 = (d_{ij})_{5 \times 5}, \text{ 其中 } d_{ij} = \left[\sum_{t=1}^5 (\chi_{it} - \chi_{jt})^2 \right]^{1/2} \quad (i, j = 1, 2, 3, 4, 5)$$

经计算,得到 1998 年度南充市五个企业(样点)之间的距离矩阵为:

$$\begin{pmatrix} 0 & & & & \\ 4.638 & 0 & & & \\ 14.998 & 4.281 & 0 & & \\ 26.326 & 10.561 & 5.184 & 0 & \\ 0.993 & 1.861 & 10.925 & 20.233 & 0 \end{pmatrix}$$

1.2.2 进行系统聚类,其步骤如下:

- (1) 初始聚类 现将每个样点(企业)各作为一类,这时类与类之间的距离总是等于样点之间的距离。
- (2) 并距离最短的两类为一个新类。
- (3) 计算新类与此时待并的各类之间的距离。

对于不同的系统聚类方法,其具体计算类间距离的递推公式各不相同,下面以常用的类平均法和重心法为例,进行说明。

设某一步将第 p 类 G_p (含有 n_p 个样点)和第 q 类 G_q (含有 n_q 个样点)合并为一个新类 G_r (含有 n_r 个样点, $n_r = n_p + n_q$),则新类 G_r 与待并类 G_k 间的距离 $D(r, k)$ 的递推公式分别为:

$$1^\circ \text{ 类平均法 } \quad D_c(r, k) = \frac{n_p}{n_r} D_c(p, k) + \frac{n_q}{n_r} D_c(q, k)$$

$$2^\circ \text{ 重心法 } \quad D_c^2(r, k) = \frac{n_p}{n_r} D_c^2(p, k) + \frac{n_q}{n_r} D_c^2(q, k) - \frac{n_p \cdot n_q}{n_r^2} D_c^2(p, q)$$

- (4) 聚类结束,即所有的样点最终合并为一个大类,否则就回到步骤(3)继续并类。

1.2.3 聚类结束后,根据并类结构,画出聚类分析关系图(谱系图)。

2 聚类分析结果及讨论

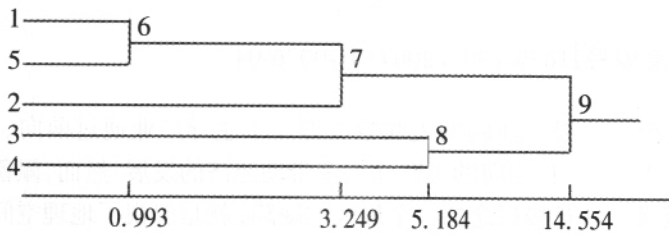
我们利用前面所述的类平均法和重心法对南充市五个丝绸企业进行了聚类分析。

2.1 聚类过程

2.1.1 类平均法

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈	G ₉
G ₁	0								
G ₂	4.638	0							
G ₃	14.998	4.281	0						
G ₄	26.326	10.561	5.184	0					
G ₅	0.993	1.861	10.923	20.233	0				
G ₆	×	3.250	12.962	23.280	×	0			
G ₇	×	×	10.068	19.04	×	×	0		
G ₈	×	×	×	×	×	×	14.554	0	
G ₉	×	×	×	×	×	×	×	×	0

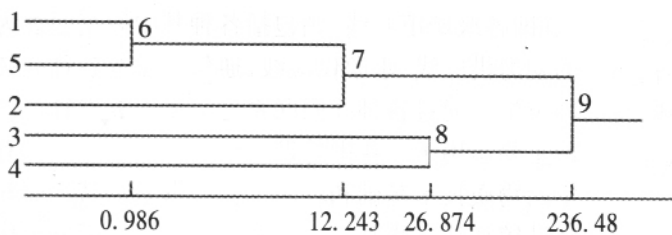
根据上述聚类过程,画出聚类谱系图如下:



2.1.2 重心法

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈	G ₉
G ₁	0								
G ₂	21.418	0							
G ₃	224.946	18.327	0						
G ₄	693.059	111.535	26.874	0					
G ₅	0.986	3.463	119.356	409.374	0				
G ₆	×	12.243	171.898	550.982	×	0			
G ₇	×	×	117.983	368.448	×	×	0		
G ₈	×	×	×	×	×	×	236.483	0	
G ₉	×	×	×	×	×	×	×	×	0

根据上述聚类过程,画出聚类谱系图如下:



2.2 结果分析

从上面得到的聚类谱系图,可以看出,在同一个度中,所用两种不同的系统聚类法得到的结果是基本一致的。根据五个丝绸企业经济效益可以将其分为三类:阆中丝绸电厂和美亚集团为一类,经济效益较好;南部丝绸厂为一类,经济效益次之;嘉丽华和南泰集团为另一类,经济效益较差。

随后几年的经济运行结果表明,阆中丝绸电厂、美亚持续好转,南部丝绸厂亏损减少,而嘉丽华和南泰集团则面临更大困境,其中嘉丽华集团宣告破产。这就进一步证明了上述分类结果的正确性。

参考文献:

[1]张尧庭,方开泰.多元统计分析[M].北京:科学出版社,1982.

(下转 79 页)

参考文献：

- [1]李斌. 嘉陵江干流四川段旅游资源开发初步探析[EB/OL]. WWW. CHINA - REGION. COM. 2006 03 17.
 [2]杨万钟主编,华东师范大学等四校合编. 经济地理学导论[M]. 上海:华东师范大学出版社,1999.
 [3]隋鹏飞,胡碧玉. 嘉陵江干流地区产业合理布局探讨[J]. 小城镇建设, 2004(9).
 [4]黄元斌. 嘉陵江流域产业结构调整及对策研究(硕士毕业论文)[D]. 2005 6 :77 - 79.

Research on the Industrial Distribution for the Jialing River Valley

HUANG Yuan - bin

(Department of Political Science & Law, Sichuan University of Science & Engineering Zigong Sichuan 643000)

Abstract: Considering industrial distribution for the Jialing River valley , we should foster growth pole , such as Mianyang and Nanchong firstly, and attach importance to the point and axis model , so as to make Chuanshan railway , Chengda railway and Jialing River axis and form distribution model like the pattern of “干”. At the same time , considering the distinguishing feature of Jialing River valley economy and industrial development , we should apply circle and gradually development model, making Mianyang and Nanchong as center, and making Jialing River and Chuanshan railway as axis , in order to promote Jialing River valley economy development healthily and lastingly.

Key words: Jialing river valley ;Industrial distribution ; Distribution model

(责任编辑:张荣萍)

(上接 75 页)

- [2]陈希儒. 数理统计引论[M]. 北京:中国科学技术大学出版社,1999.
 [3]Anderberg. M. R. Cluster Analysis for Application[M]. Academic Press, 1973.

The Application of Cluster Analysis on the Classification of Enterprises' Economical Profit

TANG Qiang

(School of Mathematics and Information, China West Normal University ,Nanchong, Sichuan 637002)

Abstract: Scientific methods are needed to classify some enterprises in the same line according to economical profit. Cluster analysis is an effective approach. Based on the data of the economical profit of enterprises, the article makes cluster analysis to five silk enterprises in Nanchong.

Key words: Economical profit; Cluster analysis

(责任编辑:张荣萍)