

doi:10.16055/j.issn.1672-058X.2015.0006.003

# 基于聚类与主成分分析的重庆库区经济发展研究\*

刘宇琴

(重庆工商大学 管理学院,重庆 400067)

**摘要:**选取重庆库区万州、云阳、江津等 17 个区县作为研究目标,以重庆统计年鉴发布的数据为基础,利用主成分分析法选取出代表性的主成分,并对主成分的得分进行聚类;结论显示:重庆库区分为重庆市区为中心和万州为中心的两个主成分,区域中心的发展辐射作用较为明显;最后提出重庆库区经济发展相关建议。

**关键词:**重庆库区;主成分分析;聚类分析;经济发展

**中图分类号:** O29      **文献标识码:** A      **文章编号:** 1672-058X(2015)06-0011-06

2014 年 4 月,李克强总理在重庆主持召开沿江 11 省市的长江经济带座谈会,他强调中西部具有经济发展最大的回旋余地,建设长江经济带,就是要构建沿海与中西部相互支撑、良性互动的棋局,他反复强调重庆库区经济带建设的重要性。为此,对重庆库区的经济发展进行研究,具有很大的意义。重庆库区东起巫山县,西至江津区,南起武隆县,北至开县,包括了重庆主城区 4 个区县在内的 17 个区县,横跨了都市经济圈和三峡库区生态经济区,经济发展不平衡,特别是郊区县经济发展相对滞后,库区内的经济发展水平表现出巨大的差异,用主成分分析的方法从代表一个城市经济发展水平的 7 个指标中选取一至两个主成分,再用聚类分析法对 17 个区县进行聚类,以此来评价地区经济发展水平和区域差异,对指导现阶段经济发展和战略、平衡区域差异是很有必要的。

## 1 指标选取

李芝芳等(2009)认为,由于经济社会发展的复杂性,使得经济发展的评价指标很复杂,认为要遵循简明性、客观性、实用性和动态性的原则<sup>[1]</sup>;惠俊刚等(2007)认为为了更好的反映区域经济的差异,指标的选取应具有综合性、代表性、可比性、可得性和简洁性,特别是在聚类分析中,一般要求样本的容量大于变量的个数<sup>[2]</sup>;熊小刚(2008)运用聚类分析对环鄱阳湖经济圈进行研究,选取了具有弱相关性的指标,比如国内生产总值、财政总收入、工业总产值、经济密度、农民人均收入、城乡居民储蓄年末余额等作为指标体系<sup>[3]</sup>;通过对文献的研究和分析,结合指标选取的可获得性、代笔性、综合性和简洁性的原则<sup>[4]</sup>选取了财政收入( $x_1$ ),地区生产总值( $x_2$ ),工业总产值( $x_3$ ),农林渔牧业产值( $x_4$ ),全社会固定资产投资( $x_5$ ),人均 GDP( $x_6$ ),人均第三产业总值( $x_7$ )和社会消费品零售总额( $x_8$ )个相关指标,地区生产总值、人均 GDP 和财政收入反映了一个地区经济发展的综合水平;全社会固定资产投资反映了社各界对生产投资的重视程度及实施;农林渔牧产值、工业产值一级人均第三产业总产值则代表了一个地区的第一、第二和第三产业的发展状况。具体指标结果见表 1。

收稿日期:2014-10-23;修回日期:2014-11-20.

\* 基金项目:重庆市科学技术委员会科技创新服务(软科学)项目(CSTC2012CX-RKXA00098).

作者简介:刘宇琴(1991-),女,四川泸州人,硕士研究生,从事人力资源管理研究.

表 1 重庆库区各区县主要经济指标

|     | $X_1$     | $X_2$     | $X_3$      | $X_4$   | $X_5$     | $X_6$  | $X_7$  | $X_8$     |
|-----|-----------|-----------|------------|---------|-----------|--------|--------|-----------|
| 江北区 | 803 328   | 5 277 599 | 5 661 935  | 27 276  | 4 435 615 | 66 519 | 42 233 | 3 406 956 |
| 南岸区 | 610 299   | 4 655 578 | 7 512 235  | 71 160  | 4 034 104 | 58 035 | 20 951 | 3 088 453 |
| 渝北区 | 1 123 342 | 8 793 249 | 18 393 690 | 354 971 | 6 102 581 | 62 372 | 21 187 | 3 402 995 |
| 巴南区 | 340 153   | 4 208 456 | 4 836 305  | 556 310 | 4 291 536 | 44 749 | 17 780 | 1 637 706 |
| 涪陵区 | 378 100   | 6 305 288 | 8 692 631  | 622 306 | 4 393 616 | 57 794 | 18 360 | 1 538 821 |
| 长寿区 | 252 376   | 3 364 092 | 5 494 899  | 452 628 | 2 993 680 | 42 852 | 13 396 | 789 843   |
| 江津区 | 307 143   | 4 260 134 | 7 069 766  | 885 738 | 3 507 630 | 34 043 | 8 752  | 1 507 456 |
| 万州区 | 417 146   | 6 628 618 | 5 097 705  | 696 568 | 4 104 232 | 42 016 | 16 660 | 1 891 863 |
| 丰都县 | 85 217    | 1 110 773 | 674 389    | 333 078 | 2 242 428 | 17 519 | 6 425  | 439 820   |
| 忠 县 | 95 570    | 1 567 970 | 436 988    | 424 108 | 1 501 773 | 21 045 | 7 759  | 484 289   |
| 开 县 | 117 092   | 2 295 507 | 1 543 837  | 637 508 | 2 035 175 | 19 768 | 7 012  | 952 247   |
| 云阳县 | 80 003    | 1 266 274 | 521 008    | 479 444 | 1 453 735 | 13 930 | 5 279  | 565 127   |
| 奉节县 | 111 590   | 1 445 675 | 192 594    | 458 877 | 1 676 958 | 17 853 | 8 045  | 409 901   |
| 巫山县 | 60 868    | 703 497   | 175 789    | 248 377 | 700 104   | 14 529 | 6 295  | 257 424   |
| 巫溪县 | 47 529    | 531 140   | 133 574    | 189 926 | 949 727   | 13 055 | 5 296  | 187 301   |
| 武隆县 | 86 876    | 984 028   | 351 631    | 235 866 | 1 259 071 | 28 188 | 13 093 | 310 890   |
| 石柱县 | 71 426    | 931 033   | 618 205    | 278 393 | 1 322 419 | 22 612 | 7 889  | 353 779   |

数据来源:重庆统计年鉴(2013)

由表 1 可知利用主成分分析和聚类分析对重庆库区经济进行评价,并给出相应的建议,对于今后库区的长远发展具有理论上的现实意义。

## 2 数据分析过程

### 2.1 主成分的选择

利用 SPSS20.0 对数据进行处理,首先对数据做标准化处理,计算相关系数矩阵,计算特征根和特征向量,方差贡献率,见表 2。

表 2 主成分特征根和贡献率

| 序号 | 合计    | 贡献率/%  | 累计贡献率/% |
|----|-------|--------|---------|
| 1  | 6.113 | 76.411 | 76.411  |
| 2  | 1.323 | 16.542 | 92.953  |
| 3  | 0.338 | 4.224  | 97.177  |
| 4  | 0.114 | 1.427  | 98.604  |
| 5  | 0.048 | 0.598  | 99.201  |
| 6  | 0.033 | 0.406  | 99.608  |
| 7  | 0.029 | 0.366  | 99.974  |
| 8  | 0.002 | 0.026  | 100     |

表 3 主成分载荷

| 项目                 | 成分     |        |
|--------------------|--------|--------|
|                    | 1      | 2      |
| 财政收入( $x_1$ )      | 0.969  | -0.098 |
| 地区生产总值( $x_2$ )    | 0.946  | 0.282  |
| 工业总产值( $x_3$ )     | 0.906  | 0.235  |
| 农林渔牧产值( $x_4$ )    | -0.037 | 0.970  |
| 全社会固定资产投资( $x_5$ ) | 0.963  | 0.209  |
| 人均 GDP( $x_6$ )    | 0.959  | -0.090 |
| 人均第三产业总值( $x_7$ )  | 0.830  | -0.412 |
| 社会消费品零售总额( $x_8$ ) | 0.959  | -0.127 |

一般取累计贡献率达到80%以上少数几个主成分就可以代表原来多个指标的绝大部分信息,由表2可见,取第一,第二个主成分的累计贡献率即可达到92%,且重庆库区的区县有很大一部分农业产业,所以选取代表农林牧渔产业总值的第二主成分也作为分析的目标,第一、第二主成分与其他成分之间的相关系数如表3。

由表3可知,第一主成分与财政收入( $x_1$ )、地区生产总值( $x_2$ )、工业生产总值( $x_3$ )、全社会固定资产投资( $x_5$ )、人均GDP( $x_6$ )、人均第三产业总值( $x_7$ )、社会消费品零售总额( $x_8$ )有很大的正相关,与农林渔牧产值( $x_4$ )呈负相关,而第二主成分与农林渔牧产值( $x_4$ )有显著的正相关性,达到0.97,与其他指标相关性较弱或呈负相关。

由第一主成分和第二主成分的特征向量矩阵可以得到两个主成分:

$$z_1 = 0.39x_1 + 0.38x_2 + 0.37x_3 - 0.01x_4 + 0.39x_5 + 0.39x_6 + 0.34x_7 + 0.39x_8$$

$$z_2 = -0.09x_1 + 0.25x_2 + 0.2x_3 + 0.84x_4 + 0.18x_5 - 0.08x_6 - 0.36x_7 - 0.11x_8$$

对于第一主成分 $z_1$ ,除了 $x_4$ (农林渔牧产值)以外,其他的指标的系数都在0.35左右,而这几个指标包括了财政收入、地区生产总值、工业总产值等7个指标,这7个指标都是经济发展中的代表性指标,包括了第二产业和第三产业的综合水平、政府财政收入和人们消费能力和财富水平等。所以第一主成分代表了一个地区的经济发展综合水平;第二主成分与 $x_4$ 的系数高达0.84,远远高于其他指标的系数,而与 $x_7$ (人均第三产业总值)的系数也为0.36,也具有较高的相关性,而农林渔牧则代表了第一产业的发展综合情况,农林牧渔业广义包括了除第一产业的传统种植行业以及饲养业、林业之外,也包含了农用机械、农业服务、新型农业发展、农副产品加工等新型农业与第二和第三产业的结合,所以,第二主成分代表了以第一产业为主的农业及其新型第三产业。根据以上两个主成分的线性组合以及标准化后的原始数据,可以得到两个主成分的得分,如表4。

表4 第一和第二主成分得分及排名

| 区县  | 第一主成分得分      | 第一主成分排名 | 第二主成分得分      | 第二主成分排名 |
|-----|--------------|---------|--------------|---------|
| 江北区 | 4.024 380 7  | 2       | -2.578 292 2 | 17      |
| 南岸区 | 2.657 316 0  | 3       | -1.494 387 2 | 16      |
| 渝北区 | 5.492 578 9  | 1       | 0.412 920 4  | 6       |
| 巴南区 | 1.176 502 8  | 6       | 0.576 030 4  | 5       |
| 涪陵区 | 2.124 239 2  | 4       | 1.105 569 2  | 3       |
| 长寿区 | 0.174 424 2  | 8       | 0.282 541 4  | 8       |
| 江津区 | 0.507 313 8  | 7       | 2.218 582 4  | 1       |
| 万州区 | 1.603 106 7  | 5       | 1.313 936 6  | 2       |
| 丰都县 | -1.842 652 6 | 12      | -0.190 666 7 | 11      |
| 忠 县 | -1.826 240 6 | 11      | 0.019 369 7  | 10      |
| 开 县 | -1.370 698 2 | 9       | 0.962 896 3  | 4       |
| 云阳县 | -2.111 282 5 | 15      | 0.316 924 4  | 7       |
| 奉节县 | -1.883 954 5 | 13      | 0.151 413 9  | 9       |
| 巫山县 | -2.481 716 5 | 16      | -0.693 911 3 | 13      |
| 巫溪县 | -2.556 184 7 | 17      | -0.843 193 9 | 14      |
| 武隆县 | -1.700 870 4 | 10      | -0.977 810 3 | 15      |
| 石柱县 | -1.986 261 6 | 14      | -0.581 920 5 | 12      |

在表 4 中,第一主成分得分最高的是渝北区,达到了 5.49 左右,第二是江北区,有 4.02 分,都远远高于其他城市,然后是南岸区、涪陵、万州等区县,经济发展的综合水平都比较高,得分最低的是巫溪县,只有 -2.55 分;第二主成分中得分最高的是江津区,其得分达到了 2.21 分,第二的是万州区,表明这两座城市的农林牧渔业较发达。

## 2.2 聚类分析

对第一主成分和第二主成分进行标准化处理后,选择欧式距离,采用最短距离法进行聚类,对重庆库区的 17 个区县进行经济发展综合水平和农林牧渔业发展综合水平进行水平聚类分析,得到最短距离聚类谱系图(如图 1、2)

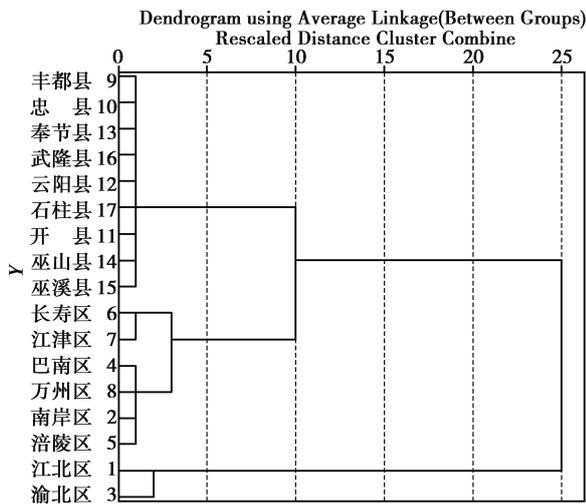


图 1 第一主成分聚类树状图

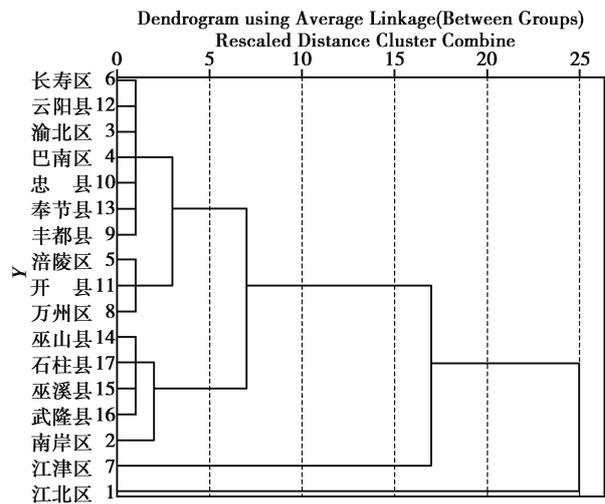


图 2 第二主成分聚类树状图

图 1 是第一主成分的聚类图,由它把库区经济发展综合水平分为四类:第一类是江北区和渝北区,第二类是巴南区、万州区、涪陵区和南岸区,第三类是长寿区和江津区,第四类是丰都县,石柱县等其他区县。根据第一主成分的得分,满足得分的依据,江北和渝北的第一主成分得分都远高于其他区县,经济发展综合水平较高,处于第二层次的长寿、万州等区经济发展综合水平较渝北和江北稍弱,但还是远远领先于其他得分较低的城市。

根据第二主成分的得分对第二主成分进行聚类,如图 2。根据图 2 对第二主成分的聚类可以把重庆库区区县的第二主成分分为四类:第一类是得分最低的江北区,第二类是得分最高的江津区,第三类是得分较低但远远高于江北的巫山县、石柱县、巫溪县、武隆县和南岸区,第四类是处于之间的长寿区、云阳县、渝北区、万州区等。

## 3 结果分析

### 3.1 第一主成分分析

根据分析,第一主成分的贡献率达到了 76%,位于主城区的渝北区和江北区排名第一和第二,为第一类,且分数都远远超过其他城市。第一主成分的第二类是巴南区、涪陵区、万州区和南岸区,这几个区在第一主成分得分方面均位于前几位,低于渝北和江北,但也远远高于其他区县,除了万州外,都是重庆 1 h 经济圈的组成部分,受到经济发展的辐射作用,经济发展迅速。第三类是长寿区和江津区,经济发展综合水平较上述城市较弱,长寿和江津都是传统的工业城市,特别是江津,工业较发达,有著名的珞璜工业园和德感工业园,其机械、汽摩产业、农用机械、造纸等均在西部处于领先地位。

最后一类是得分较低的其他区县,这些城市基本都位于渝东南翼和渝东北翼,是传统的经济较薄弱的城市,工业基础差,现代化程度较低,交通不便利。

### 3.2 第二主成分分析

第一类是得分最低的江北区,由于其处于主城区核心,农林牧渔类产业面积较少,且农村人口少,城市化水平很高,其分为第一类。

第二类是得分最高的江津区,江津区的农林牧渔产业发展水平很高,最近新建的江津现代农业园区是全国性的农业园区,农业现代化水平高,农产品附加值较高,产业集中规模化发展,是西南地区乃至全国领先的农业园区,大大促进了现代农业的发展。

第三类是处于中间水平的长寿区、云阳县、渝北区、万州区、涪陵区等,处于渝东北翼的万州、云阳、奉节、忠县等是全国柑橘的重要产地,占据了农林牧渔类产值的大部分,而且受万州经济的影响辐射,这几个城市的农业现代化程度有了明显的提高。

第四类是发展水平较低的巫山县、石柱县、巫溪县、武隆县和南岸区,南岸区是处于重庆主城核心,农业面积小,发展较弱,其余的巫山、石柱、巫溪和武隆农业人口较多,但受其经济综合水平的影响,农业现代化程度很低,并没有形成产业化,水平较低。

## 4 结论及建议

对于第一主成分和第二主成分分析的结果和聚类的结果来看,渝东北的重庆库区腹心的城市经济水平都偏低,除了万州外,其余的区县都与1 h经济圈的城市呈现巨大的差异,处于1 h经济圈的区县的农林牧渔产业也明显高于渝东北库区腹心,江津更是处于第一位;三峡库区重庆段的全县在经济、农业方面的差异甚至大于全国平均的差异程度,是典型的大城市、大农村的格局。对此提出相关建议:

### 4.1 加大对重庆库区的投资力度,特别是库区腹心

以投资来带动库区经济、农业协调发展,对外部投资依赖很强的重庆库区要实现区域经济、农业协调发展,国家政策性的投资的方式和力度仍然是促进区域协调发展的主要因素。特别是对于渝东北的区县城市,更应该加大投资力度,形成区县间资源互补,协调发展。

### 4.2 发挥好区域中心城市的辐射作用

重庆库区的重庆1 h经济圈内,主城区的快速发展明显带动了1 h经济圈的城市快速发展,取得了非常明显的效果。而渝东北远离主城区,周围没有像重庆一样的特大城市,万州作为唯一的区域中心城市,经济综合水平和农业综合水平都远高于周围的开县、云阳、石柱、忠县等城市,要充分发挥好其大中城市的辐射作用,加大对万州的扶持力度,加大万州的政策支持和优惠。

### 4.3 协调好库区城市的产业结构,加强对库区的产业扶持

三峡库区落后,自我发展能力很弱,国家应加强对库区的产业扶持,库区各个城市要有自己城市的支撑产业,同时也要主要好产业的协调<sup>[8]</sup>,1 h经济圈的库区城市大都经济基础较好,特别是江津和长寿、涪陵,工业基础好,产业协调完善,而渝东北的库区城市没有自身独特的产业,大都是三个产业都比较落后。国家在对渝东北库区腹心的扶持中,不能忽视工业的发展,工业化是经济发展不能逾越的阶段,而农林牧渔等第一产业也要协调发展,同时,可以发挥自身的优势产业,找到经济发展的突破口。

### 4.4 突出区位优势,促进库区现代农业发展

城市的郊区如渝北、万州、涪陵等区县区位优势较优越,农业基础好,以农村城镇化为突破口,形成都市农业和现代农业经济开发区,而渝东北的云阳、奉节等城市现代农业基础薄弱,应充分利用好其农业资源丰富特点,形成以立体的农业经济综合区,开展观光农业,与一二产业协调发展。

(下转第35页)