

基于聚类分析模型的重庆区域经济差异性分析

王继梅

贵州大学, 数学与统计学院, 贵州 贵阳

收稿日期: 2022年6月19日; 录用日期: 2022年7月14日; 发布日期: 2022年7月21日

摘要

我国各地区在经济发展过程中, 经济发展不协调。适当的区域经济差异能够使资源得到合理的优化和配置, 促进经济的发展。然而, 区域经济差异过大, 地区的基尼系数会提高, 容易造成社会矛盾, 不利于和谐社会的建设。本文以重庆市38个区(县)所辖范围为研究对象, 首先, 从影响区域经济发展的众多因素中, 基于不同的变量选择方法构建不同的指标体系, 通过Lasso变量选择法、逐步回归法以及自适应Lasso这三种变量选择方法的比较, 发现Lasso变量选择法简化模型的同时, 预测精度也更好, 故将其应用在重庆区域经济发展研究中。其次, 利用主成分分析法和聚类分析对区域的经济差异进行了深入研究, 结果表明, 以涪陵为界限, 重庆市东西部经济发展呈现两极分化。最后, 本文基于重庆市目前的经济发展状况, 给出了一些区域经济协调发展的建议和对策。

关键词

区域经济, Lasso, 主成分分析, 聚类分析

Analysis of Chongqing's Regional Economic Differences Based on Cluster Analysis Model

Jimei Wang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Jun. 19th, 2022; accepted: Jul. 14th, 2022; published: Jul. 21st, 2022

Abstract

In the process of economic development, the economic development of various regions in China is

not coordinated. Appropriate regional economic differences can make resources get reasonable optimization and allocation, and promote economic development. However, if the regional economic difference is too large, the regional Gini coefficient will increase, which is easy to cause social contradictions and is not conducive to the construction of a harmonious society. In this paper, 38 districts (counties) of Chongqing municipality are taken as the research object. Firstly, from the many factors affecting regional economic development, different index systems are constructed based on different variable selection methods. Through the comparison of three variable selection methods, Lasso variable selection method, stepwise regression method and adaptive Lasso, it is found that the Lasso variable selection method simplifies the model and has better prediction accuracy at the same time, so it is applied in the study of Chongqing regional economic development. Secondly, principal component analysis and cluster analysis are used to study the regional economic differences. The results show that with Fuling as the boundary, the economic development of the east and the west of Chongqing presents polarization. Finally, based on the current economic development of Chongqing, this paper gives some suggestions and countermeasures for the coordinated development of regional economy.

Keywords

Regional Economy, Lasso, Principal Component Analysis, Cluster Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

区域经济发展不平衡是国际性的普遍现象,世界各国都不同程度地面临经济发展不平衡问题的困扰。众所周知,经济社会发展的平衡是相对的,其不平衡是绝对的。暂时的经济不平衡有利于保持竞争的压力与活力,进而促进生产要素的合理化。但长期的经济发展不平衡,会影响社会的公平,会阻碍国民经济的平稳发展[1]。

中国是一个地域辽阔、人口众多的发展中国家,也是世界上自然地理、经济社会、人口资源差异最大的国家之一。我国区域经济发展的差异性是由自然、历史、政治、地理、社会、空间差异等诸多因素长期作用而成[2][3]。改革开放以来,我国东部地区经济实力明显比西部地区的经济力强,这与我国当时“两个大局”的战略思想有关。当时,邓小平提出在生产布局上向东部倾斜的政策,形成东-中-西梯度发展的经济发展模式。随着国家经济的不断发展,当初的发展模式已不能适应我国现在的国情。我国东西部经济发展的差异化不利于社会的稳定和全国统一市场的形成,从而影响了经济的可持续发展[4][5][6]。

因此,加快西部地区经济发展成为我国解决社会、经济矛盾从而实现共同富裕的重大问题。如何加快西部的发展,缩小区域经济差异,是一个值得深入研究的重大战略问题,也成为区域经济学研究的热点内容之一。重庆是中国大规模开发西部战略的重点和前沿阵地,也是西部地区新的经济增长点[7]。

重庆的腾飞,是西部大开发的一个重要组成部分。重庆从1997年直辖以来,推行了一系列的对外开放和经济发展措施,实现了国民经济的快速发展。与此同时,重庆市内部各区域间的经济差异也日趋明显。重庆主城区经济相对较为发达,处于特大城市经济生活水平。但是其中某些区(县)经济发展相对较慢,

生活条件也没有得到很好的改善, 这不利于重庆市整体经济的发展, 也不利于提高人民的生活水平和社会的安定团结。

因此研究重庆市的区域经济差异, 不但可以提高重庆市的整体经济水平、缩小城乡差距, 加快重庆市的城市化进程, 维护社会安定团结, 还可以作为区域经济差异研究的典范, 具有较好的理论研究意义和现实意义[8]。而研究区域经济的差异性, 需要有一套科学有效的区域经济发展评价指标。这些指标的评价结果是政府部门判断该区域现有的经济水平和制定该区域经济发展策略的重要依据。对于各级政府部门而言, 想制定推动当地经济发展的相关策略, 提高居民生活水平和生活幸福指数, 就必须了解当地城区经济的发展状况, 因此很有必要提出一些合理的统计方法来对城区经济发展指标进行研究。

目前区域经济分析常用的数学方法有线性规划、决策论和多元统计分析等。其中多元统计主要的分析方法有聚类分析、主成分分析、因子分析和回归分析等[9] [10]。本文利用 Lasso 变量选择方法[11] [12]、主成分分析[13]和聚类分析, 从现实的角度分析重庆区域经济差异的现状及成因, 进而对缩小区域经济差异、加快西部地区经济发展提出有针对性的政策措施。

2. 相关理论介绍

2.1. Lasso 变量选择

在大数据时代, 数据来源呈现多源、海量和高维的特征, Lasso (Least absolute shrinkage and selection operator)方法因能从高维变量中高效提取出关键变量, 并能有效解决变量间的多重共线性问题、提高模型的解释精度而获得高度评价和广泛应用[11] [12]。

给定观测数据 $(Y_i; X_{i1}, \dots, X_{ip})$, $i = 1, 2, \dots, n$, 一个多元线性回归模型可设为

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

其中误差项 ε_i 满足 Gauss-Markov 条件, 令

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2)$$

则公式(1)可表示为如下的矩阵形式:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

Lasso 方法的基本思想是在普通最小二乘估计的基础上引入惩罚因子对回归系数 $\boldsymbol{\beta}$ 添加 L_1 范数作为约束, 公式如下:

$$\hat{\boldsymbol{\beta}} = \arg \min \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad s.t. \quad \sum_{j=0}^p |\beta_j| \leq s \quad (4)$$

等价地, 可表示为

$$\hat{\boldsymbol{\beta}} = \arg \min \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|_1 \quad (5)$$

其中 $s \geq 0$ 是一个调节参数, 控制压缩的强度, 且由于 L_1 范数约束的自然属性, 使得回归模型的系数可能为 0, 回归系数等于 0 所对应的自变量与因变量相关性很弱, 将被剔除, 从而实现变量选择。

2.2. 主成分分析

主成分分析(principal component analysis)是将多指标化为少数几个综合指标的一种统计分析方法,是由 Pearson (1901)针对非随机变量提出,后来 Hotelling (1933)将此方法推广到随机向量的情形。主成分分析将多个变量化成少数几个主成分实现数据降维,这些主成分能够反映原始变量的绝大部分信息,通常表示为原始变量的线性组合[10]。

设 \mathbf{X} 是 p 维随机变量,并假设 $\boldsymbol{\mu} = E(\mathbf{X})$, $\boldsymbol{\Sigma} = \text{var}(\mathbf{X})$ 。考虑如下线性变换

$$\begin{cases} Z_1 = \mathbf{a}_1^T \mathbf{X} \\ Z_2 = \mathbf{a}_2^T \mathbf{X} \\ \vdots \\ Z_p = \mathbf{a}_p^T \mathbf{X} \end{cases} \quad (6)$$

易见

$$\text{var}(Z_i) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_i, \quad i = 1, 2, \dots, p \quad (7)$$

$$\text{cov}(Z_i, Z_j) = \mathbf{a}_i^T \boldsymbol{\Sigma} \mathbf{a}_j, \quad i, j = 1, 2, \dots, p, \quad i \neq j \quad (8)$$

若希望 Z_1 的方差达到最大,那么 \mathbf{a}_1 是约束优化问题

$$\max \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a} \quad \text{s.t.} \quad \mathbf{a}^T \mathbf{a} = 1 \quad (9)$$

的解。因此, \mathbf{a}_1 是 $\boldsymbol{\Sigma}$ 最大特征值(不妨设为 λ_1)的特征向量。此时,称 $Z_1 = \mathbf{a}_1^T \mathbf{X}$ 为第一主成分。类似地,希望 Z_2 的方差达到最大,且要求 $\text{cov}(Z_1, Z_2) = \mathbf{a}_1^T \boldsymbol{\Sigma} \mathbf{a}_2 = 0$ 。由于 \mathbf{a}_1 是 λ_1 的特征向量,故选择的 \mathbf{a}_2 应与 \mathbf{a}_1 正交。类似于前面的推导, \mathbf{a}_2 是 $\boldsymbol{\Sigma}$ 第二大特征值(不妨设为 λ_2)的特征向量,称 $Z_2 = \mathbf{a}_2^T \mathbf{X}$ 为第二主成分。

一般情况下,对于协方差阵 $\boldsymbol{\Sigma}$,存在正交阵 \mathbf{Q} ,将它化为对角阵,即

$$\mathbf{Q}^T \boldsymbol{\Sigma} \mathbf{Q} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix} \quad (10)$$

且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 。则矩阵 \mathbf{Q} 的第 i 列就对应于 \mathbf{a}_i ,相应的 Z_i 为第 i 主成分。

2.3. 聚类分析

聚类分析(cluster analysis)是一类将数据所研究对象进行分类的统计方法。这一类方法的共同特点是:事先不知道类别的个数与结构;据以进行分析的数据是对象之间的相似性(similarity)或相异性(dissimilarity)的数据。将这些相似(相异)性数据看成是对象之间的“距离”远近的一种度量,将距离近的对象归入一类,不同类别之间的对象距离较远。几乎所有的聚类算法均遵循以下 5 个步骤:对初始数据集进行预处理;分析数据提取有效特征,使得数据能够保留更多信息;根据提取出的特征以及数据结构选择或设计合适的聚类算法进行聚类;对聚类结果使用相关的评价指标进行有效性验证;分析聚类得出的结果,得到一定的结论。目前相似性度量主要有三种,分别是距离测度、相似度测度及匹配测度。

聚类分析根据分类对象不同分为 Q 型聚类分析和 R 型聚类分析。 Q 型聚类分析是指对样本进行聚类, R 型聚类分析是指对变量进行聚类分析。其中,系统聚类法(hierarchical clustering method)属于 Q 型聚类分析方法,它是聚类分析诸方法中最常用的一种方法,其基本思想是:开始将 n 个样本各自作为一类,

并规定样本之间的距离和类与类之间的距离，然后将距离最近的两类合并成一个新类，计算新类与其他类的距离；重复进行最近两类的合并，每次减少一类，直至所有的样本合并为一类[11][12]。

3. 数据的探索性分析

3.1. 区域经济发展评价指标的选取

通过相关文献研究发现[1][2][7]，影响区域经济发展的因素主要包括两个方面：一是利于经济增长的因素，它主要包括经济效益产业结构、劳动力资源、自然资源状况、基础设施状况等；二是居民收入水平、消费水平等。因此，在遵循区域经济指标体系建立原则的基础上，结合相关的研究成果，本文以2018年重庆市各城区的经济发展为研究对象，数据来源于2018年《重庆统计年鉴》的统计数据。经过多年的行政区划变动，现在重庆市下辖19个区、15个县和4个自治县，总计38个区县。该数据包含38个样本以及11个特征，这11个特征是反映该城区经济发展综合实力的11个指标，如表1所示。

Table 1. Economic indicators

表 1. 经济指标

指标	单位	指标代码
地区生产总值	万元	Y
常住人口	万人	X ₁
人均地区生产总值	元	X ₂
农业		X ₃
工业总产值	万元	X ₄
建筑业总产值	万元	X ₅
全社会固定资产投资	万元	X ₆
社会消费品零售总额	亿元	X ₇
城镇非私营单位就业人员年平均工资	元	X ₈
全体居民人均可支配收入	元	X ₉
全体居民人均生活消费支出	元	X ₁₀

3.2. 研究区域概况

重庆市，简称“渝”，地处105°11'~110°11'E和28°10'~32°13'N之间，位于中国大陆西南部，介于青藏高原和长江中下游平原的过渡地带，处于长江上游地区，渝东北连接陕西省，渝南接贵州，东北部和西部与四川接壤，渝东南抵湖北省。该市辖区面积为8.24万平方千米，主要包括渝中区、万州区、开县和武隆县等38个县(区)，见图1。重庆市人力资源丰富，农业基础较好，工业以重工业为主，金融业和商业较发达。从经济区位来看，重庆是川渝经济圈的重要组成部分，东部与中部五省经济区相邻，联合四川实现区域经济共同发展，充分利用本市的水路和陆路交通优势，在国家西部大开发战略背景下，实现其区域经济的快速发展，同时，具有巨大的经济吸引力和发展潜力[8]。

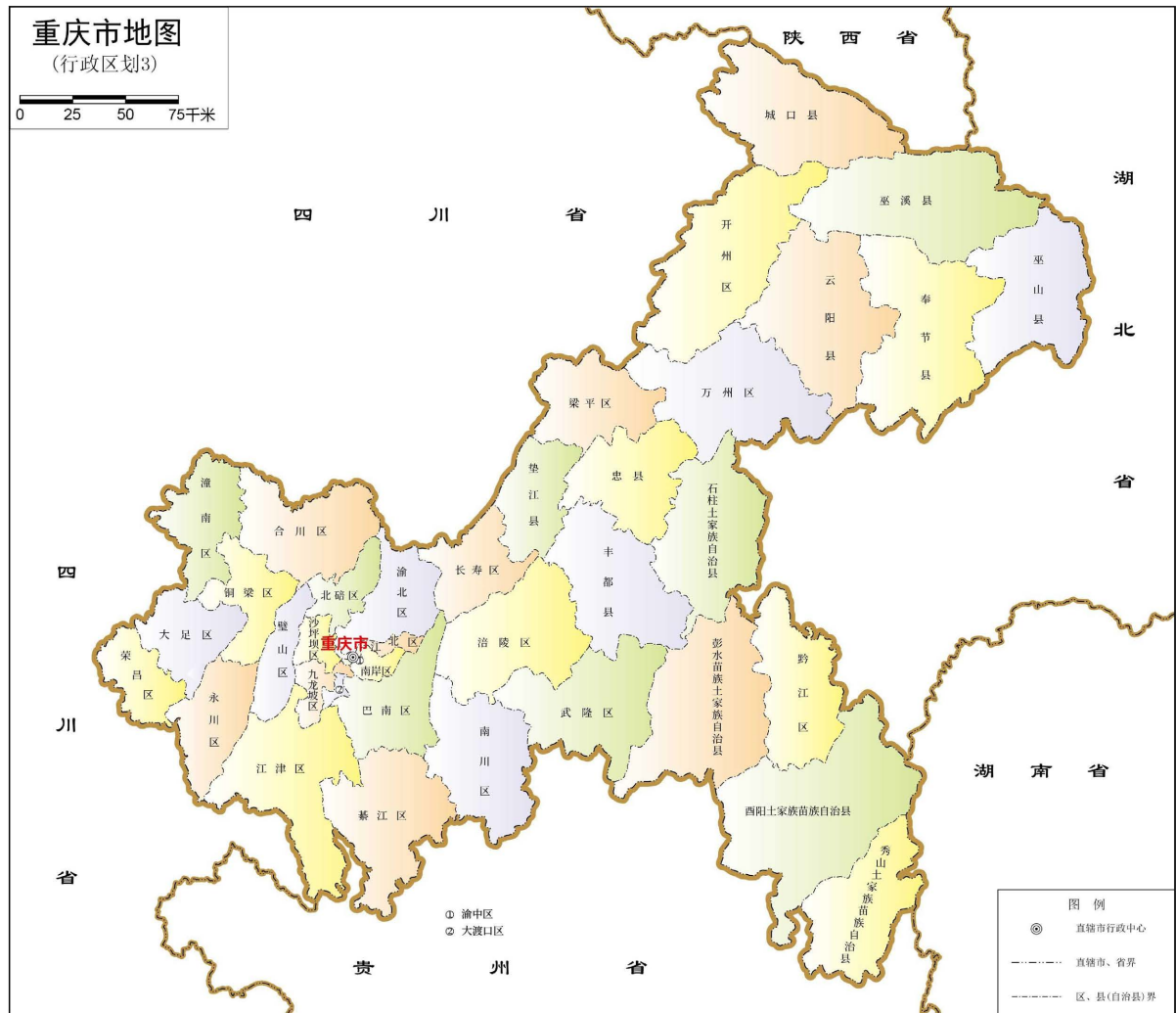


Figure 1. Map of Chongqing

图 1. 重庆市地图

3.3. 数据的初步探索

3.3.1. 描述性统计分析

本文首先考虑对 11 个经济指标进行初步探索，计算得到各个变量的主要描述统计量如表 2 所示。

由表 2 可知，重庆市平均地区生产总值是 5,318,419 万元；城口县的地区生产总值是最低的，为 557,772 万元；渝北区的地区生产总值是最高的，为 15,430,945 万元。重庆各区域的平均常住人口约为 74 万人，其中渝北区的常住人口最多，约为 166 万人，城口县的常住人口最少，约为 18 万人。通过重庆统计年鉴多年来的数据可以发现，渝中区的农业一直是无数值的，查阅资料可知，在所有的区县中，仅渝中区没有农业生产。于是把空值改成了 0；江津区的农业总产值居全市第一，近年来，江津区致力于建设国家现代农业示范区，充分发挥富硒资源，突出发展富硒产业，着力构建现代农业发展格局，加快转变农业发展方式，有效开辟了一条“富硒为民、富硒富民”的现代特色效益农业发展之路，实现了传统农业向现代农业的转变。社会发展到今天，工业仍然是经济社会发展的重要支柱。2019 年，重庆工业增加值达到 6656.72 亿元，仅次于上海、深圳、苏州，排全国第四位。支撑起重庆这个工业大市的，是多个工业强区强县，排名第一的是渝北区，工业总产值为 27,974,422 万元。

Table 2. The main descriptive statistics of each variable
表 2. 各变量的主要描述统计量

指标代码	最小值	中位数	均值	最大值	标准差
Y	557,772	4,650,556	5,318,419	15,430,945	3,688,049
X ₁	18.44	73.61	81.63	166.17	36
X ₂	26,995	52,642	61,198	182,540	29,352
X ₃	0	304,236	340,178	960,117	224,917
X ₄	22,419	2,561,524	5,444,748	27,974,422	6,232,139
X ₅	22,419	1,793,152	2,057,743	6,929,296	1,608,877
X ₆	682,033	4,442,446	4,807,286	14,755,900	3,045,715
X ₇	13.16	123.71	198.47	665.56	185
X ₈	61,929	76,503	76,643	94,778	7131
X ₉	14,835	26,507	26,798	40,484	7375

接下来通过半幅星图更直观地刻画数据的特征，如图 2 所示。

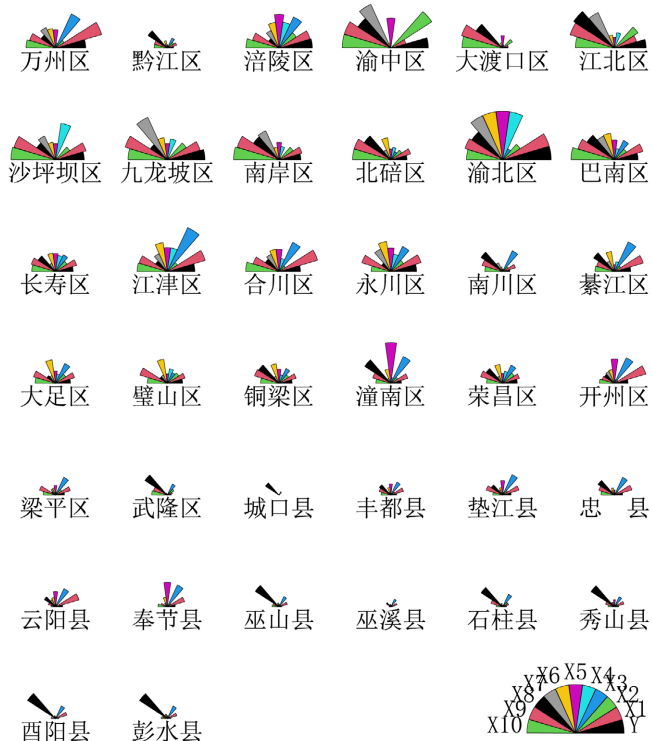


Figure 2. Half star diagram of economic indicators of 38 districts (counties)
图 2. 38 个区(县)经济指标的半幅星图

可以很明显地看出,渝北区除了农业发展水平方面稍微低一些,其他的都具有明显优势;重庆主城区,即重庆市城乡总体规划中所称的都市区,范围包括渝中区、大渡口区、江北区、南岸区、沙坪坝区、九龙坡区、北碚区、渝北区和巴南区行政区域,这些区域的社会消费品零售总额、全体居民人均可支配收入和全体居民人均生活消费支出都是比较高的。因为都市区是国家中心城市的核心载体,是全市的政治、经济、文化、交通、金融中心,并且渝中区是整个重庆主城区的核心,尤其是解放碑和朝天门商圈。虽然这些年来,重庆东部欠发达地区(简称渝东欠发达地区)经济社会达到了历史新水平。然而,从图 2 中可看出,同全市经济发达地区相比,渝东欠发达地区目前的发展仍存在相当大的差距。从下面的轮廓图也可以得到相似的结论,如图 3 所示。

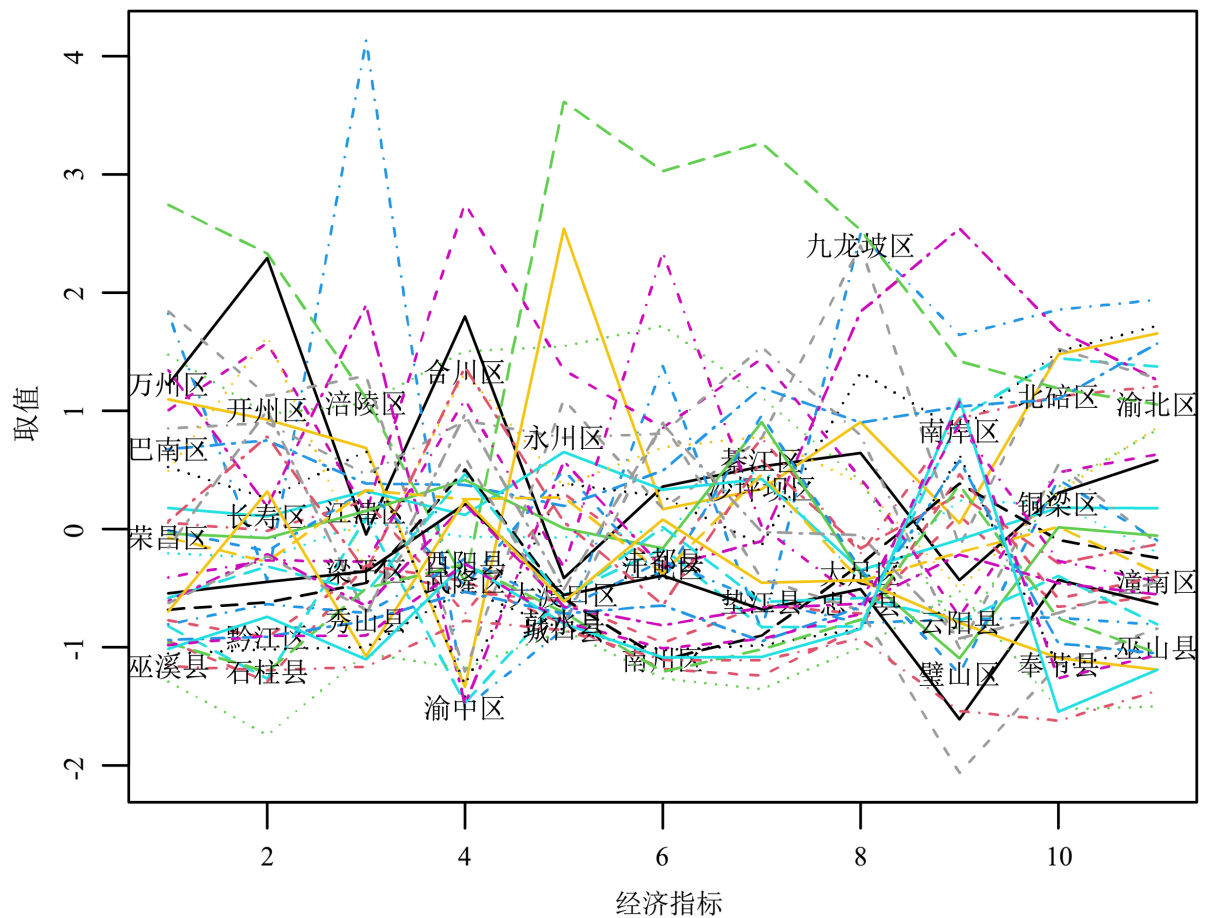


Figure 3. Skeleton diagram
图 3. 轮廓图

从图 3 可知,重庆东部地区的轮廓线都偏低,即各个经济指标偏低;相反,位于市中心附近的区域轮廓线大多数情况下都偏高,即各经济指标较高。

3.3.2. 变量间的相关关系探索

本文要利用 Lasso 变量选择、主成分分析和聚类分析,来分析重庆区域经济差异的现状和成因,这些方法在数据的变量之间有较大相关关系时是有效的。于是,我们要先检测变量间的相关性,利用 R 软件对数据进行计算,得到相关系数热力和散点图如图 4、图 5 所示。

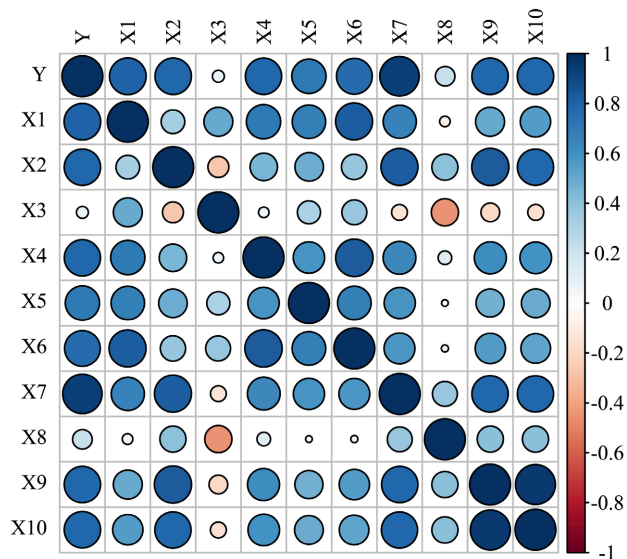


Figure 4. Correlation coefficient thermal diagram

图 4. 相关系数热力图

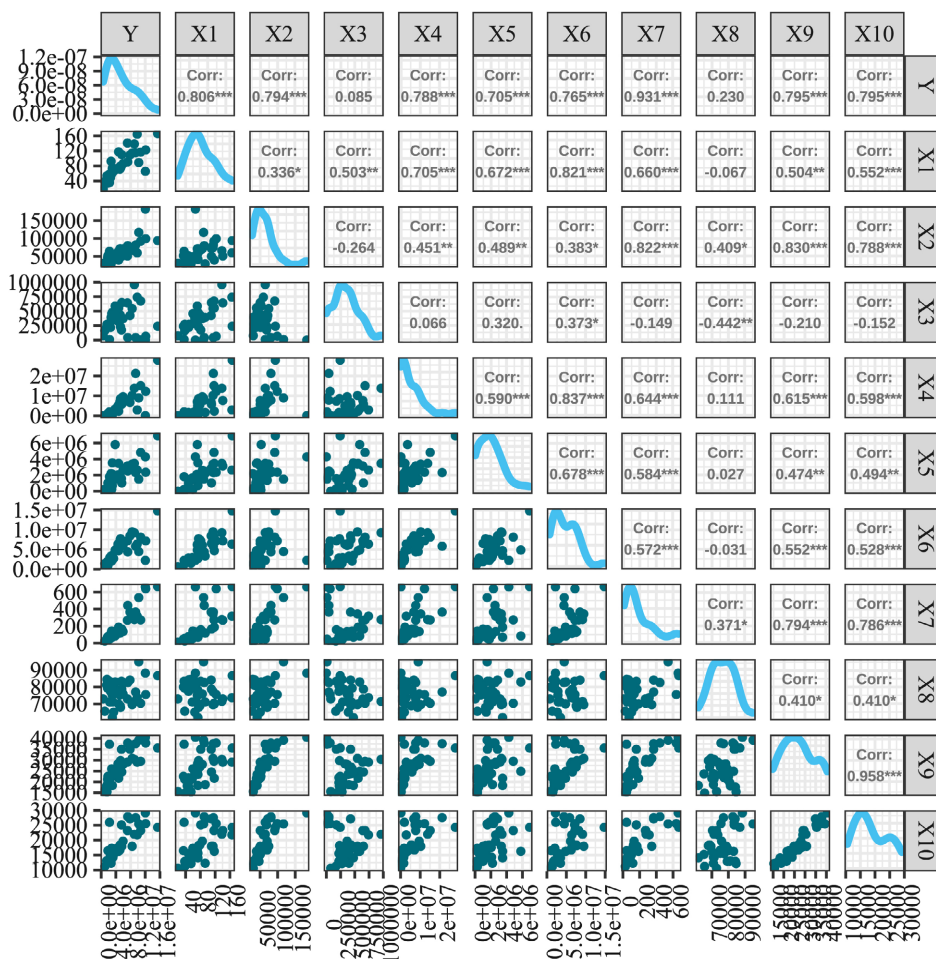


Figure 5. Scatter diagram

图 5. 散点图

由图 4 和图 5 可知, 大部分变量间的相关系数都在 0.3 以上, 表明各变量间具有较高的相关性, 适合本文要使用的方法。

4. 实证分析

影响区域经济发展的因素有很多, 结合相关文献研究成果, 本文选择了反映重庆市各城区经济发展综合实力的 11 个指标。在进行主成分分析和聚类分析之前, 本文通过进行变量选择的 Lasso、逐步回归和自适应 Lasso 与不进行变量选择的线性模型作比较, 得到 Lasso 变量选择法选取的指标在重庆区域经济发展研究中更具代表性, 预测精度也更高。

4.1. 基于 Lasso 方法的变量选择

以 2018 年重庆各区县地区生产总值作为被解释变量, 记为 Y , 运用 Lasso 方法建立模型:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{10} X_{10} + \varepsilon \quad (11)$$

其中 β_0 为常数项, $\beta_i (i=1,2,\dots,10)$ 为各变量系数, ε 为随机干扰项。

图 6 是利用 R 软件运行得到的变量选择路径图。左图显示了系数随参数变化的结果, 横轴已由最小二乘解给出的 $\sum_{j=0}^p |\beta_j|$ 的最大可能值(即 s)进行了缩放。右图显示了使用交叉验证(cross validation, CV)法来选择 s , 可看到 s 取值建议在 0.8 左右, 查看左图, 在 s 为 0.8 左右产生了一个包含 7 个自变量的模型。也就是基于 Lasso 方法最终选取出 X_1 、 X_2 、 X_4 、 X_5 、 X_6 、 X_7 、 X_9 这 7 个主要变量, 如表 3 所示。

由表 3 可知, 基于 Lasso 变量选择法, 影响重庆区域经济发展的主要因素即为常住人口、人均地区生产总值、工业总产值、建筑业总产值、全社会固定资产投资和全体居民人均可支配收入。

Table 3. Parameter estimation results of Lasso method

表 3. Lasso 方法的参数估计结果

变量系数	估计值
X_1	39258.541
X_2	54.506
X_3	0
X_4	0.106
X_5	0.004
X_6	0.045
X_7	5197.399
X_8	0
X_9	-50.150
X_{10}	0

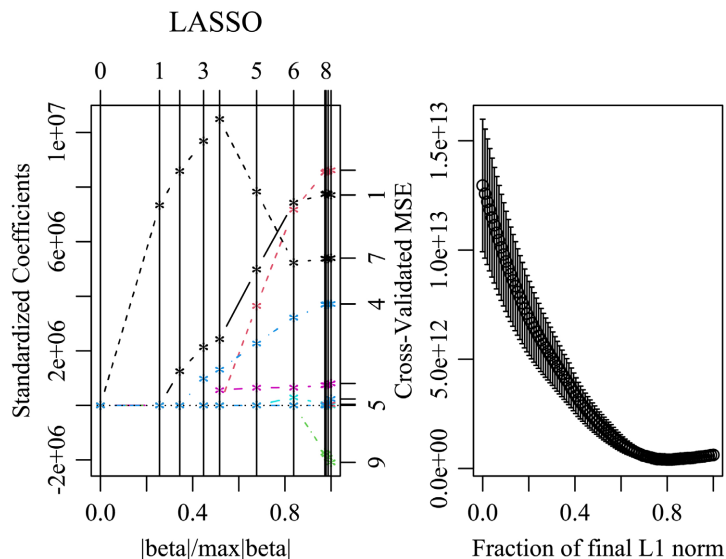


Figure 6. Variable selection path diagram for Lasso method
图 6. Lasso 方法的变量选择路径图

4.2. Lasso 方法与其他变量选择方法的比较

本文通过将 Lasso 方法、逐步回归法和自适应 Lasso 进行对比[14], 发现 Lasso 变量选择法既达到简化模型的目的, 又提高了预测精度。为了说明变量选择起到简化模型的作用, 首先用原始自变量做一个线性回归模型, 再利用三种变量选择方法简化模型。按 8:2 的比例将数据分为训练集和测试集, 选择平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)作为预测评价指标, 整理如表 4 所示。

Table 4. Comparison of models
表 4. 模型比较

模型	MAPE (%)	剔除的自变量
线性回归	10.461	
Lasso	10.136	X_3 、 X_8 、 X_{10}
逐步回归	10.730	X_3 、 X_5 、 X_6 、 X_8 、 X_{10}
自适应 Lasso	11.323	X_3 、 X_5

从表 4 可知, Lasso 方法的 MAPE 是最低的, 为 10.136%, 因此选择剔除 X_3 、 X_8 、 X_{10} 是较合适的。

4.3. 主成分分析

4.3.1. 相关性检验

经过变量选择, 从原始 11 个变量中剔除了 3 个变量。在进行主成分分析之前, 先检测变量间的相关性, 以剔除变量后的数据为依据, 利用 R 软件对这些数据进行计算, 得到相关系数热力图如图 7 所示。

由图可以发现, 各变量间具有较高的相关性, 说明可以使用主成分分析方法对重庆市的区域经济差异进行分析。

4.3.2. 提取主成分

提取特征值大于 1 的主成分，结果如表 5 所示。本文将 8 个指标转化为 2 个主成分，其中第 1 个主成分的方差贡献率为 71.05%，第 2 个主成分的方差贡献率为 14.22%，合计 85.27%，说明两个主成分包含 85.27% 的原始信息。由于前两个主成分的累计方差贡献率达到 85.27% 大于 80%，因此，可选取前两个主成分作为重庆市城区经济发展的综合指标。

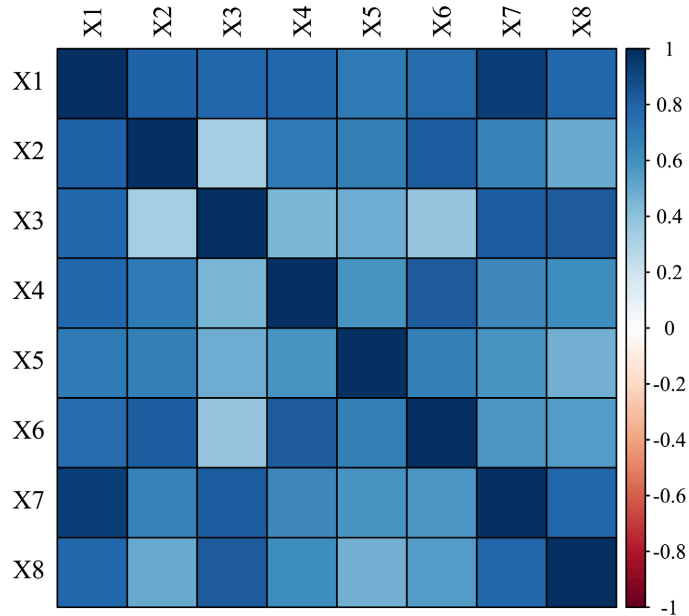


Figure 7. Thermodynamic diagram

图 7. 热力图

Table 5. Eigenvalues and variance contribution rates of principal components

表 5. 主成分的特征值和方差贡献率

成分	初始特征值			提取平方和载入		
	特征值	方差贡献率 (%)	累计贡献率 (%)	合计	方差贡献率 (%)	累计贡献率 (%)
1	5.684	71.048	71.048	5.684	71.048	71.048
2	1.138	14.220	85.268	1.138	14.220	85.268
3	0.466	5.830	91.098			
4	0.320	3.999	95.097			
5	0.197	2.467	97.564			
6	0.126	1.576	99.140			
7	0.061	0.768	99.908			
8	0.007	0.092	100			

得到的碎石图如图 8 所示。

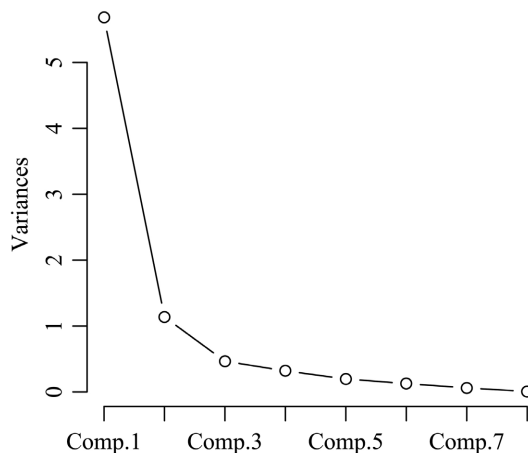


Figure 8. Scree plot
图 8. 碎石图

由图 8 可以看出，从第三个主成分开始，方差下降速度变缓，因此保留两个主成分是合理的。因子载荷矩阵如表 6 所示。

Table 6. Factor load matrix
表 6. 因子载荷矩阵

	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈
X ₁	0.411			0.213	0.184	0.153	0.230	0.819
X ₂	0.344	-0.396		0.612	-0.274	-0.132	0.392	-0.322
X ₃	0.318	0.571	0.141	-0.156		0.449	0.417	-0.386
X ₄	0.352	-0.260	-0.462	-0.404	0.512	-0.319	0.204	-0.159
X ₅	0.321	-0.227	0.810	-0.356		-0.238		
X ₆	0.350	-0.415	-0.202	-0.223	-0.278	0.630	-0.374	
X ₇	0.376	0.276		0.417	0.365	-0.106	-0.651	-0.201
X ₈	0.347	0.380	-0.261	-0.205	-0.643	-0.437	-0.100	0.102

由表 6 则可以得到主成分的表达式：

$$Z_1 = 0.411X_1 + 0.344X_2 + 0.318X_3 + 0.352X_4 + 0.321X_5 + 0.350X_6 + 0.376X_7 + 0.347X_8$$

$$Z_2 = -0.396X_2 + 0.571X_3 - 0.260X_4 - 0.227X_5 - 0.415X_6 + 0.276X_7 + 0.380X_8$$

由此可以计算主成分得分并进行排序。在主成分分析中，主成分得分是根据主成分表达式将标准化后的数据代入而得到的。先求出前 2 个主成分的分 f_1 、 f_2 ，以各主成分的方差贡献率 v_i 作为权数，构造综合评价函数 $f = v_1f_1 + v_2f_2$ ，就可以得到主成分的综合得分 f 并对各城区经济发展进行排序，主成分得分及排名如表 7。

Table 7. Principal component score ranking
表 7. 主成分得分排名

区县	f_1	f_1 排名	f_2	f_2 排名	f	综合排名
渝北区	7.128	1	-1.959	38	5.613	1
渝中区	3.388	3	4.375	1	3.553	2
九龙坡区	3.749	2	1.035	5	3.296	3
江北区	2.592	7	2.246	2	2.534	4
沙坪坝区	2.946	5	0.075	13	2.467	5
涪陵区	3.084	4	-0.689	29	2.454	6
江津区	2.616	6	-1.331	37	1.958	7
永川区	2.395	8	-0.638	28	1.889	8
巴南区	2.052	9	-0.022	16	1.706	9
南岸区	1.731	11	1.157	4	1.635	10
万州区	1.804	10	-0.758	30	1.377	11
合川区	1.550	12	-1.177	35	1.095	12
北碚区	0.689	13	0.194	11	0.606	13
长寿区	0.636	14	-0.300	25	0.480	14
璧山区	0.428	15	-0.105	20	0.339	15
大足区	0.138	16	-0.327	26	0.060	16
荣昌区	-0.046	19	0.002	15	-0.038	17
铜梁区	-0.091	20	-0.120	21	-0.096	18
綦江区	0.036	17	-0.788	31	-0.101	19
潼南区	-0.026	18	-0.820	33	-0.158	20
开州区	-0.118	21	-1.178	36	-0.295	21
大渡口区	-1.106	22	1.366	3	-0.694	22
垫江县	-1.317	23	-0.124	22	-1.118	23
梁平区	-1.412	25	0.149	12	-1.151	24
奉节县	-1.366	24	-0.804	32	-1.272	25
南川区	-1.684	28	0.606	7	-1.302	26
云阳县	-1.429	26	-1.007	34	-1.359	27
忠县	-1.675	27	-0.048	17	-1.404	28

Continued

丰都县	-1.896	29	-0.227	24	-1.618	29
黔江区	-2.014	30	0.321	9	-1.624	30
秀山县	-2.353	31	0.030	14	-1.956	31
武隆区	-2.581	33	0.820	6	-2.014	32
石柱县	-2.550	32	0.514	8	-2.039	33
彭水县	-2.671	34	-0.084	19	-2.240	34
巫山县	-2.835	35	-0.056	18	-2.372	35
酉阳县	-2.926	36	-0.328	27	-2.493	36
巫溪县	-3.259	37	-0.213	23	-2.751	37
城口县	-3.602	38	0.216	10	-2.966	38

从表 7 中综合得分排名可以看出, 渝北区、渝中区、九龙坡区的经济发展排名前三, 而巫山县、酉阳县、巫溪县、城口县经济发展排名靠后。从表中还可以发现, 行政级别为区的城市比行政级别为县的经济发展情况好。

4.4. 聚类分析

将表 7 的两个公共因子采用系统聚类法进行聚类分析, 得到聚类树状图如图 9 所示。

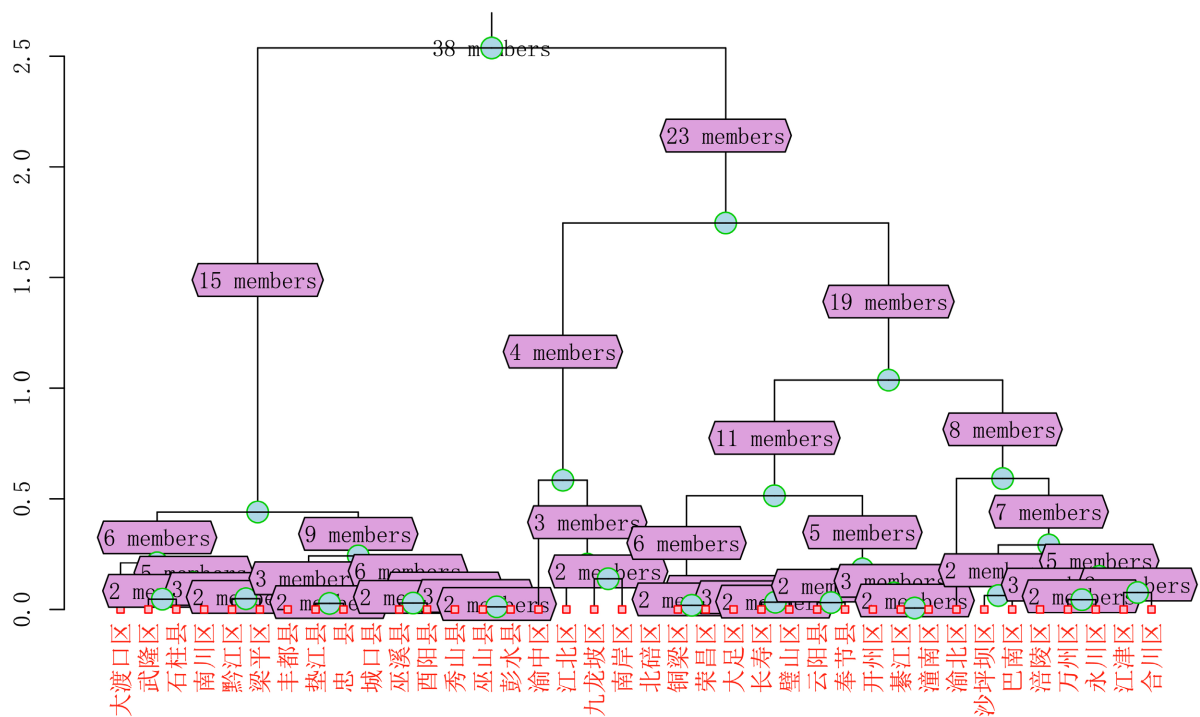


Figure 9. Cluster dendrogram
图 9. 聚类树状图

以 4 类划分为例，其聚类情况如表 8 所示。

Table 8. Classification table
表 8. 分类表

区县	类别	区县	类别
渝北区	1	巫山县	2
沙坪坝区	1	酉阳县	2
涪陵区	1	巫溪县	2
江津区	1	城口县	2
永川区	1	渝中区	3
巴南区	1	九龙坡区	3
万州区	1	江北区	3
合川区	1	南岸区	3
大渡口区	2	北碚区	4
垫江县	2	长寿区	4
梁平区	2	璧山区	4
南川区	2	大足区	4
忠县	2	荣昌区	4
丰都县	2	铜梁区	4
黔江区	2	綦江区	4
秀山县	2	潼南区	4
武隆区	2	开州区	4
石柱县	2	奉节县	4
彭水县	2	云阳县	4

从表中可以看出，渝中区、江北区、九龙坡区、南岸区综合经济实力最强、经济发展最好。这四大区位于重庆城市圈的中心，它们的地理位置优越，交通便利，经济基础坚实，综合实力强大。渝北区、沙坪坝区、涪陵区、江津区、永川区、巴南区、万州区和合川区在经济发展方面属于第二梯队。这几个区距离重庆市中心距离较近，地理优势相较于其他城区明显，而且各城区的特色经济发展良好。北碚区、长寿区、璧山区、大足区、荣昌区、铜梁区、綦江区、潼南区、开州区、奉节县和云阳县属于第三梯队，大渡口区、垫江县、梁平区、南川区、忠县、丰都县、黔江区、秀山县、武隆区、石柱县、彭水县、巫山县、酉阳县、巫溪县和城口县属于第四梯队。从以上分析可以看出，这 4 类城区的经济发展差异是非常明显的。计算 4 个类别中各城区各主成分得分和综合得分的均值，结果如表 9 所示。

Table 9. Regional economic differences in Chongqing
表 9. 重庆市区域经济差异

	第 1 类	第 2 类	第 3 类	第 4 类
f_1 均值	2.947	-2.259	2.865	-0.105
f_2 均值	-0.812	0.196	2.203	-0.478
f 均值	2.320	-1.849	2.755	-0.167

可以看出,第 3 类地区的第 2 主成分得分和综合得分最高,分别为 2.203 和 2.755,综合经济发展水平最好;第 1 类地区的第 1 主成分得分最高,为 2.947,经济发展水平较高,综合得分为 2.320;第 2 类地区的第 1 主成分得分和综合得分均为负数,综合得分为-1.849,综合经济发展水平最差。综合经济发展水平较好的第 1 类和第 3 类,说明重庆市区域经济发展水平存在很大差异。

5. 结论和建议

重庆市东西部经济发展呈现两极分化,这主要与区域的自然地理条件、经济基础和交通区位因素有关。以涪陵为界限以西的地区较为发达,中部以万州区隔开分成两个大型的农业区域,这种“三分天下”的大势明显使重庆市的经济发展极不均衡,这种不均衡对重庆未来整体发展是非常不利的。因此,改善东南和东北两个方向的农业区是极为关键的。众所周知,重庆素有“山城”之称,在重庆修建公路和铁路本要比平原地区困难许多,再加之重庆东北部巫溪、巫山等地区多暗河,完善的铁路和公路的耗资非常大,直接阻碍了该地区对外联系和经济流通。因此,对于重庆市区域经济的发展,要根据重庆各城区的实际情况,采取相应的措施发挥各城区的优势,重点扶持经济相对落后的城区,使各城区经济协调发展。具体来说,可分为以下三个方面。

加强政府宏观调控和布局,促进经济社会的发展。加强和改善宏观调控,保持经济平稳较快发展。深化金融改革,改善金融服务,优化信贷结构,大力扶持高新技术产业。加大对贫困地区的农业、教育、社会保障等投入。加强交通网络的建设,加快机场、水运、铁路、公路的建设。

加快经济结构调整,大力发展高新技术产业。从经济结构看,目前重庆各产业结构比例相差很大,需要加快产业结构的调整,以“稳定巩固第一产业、优化提高第二产业、加快发展第三产业”的方针,来壮大支柱产业,加快发展高新技术产业。

加快经济落后地区的经济发展。根据各区县的资源状况和条件,发挥优势,突出特色。对各区县进行分类指导,打造各自的主导产业,防止产业趋同,培育区县经济新的增长点。

参考文献

- [1] 张震,王泽宇,李瑛.中国区域经济高质量发展研究综述[J].资源开发与市场,2021,37(8):928-933.
- [2] 肖金成,沈体雁,左万水.中国经济南北差距扩大的原因与趋势分析——中国区域经济 50 人论坛第二十次专题研讨会综述[J].经济与管理,2022,36(1):40-47.
- [3] 杨万平,李冬.中国八大区域经济发展质量的空间差异及其形成机制[J].当代经济科学,2022,44(2):51-65.
- [4] 孙久文,蒋治.“十四五”时期中国区域经济发展格局展望[J].中共中央党校(国家行政学院)学报,2021,25(2):77-87.
- [5] 李依凝.基于多元统计分析的京津冀一体化区域经济发展研究[J].商场现代化,2021(4):169-171.
- [6] 贾婷婷,张宾.安徽省区域经济差异作用因素研究——基于市域面板数据的实证分析[J].重庆科技学院学报(社会科学版),2021(3):48-54.

-
- [7] 高志刚, 张毅. 区域经济差距对西部地区经济高质量发展的影响研究[J]. 宁夏社会科学, 2021(1): 99-110.
- [8] 张彬, 谢贤健, 杨联安, 郭金铭. 基于 PCA-TOPSIS 的重庆市县域经济密度综合测度及时空演变[J]. 干旱区地理, 2020, 43(2): 546-556.
- [9] 杨虎, 杨玥含. 金融大数据统计方法与实证[M]. 北京: 科学出版社, 2016.
- [10] 薛毅, 陈立萍. 统计建模与 R 软件[M]. 北京: 清华大学出版社, 2007.
- [11] 农秋红, 韦程东, 罗文婷. Lasso 变量选择法在广西区域经济发展影响因素选取中的应用[J]. 中国商论, 2021(10): 162-164.
- [12] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [13] 李炳军, 张一帆, 张淑华. 主成分分析法的改进及其在河南区域经济发展评价中的应用[J]. 河南农业大学学报, 2022, 56(3): 479-489. <https://doi.org/10.16445/j.cnki.1000-2340.20210618.001>
- [14] Zou, H., Hastie, T. and Tibshirani, R. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, **15**, 265-286. <https://doi.org/10.1198/106186006X113430>