

基于多元统计分析的山东省经济发展研究

梁佳慧

曲阜师范大学, 山东 曲阜
Email: jiayouljhljh@qq.com

收稿日期: 2021年6月26日; 录用日期: 2021年7月19日; 发布日期: 2021年7月28日

摘要

山东省是全国的经济大省、人口大省、文化大省, 所以山东省的发展一直是国家关注的重点对象之一, 为此本文将针对山东省16个城市的经济发展现状进行研究。首先本文选取与经济发展相关的11个指标, 并对这11个指标的数据进行因子分析, 将11个指标的信息概括为经济总量因子(第一主因子)和人民生活水平因子(第二主因子), 再根据主因子计算综合得分, 对16个城市进行排名, 给出综合评价模型。再使用聚类分析, 将不同经济发展水平的城市进行分类, 研究山东省地区的经济发展结构, 最后使用Fisher判别检验聚类结果的正确性。通过以上方法, 本文指出山东省城市经济发展现状, 并为山东省经济发展提供优化措施和建议。

关键词

因子分析, 聚类分析, Fisher判别, 经济发展, 综合评价

Research of Economic Development in Shandong Province Based on Multivariate Statistical Analysis

Jiahui Liang

Qufu Normal University, Qufu Shandong
Email: jiayouljhljh@qq.com

Received: Jun. 26th, 2021; accepted: Jul. 19th, 2021; published: Jul. 28th, 2021

Abstract

Shandong Province is a large economic province, a large population province, and a large cultural

province in the country. Therefore, the development of Shandong Province has always been one of the key objects of national concern. For this reason, this paper will focus on the current economic development of 16 cities in Shandong Province. First of all, this paper selects 11 indicators related to economic development, and performs factor analysis on the data of these 11 indicators, and summarizes the information of the 11 indicators into the economic aggregate factor (the first main factor) and the people's living standard factor (the second Main factor), and then calculate the comprehensive score based on the main factor, rank 16 cities, and give a comprehensive evaluation model. Then use cluster analysis to classify cities with different economic development levels, study the economic development structure of Shandong Province, and finally use Fisher discrimination method to test the correctness of the clustering results. Through the above methods, this paper points out the current situation of urban economic development in Shandong Province, and provides optimization measures and suggestions for the economic development of Shandong Province.

Keywords

Factor Analysis, Cluster Analysis, Fisher Discriminant, Economic Development, Comprehensive Evaluation

Copyright © 2021 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

衡量一个城市的强弱与否，经济发展水平都是一个至关重要的指标。近年来，为解决山东省经济发展问题，山东省下发多个与经济相关的政策文件，落实“六稳”“六保”促进高质量发展政策清单，鼓励支持扩大制造业和技术改造投资方面，充分激发社会消费潜力方面。有鉴于此，本文将针对山东省 16 个城市的经济发展水平进行研究。

张震、覃成林[1]构建了由经济发展动力、新型产业结构、交通信息基础设施等 7 个维度组成的经济高质量发展指标体系，并构建城市群经济高质量发展指数对粤港澳大湾区经济高质量发展进行研究；朱家明、孙榕竟[2]选取了地区生产总值、固定资产投资价格指数、货运量总计、工业污染治理完成投资、国内专利申请受理项数等指标，使用因子分析对国内地区经济发展水平进行了综合评价与分析，并给出区域协调发展与优化资源配置的建议；代美华[3]针对川渝区域经济高质量发展的影响因素，提取川渝区域经济发展过程中的变量，确定了川渝区域经济高质量发展统计评价模型指标及其权重，并使其结果量化；郭杨[4]使用因子分析对山东省各地区综合实力进行研究评价，其使用因子分析得出了四个主因子，并且对四个主因子都进行了聚类分析，从四个方面对于山东省的地区综合实力进行了分析。李奇、胡世雄[5]等建立了经济总量、经济结构、经济效益三方面的经济发展评价体系，使用主成分分析研究山西省经济综合实力。

本文通过查阅参考文献，最终选择了固定资产投资、社会消费零售总额、就业人员数、生产总值、全体居民可支配收入、全体居民人均消费支出、公共预算收入、第二产业值、第三产业值、出口总值、进口总值 11 个衡量经济发展水平的指标，使用因子分析方法对其进行研究，再进一步使用聚类分析对不同经济发展水平的城市进行分类。经过翻阅大量参考文献，发现大部分学者使用因子分析和聚类分析对经济发展水平进行研究时，并没有使用统计方法对聚类分析结果进行验证，所以本文为验证聚类分析结

果的正确性,使用 Fisher 判别方法直观判断聚类分析结果的正确性,保证过程的完整性和准确性。本文基于以上方法,将对山东省经济发展水平进行研究,指出山东省的经济发展现状,并给出相关建议。

2. 材料与方法

2.1. 研究经济发展的指标选取原则

要建立全面、准确、合理的区域指标,需要遵守以下基本原则[6]:

1) 科学性:选取指标需具有理论依据,可以经得起实践检验,能客观反映经济发展状况,具备足够的科学性。

2) 代表性:选取的指标应该具有代表性,与经济发展相关。这并不意味着选取的指标越多越好,应当以深入了解城市的经济发展为根本,选择能够区分不同城市经济发展水平,并且可以反映问题,具有代表性的指标。

3) 可行性:选取指标时,要充分考虑所选取的指标是否适合用来定量分析,其不会导致评价过程变得困难,可确保分析的可行性和分析结果的客观性。

4) 可获得性:选取指标时要考虑所需数据的来源,在遵循代表性和可行性的基础上,选择可通过统计年鉴、单位部门等有关途径获得准确数据的指标。

2.2. 指标的选取

基于以上选取指标的原则并参考文献,本文以 2019 年山东省 16 城市的经济发展为对象,选择了 11 项能够反映城市经济发展的主要经济指标,具体指标如下表 1 所示。(数据来源于 2020 年山东省统计年鉴以及 2020 年山东省 16 个城市的统计年鉴,具体数据见附录。)

Table 1. Selected indicators

表 1. 选取的指标

记号	变量	记号	变量	记号	变量
x_1	固定资产投资	x_2	社会消费零售总额	x_3	就业人员数
x_4	生产总值	x_5	全体居民可支配收入	x_6	全体居民人均消费支出
x_7	公共预算收入	x_8	第二产业值	x_9	第三产业值
x_{10}	出口总值	x_{11}	进口总值		

2.3. 地区经济发展研究的方法

本文首先选取 11 个影响城市经济发展的主要经济指标,构建了山东省经济发展评价指标体系,并通过借助 R 软件,运用因子分析、聚类分析, Fisher 判别对数据进行分析。通过因子分析建立山东省经济发展的评价模型,聚类分析将山东省不同经济发展水平的城市进行分类,最后通过 Fisher 判别直观检验聚类结果的正确性。

2.3.1. 因子分析

因子分析是一种将较多数据实现降维的方法,它试图用几个潜在,不可预测的随机变量(因子)来描述原始变量间的协方差或相关关系。这些随机变量的个数会尽可能地比原始变量的个数少,并且涵盖原始变量的较多信息[7]。

因子分析的矩阵形式的模型如下:

$$x = \mu + Af + \varepsilon$$

即：

$$\begin{cases} x_1 = \mu_1 + \alpha_{11}f_1 + \alpha_{12}f_2 + \cdots + \alpha_{1m}f_m + \varepsilon_1 \\ x_2 = \mu_2 + \alpha_{21}f_1 + \alpha_{22}f_2 + \cdots + \alpha_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_p = \mu_p + \alpha_{p1}f_1 + \alpha_{p2}f_2 + \cdots + \alpha_{pm}f_m + \varepsilon_p \end{cases}$$

其中 f_1, f_2, \dots, f_m 称为公共因子, ε_i 为 x_i 的特殊因子, a_{ij} 为因子载荷, 是第 i 个变量在第 j 个因子上的负荷。并且符合以下条件

- 1) $m \leq p$;
- 2) $E(f) = E(\varepsilon) = 0$;
- 3) $V(f) = I$, 即公共因子之间不相关且方差为 1;
- 4) $V(\varepsilon) = D = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$, 特殊因子不相关且方差不同;
- 5) $\text{cov}(f, \varepsilon) = 0$, 即公共因子与特殊因子不相关;

2.3.2. 聚类分析

常见的聚类分析方法有系统聚类法、动态聚类法、最优分割法等。本文使用系统聚类法进行聚类分析。系统聚类法是将最初 n 个样本看成一类, 然后计算类与类之间的距离, 将类与类之间距离最小的归为一类, 为新类, 再计算新类与其他类的距离, 重复上述过程, 直到所有样品归为一类。

2.3.3. Fisher 判别

Fisher 判别是一种降维的方法。用原始变量的少数几个线性组合(Fisher 判别函数)来代替原始变量, 并通过判别函数对样品的归属做出判别或将各组分离。当判别函数为 2 个, 可计算各样品的两个判别函数得分, 并画出判别函数得分的散点图, 使用目测法对样品的分离情况及结构进行观察。通过判别函数得分的散点图, Fisher 判别在本文当中可用来检验聚类分析的结果是否正确。

3. 结果与分析

3.1. 因子分析

3.1.1. KMO 和 Barlett 检验

Table 2. KMO Bartlett Ball degree test

表 2. KMO 和 Barlett 的检验

KMO 和巴特利特检验		
KMO 取样适切性量数。		0.8
	近似卡方	335.1679
巴特利特球形度检验	自由度	55
	显著性	0.000

由表 2 可知, KMO 的值为 0.8, Barlett 的球形检验的显著性小于 0.01, 意味着原始变量之间的相关性很强, 说明原始变量适合做因子分析。

3.1.2. 因子分析过程

利用 R 软件进行因子分析, 结果如下:

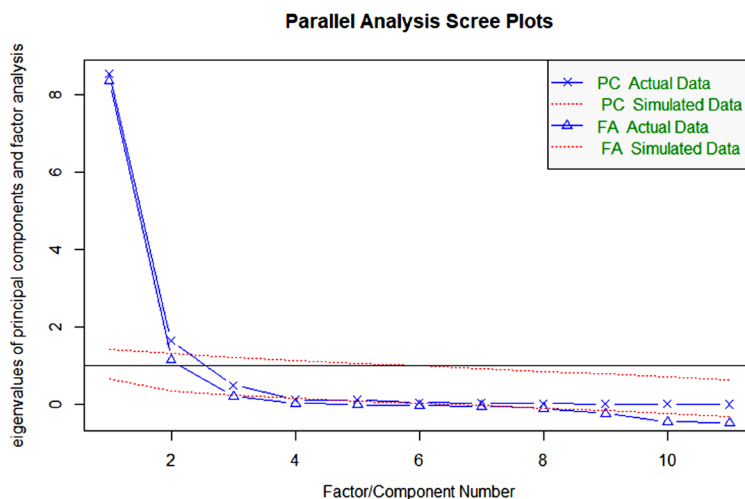


Figure 1. Factor analysis

图 1. 因子分析图

图 1 中以因子的特征值为纵坐标，以因子个数为横坐标，从图中可以看出前两个特征值非常陡峭，而在 2 个以后，变化就不是很明显，由此可以说明 2 个因子就能涵盖原始变量的大部分信息。将所得结果进行整理，所得数据如下表：

Table 3. Explanation of total variance

表 3. 总方差解释表

	第一主因子	第二主因子
初始特征值	8.30	1.62
初始特征值方差百分比	0.75	0.15
累积方差百分比	0.75	0.9
解释原始变量百分比	0.84	0.16
累积解释百分比	0.84	1

Table 4. Factor load table

表 4. 因子载荷表

	第一主因子	第二主因子	主成分对每个变量的方差解释度	方差无法被主成分解释的比例
x_1	0.93	0.23	0.91	0.09
x_2	0.91	0.39	0.98	0.02
x_3	0.47	0.78	0.83	0.17
x_4	0.97	0.25	1.00	0.004
x_5	0.82	-0.53	0.96	0.04
x_6	0.84	-0.54	1.00	0.005
x_7	0.97	0.19	0.97	0.03
x_8	0.97	0.14	0.96	0.04
x_9	0.96	0.27	0.99	0.01
x_{10}	0.84	0.07	0.71	0.29
x_{11}	0.76	-0.19	0.61	0.39

表 3 是因子提取后的结果。初始特征值是衡量因子重要程度的指标。第一主因子的特征值为 8.3，表明第一个因子解释了原有方差的 75%，第二主因子的特征值为 1.62，表明第二个主因子解释了原有方差的 1.62，两者的方差累积贡献率为 90%，所以保留两个主因子。由于未旋转的因子载荷矩阵可能某一行有多个载荷系数比较大，也有可能某一列有多个载荷系数比较大，这表明公共因子可能解释了多个变量的信息，导致该因子代表性不强，所以为了更好的解释每一个公共因子所代表的实际含义，进行因子旋转。利用 R 软件所得结果如下：

Table 5. Explanation table of total variance after factor rotation

表 5. 因子旋转后总方差解释表

	第一主因子	第二主因子
初始特征值	5.58	4.33
初始特征值方差百分比	0.51	0.39
累积方差百分比	0.51	0.9
解释原始变量百分比	0.56	0.44
累积解释百分比	0.56	1

Table 6. Factor load table after factor rotation

表 6. 因子旋转后因子载荷表

	第一主因子	第二主因子	主成分对每个变量的方差解释度	方差无法被主成分解释的比例
x_1	0.81	0.51	0.91	0.09
x_2	0.91	0.39	0.98	0.02
x_3	0.89	-0.20	0.83	0.17
x_4	0.85	0.53	1.00	0.004
x_5	0.19	0.96	0.96	0.04
x_6	0.18	0.98	1.00	0.005
x_7	0.80	0.57	0.97	0.03
x_8	0.77	0.60	0.96	0.04
x_9	0.86	0.51	0.99	0.01
x_{10}	0.63	0.56	0.71	0.29
x_{11}	0.39	0.68	0.61	0.39

对比表 3，表 4，表 5，表 6 可知：旋转前后因子的方差比例不变，但在各观测值的载荷上发生了改变。由表 6 可知：

第一主因子在固定资产投资、社会消费零售总额、就业人员数、生产总值、公共预算收入、第二产业值、第三产业值、出口总值、进口总值有比较大的载荷，命名为经济总量因子。

第二主因子在全体居民可支配收入、全体居民人均消费支出有较大的载荷，命名为生活水平因子。

由表 6 旋转后因子载荷表将 2 个主因子用原始的 11 个变量表示出来，可得如下综合评价模型：

第一主因子得分函数：

$$F1 = 0.81x_1 + 0.91x_2 + 0.89x_3 + 0.85x_4 + 0.19x_5 + 0.18x_6 \\ + 0.8x_7 + 0.77x_8 + 0.86x_9 + 0.63x_{10} + 0.39x_{11};$$

第二主因子得分函数：

$$F2 = 0.51x_1 + 0.39x_2 - 0.2x_3 + 0.53x_4 + 0.96x_5 + 0.98x_6 + 0.57x_7 + 0.6x_8 + 0.51x_9 + 0.56x_{10} + 0.68x_{11};$$

综合得分： $F3 = 56\%F1 + 44\%F2$

通过综合评价模型，计算出山东省各个城市各个因子的得分，得到结果如下：

Table 7. Main factor scores
表 7. 主因子得分

	F1	第一主因子得分排名	F2	第二主因子得分排名	综合得分 F3	综合得分排名
济南	9.21	2	7.09	2	8.27	2
青岛	18.20	1	15.33	1	17.00	1
淄博	-1.98	7	0.57	7	-0.86	7
枣庄	-6.039	16	-4.98	15	-5.58	16
东营	-3.59	10	1.29	5	-1.44	9
烟台	6.85	3	6.17	3	6.55	3
潍坊	3.54	4	1.71	4	2.73	4
济宁	0.21	6	-1.73	8	-0.64	6
泰安	-3.81	12	-3.28	11	-3.58	11
威海	-2.89	9	0.87	6	-1.23	8
日照	-5.56	15	-3.63	12	-4.7	14
临沂	1.26	5	-2.24	9	-0.28	5
德州	-3.637	11	-4.31	13	-3.934	13
聊城	-4.87	14	-5.50	16	-5.15	15
滨州	-4.62	13	-2.83	10	-3.83	12
菏泽	-2.28	8	-4.54	14	-3.27	10

第一主因子代表山东省各个城市的经济总量因子，第二主因子代表山东省各个城市的人民生活水平因子，由表 7 可以看出无论是第一主因子得分、第二主因子得分，还是综合得分都存在较大的差异，其中青岛市的第一主因子得分，第二主因子得分和综合得分最高，枣庄市的第一主因子得分、综合得分最低，聊城市的第二主因子成分最低，每一种得分的极差都在 20 以上。第一主因子、第二主因子、综合得分的平均值都分别近似为 0，其中每个得分都高于平均值的城市有：青岛市、济南市、烟台市、潍坊市，低于平均值的城市有：淄博市、枣庄市、东营市、济宁市、泰安市、威海市、日照市、临沂市、德州市、聊城市、滨州市、菏泽市，这说明山东省地区经济发展差距很大，发展不平衡，不协调。

下图 2 和图 3 是利用 R 软件进行数据可视化的结果：

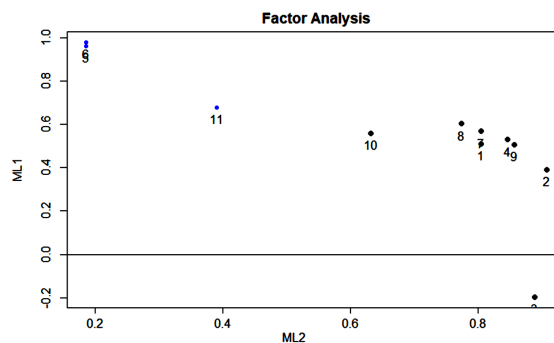


Figure 2. Factor analysis data analysis
图 2. 因子分析数据分析

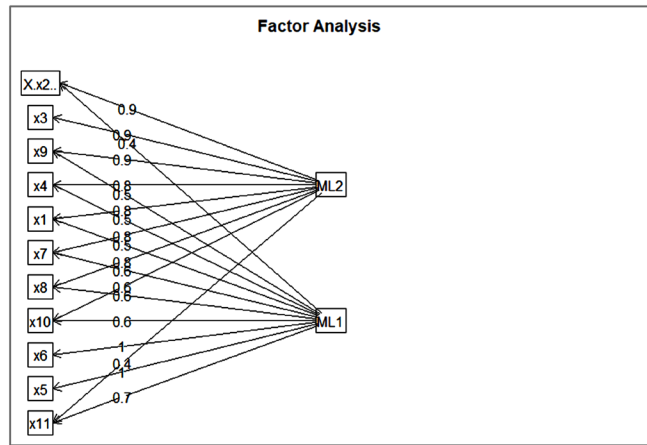


Figure 3. Results of factor analysis
图 3. 因子分析数据结果

3.2. 对原始变量进行聚类分析

将所得标准化数据利用 R 软件按照最短距离法进行聚类分析，得到分类结果，见表 8。

Table 8. Results of cluster analysis
表 8. 聚类分析结果表

城市	济南市	青岛市	淄博市	枣庄市	东营市	烟台市	潍坊市	济宁市	泰安市	威海市	日照市
类别	1	2	3	4	3	1	3	4	4	3	4
城市	临沂市	德州市	聊城市	滨州市	菏泽市						
类别	4	4	4	4	4						

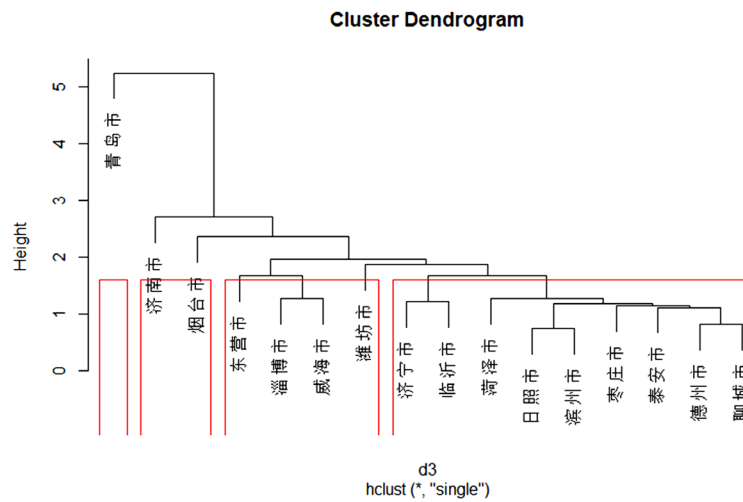


Figure 4. Cluster analysis result chart
图 4. 聚类分析结果图

由上图 4 可以看出聚类结果为：

第一类：青岛市

第二类：济南市、烟台市

第三类：东营市、潍坊市、淄博市、威海市

第四类：枣庄市、东营市、济宁市、日照市、临沂市、德州市、聊城市、滨州市、菏泽市

通过对原始变量的聚类，将山东省 16 个城市分为了四类，该种分类结果也符合实际情况，青岛市是一个港口城市，贸易发达、招商引资，有很强的商业基础。济南市为山东省会城市、是南北东西的交通枢纽。烟台是一个沿海城市，地理位置好，气候环境好，富有盛名的水果之乡，所以第一类、第二类城市的经济发展水平都比较高。而第三类城市是省内经济发展质量一般的城市，有些城市的地理位置好，但缺乏地方特色产业，经济发展道路中规中矩。最后一类城市的经济发展水平最低，在经济结构及金融发展维度的表现更是差强人意[8]。

3.3. Fisher 判别

本文针对原始变量聚类分析的结果使用 R 进行 Fisher 判别。

Table 9. Contribution rate of discriminant function

表 9. 判别函数贡献率

判别函数	第一判别函数	第二判别函数	第三判别函数
贡献率	0.8134	0.1645	0.022

由表 9 可以看出第一判别函数和第二判别函数的累积贡献率为 98%，涵盖了原始变量大部分信息，所以选取两个判别函数。

第一判别函数为：

$$y_1 = 19.46x_1 + 7.4x_2 + 7.19x_3 - 176x_4 - 10.35x_5 + 11.43x_6 + 7.52x_7 + 46.6x_8 + 97.9x_9 - 0.829x_{10} + 7.51x_{11}$$

第二判别函数为：

$$y_2 = 12.13x_1 + 5.66x_2 + 6.39x_3 - 90x_4 - 4.15x_5 + 3.59x_6 + 6.85x_7 + 18.4x_8 + 41.7x_9 - 4.17x_{10} + 5.06x_{11}$$

根据第一判别函数和第二判别函数计算出山东省 16 个城市的第一判别函数得分和第二判别函数得分，并以第一判别函数得分为横坐标，第二判别函数得分为纵坐标，画出散点图。

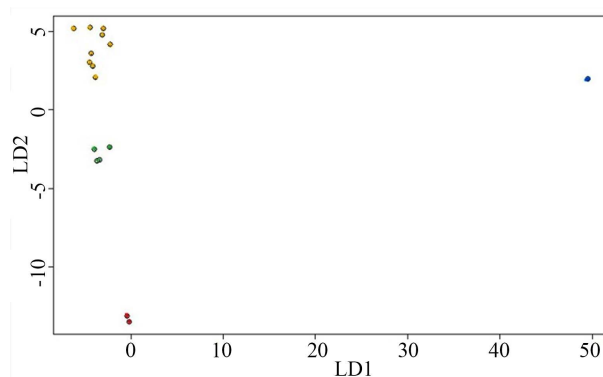


Figure 5. Scatter plot of Fisher discriminant function

图 5. Fisher 判别函数得分散点图

根据散点图 5 可以看出每一类的分离效果都十分明显，这说明该种聚类结果是合理正确的。

4. 讨论

4.1. 总结

本文通过因子分析将多个有关经济发展的指标转化为第一主因子(经济总量因子)和第二主因子(生活水平因子),并计算其综合得分来探究山东省不同城市的经济发展水平。而后通过聚类分析将山东省不同经济水平的城市进行分类,最后再通过 Fisher 判别验证聚类结果是否准确。

4.2. 政策建议

4.2.1. 区域协调发展, 缩小城市之间差距

由数据分析结果可以看出,山东省不同城市之间经济发展水平存在明显的差距,所以山东省不同城市地区应协调发展,缩小不同城市之间的差距。对于经济发展相对落后的城市应充分挖掘内部潜力,加大基础设施建设力度,调整人民收入分配格局,实现城乡居民收入增长和经济增长同步,劳动报酬增长和劳动生产率提高同步,培养投资需求和消费需求[9]。周边城市也可以发挥辐射作用,带动经济发展质量较低的城市一同发展。

4.2.2. 打造区域特色, 发展重点经济

政府应当提高对各城市自身发展的条件和可以利用的资源认识,建立不同城市在不同资源领域、不同技术方向上的竞争优势。东营市可利用好现有优势,大力发展高效生态农业,环境友好型工业,现代服务业,海洋产业等。济宁市可主导现代医药,现代农业等,泰安市可主导旅游业等。各城市相互借鉴,取长补短,提高山东省经济发展水平。

参考文献

- [1] 张震,覃成林. 粤港澳大湾区经济高质量发展分析[J]. 经济体制改革, 2021(3): 39-46.
- [2] 朱家明,孙榕竟. 基于因子-聚类分析对地区经济发展动力水平的综合评价[J]. 廊坊师范学院学报(自然科学版), 2021, 21(1): 62-67.
- [3] 代美华. 川渝区域经济高质量发展统计评价方法研究[J]. 时代金融, 2021(13): 39-41.
- [4] 郭杨. 基于因子分析和聚类分析的山东省各地区综合实力评价研究[D]: [硕士学位论文]. 济南: 山东大学, 2019.
- [5] 李奇,胡世雄,王争磊. 基于主成分分析的山西省各城市经济综合实力评价[J]. 对外经贸, 2021(4): 80-83.
- [6] 文乐. 基于多元统计分析的武汉市城区经济发展研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2016.
- [7] 王学民. 应用多元统计分析[M]. 第5版. 上海: 上海财经大学出版社, 2017: 231-256.
- [8] 马国强,陈锋. 山东省各地级市区域经济发展质量评价研究[J]. 青岛大学学报(自然科学版), 2020, 33(4): 110-114.
- [9] 促进西部和欠发达地区与全国缩小差距[N]. 人民日报, 2012-03-13(019).

附录

山东省 16 个城市未标准化数据

2019	各市固定资产投资 (亿元)	社会消费零售总额 (亿元)	各市年底就业人员 数(万人)	各市生产 总值(亿 元)	各市全体 居民可支 配收入(元)	各市全体居 民人均消费 支出(元)	各市一般公 共预算收入 (万元)	第二产业 值	第三产业 值	各市出口 总值(万美 元)	各市出口 总值(万美 元)
济南市	5385.10	4420.41	510.3	9443.37	41,472	26,639	8,741,898	3267.406	5836.003	934,788	695,984
青岛市	10204.00	5126.60	580.2	11741.31	45,452	29,501	12,417,359	4179.906	7150.458	4,943,258	3,649,250
淄博市	1851.50	1155.84	274.9	3642.42	37,543	24,537	3,686,683	1817.568	1675.513	605,733	672,788
枣庄市	1213.80	921.45	238.2	1693.91	26,291	15,824	1,472,629	736.8509	797.8316	200,853	10,544
东营市	1551.40	665.10	129.5	2916.19	40,602	25,360	2,451,019	1673.893	1096.487	497,315	1,866,571
烟台市	6226.40	2805.00	436.0	7653.45	37,783	25,113	5,954,203	3183.835	3918.566	2,510,602	1,702,423
潍坊市	3893.10	2388.00	551.8	5688.50	32,518	20,575	5,710,571	2292.466	2878.381	1,637,051	952,781
济宁市	3586.20	2153.70	495.9	4370.17	28,055	17,165	4,050,125	1761.179	2106.422	410,928	256,535
泰安市	2428.40	1054.11	341.7	2663.59	29,690	18,102	2,247,115	1036.137	1339.786	191,794	54,662
威海市	2688.20	1163.18	170.5	2963.73	39,593	25,421	2,498,531	1197.347	1478.901	1,336,251	698,916
日照市	1501.10	604.50	179.1	1949.38	27,577	16,361	1,703,736	832.3853	949.3481	590,263	937,008
临沂市	3105.50	2523.28	637.0	4600.25	27,619	14,483	3,300,090	1743.495	2447.333	982,973	224,314
德州市	2448.40	1114.33	341.3	3022.27	22,608	15,276	2,062,687	1263.309	1447.667	295,105	205,904
聊城市	1390.00	827.22	360.0	2259.82	21,602	13,823	1,966,287	806.7557	1134.43	300,099	292,517
滨州市	1392.60	710.44	232.0	2457.19	28,517	18,852	2,429,629	1041.849	1184.366	452,084	814,833
菏泽市	1545.20	1618.02	509.5	3409.98	20,673	14,444	2,219,223	1452.651	1633.38	254,895	449,439

因子分析 R 代码

```
setwd("D:/R1")
x<-read.csv("论文数据.csv",sep="," ,header=T)
x
x<-scale(x[-1])
x_cor<-cor(x)
x_cor
library(psych)
KMO(x)
cortest.bartlett(x)
fa.parallel(x.obs=176,fa="both",n.iter=100,show.legend=T)
x<-scale(x[-1])
fa.parallel(x_cor,n.obs=176,fa="both",n.iter=100,show.legend=T)
model1<-fa(x_cor,nfactors=2,rotate="none",fm="ml")
model1
model2<-fa(x_cor,nfactors=2,rotate="varimax",fm="ml")
model2
factor.plot(model2)
fa.diagram(model2,simple=F)
factor.plot(model2)

$$F1 < -0.81 * x[,1] + 0.91 * x[,2] + 0.89 * x[,3] + 0.85 * x[,4] + 0.19 * x[,5] + 0.18 * x[,6] + 0.8 * x[,7] + 0.77 * x[,8] + 0.86 * x[,9] + 0.63 * x[,10] + 0.39 * x[,11]$$

F1

$$F2 < -0.51 * x[,1] + 0.39 * x[,2] - 0.2 * x[,3] + 0.53 * x[,4] + 0.96 * x[,5] + 0.98 * x[,6] + 0.57 * x[,7] + 0.6 * x[,8] + 0.51 * x[,9] + 0.56 * x[,10] + 0.68 * x[,11]$$

F3<-0.56*F1+0.44*F2
```

聚类分析 R 代码

```
region<-x[-1]
X<-data.frame(row.names=region,x)
d3<-dist(X)
hc3<-hclust(d3,"single")
plot(hc3)
rect.hclust(hc2,k=4)
cutree(hc2,k=4)
```

Fisher 判别 R 代码

```
g<-c(1,2,3,4,3,1,3,4,4,3,4,4,4,4,4)
data1<-as.data.frame(cbind(g,x))
library(MASS)
```

```
ld<-lda(g~x1+X.x2..+x3+x4+x5+x6+x7+x8+x9+x10+x11, data1)
Z<-predict(ld)
round(Z$x, 3)
plot(Z$x, cex=1.4)
text(Z$x[, 1],Z$x[, 2],cex=0.5, data1$g)
g<-c(1,2,3,4,3,1,3,4,4,3,4,4,4,4,4)
data1<-as.data.frame(cbind(g,x))
library(MASS)
ld<-lda(g~x1+X.x2..+x3+x4+x5+x6+x7+x8+x9+x10+x11, data1)
ld
round(Z$x, 3)
plot(Z$x, cex=1.4)
text(Z$x[, 1],Z$x[, 2],cex=0.5, data1$g) #为散点标号
```