ISSN 2616-5902 Vol. 4, Issue 5: 47-51, DOI: 10.25236/AJBM.2022.040510

Study on Economic Impact Indicators of Different Cities Based on Principal Component Analysis and Cluster Analysis

Tiantian Li[1], Xinyue Chen[2]

[1]School of Mathematics and Statistics, Hubei University of Arts and Science, Xiangyang, Hubei, China
[2]Cryptographic Engineering Institute, Inforamation Engineering University, Zhengzhou, Henan, China

Abstract: *Economic strength has always been an important indicator for evaluating a county and a city. In this paper, a total of 19 indicators are used to construct an evaluation index system for regional economic. Taking 21 cities and autonomous prefectures in Sichuan Province as samples, three principal component indexes are selected to reflect most of the information of the original indexes, and the ranking of their comprehensive economic strength is obtained. According to the cluster analysis model, the 19 selected economic impact indicators are divided into five categories through correlation. In each sub-category, combined with principal component analysis, the ranking of cities is obtained, and the final ranking of Sichuan's comprehensive economic strength is given by comparing the results of two rankings. Then we gave Suggestions for adjustment and development.*

Keywords: *Comprehensive economy; Principal component analysis; Cluster analysis; KMO and Bartlett test*

1. Introduction

Economic strength has always been an important indicator in assessing a country. Sichuan Province is a strong economic province in the western region of China, but the development between regions is unbalanced and the level of development varies greatly. It is of great practical significance to study the factors that measure the economic development status, analyse the contribution of indicators to economic development, and get to evaluate the economic status of cities to promote regional economic development and so on.

At present, the research on the evaluation of economic impact indicators is rich in content, evaluation standards, methods and methods are also diverse. Some scholars use factor analysis to evaluate the level of economic development. For example, Huang Yuting [1] evaluated the level of economic development of the Yangtze River Economic Belt based on clustering and factor analysis. Other scholars use principal component analysis and analytic hierarchy process (AHP) to evaluate urban economy. For example, Zhang Dong et al. [2] evaluated and analyzed the urban development of Cangzhou city by combining the analytic hierarchy Process and principal component analysis. Yu Na et al. [3] set up the evaluation index system of urban agglomeration economic development in the Yangtze River Delta and conducted the evaluation research using principal component analysis. This paper combines principal component analysis and cluster analysis to evaluate and analyze the level of urban economic development of the subordinate cities in Sichuan Province.

This paper takes all 21 cities and autonomous prefectures under the jurisdiction of Sichuan Province as the research object, which comprehensively reflects the economic development level of each region in Sichuan Province, and the sample itself is universal. At the same time, the selection is universal and random as the economic development of each region is relatively stable, with little variation in data from year to year and high correlation.

In this paper, the sample data are from Sichuan Statistical Yearbook 2017[4].

2. Method and Model

Principal Component Analysis. It uses the theory of dimensionality reduction thinking[5] to synthesize multiple relevant variables into one or a few comprehensive indicators with little loss of information, and

ISSN 2616-5902 Vol. 4, Issue 5: 47-51, DOI: 10.25236/AJBM.2022.040510

this one or a few indicators can largely reflect the information of the original variables as a multivariate statistical method. It can not only eliminate the influence of the scale and correlation among the indicators, but also to a certain extent overcome the problem of subjectivity brought about by the artificial determination of the weight coefficients of each indicator, highlight the advantages of the performance of principal components, and achieve the purpose of simplifying the system structure and solving the problem in essence.

Cluster Analysis. The data samples are grouped according to the degree of similarity or dissimilarity of the features of the patterns to be classified, so that the data in the same group are as similar as possible and the data in different groups are as dissimilar as possible. It is intended to be used for the discovery of characteristics of data samples rather than for prediction [6].

Conclusion of Method. When using statistical analysis methods to study and analyse an issue, if the number of variables involved is large and there is a certain correlation between the variables, if the original data is used to model directly, it will increase the complexity of the research [7]. Economic development involves a wide range, covers a lot of content, and there is a certain correlation between economic impact indicators, and the indicator data will be repeatedly used, thus leading to inefficient evaluation. In this paper, the principal component analysis method is chosen, which is to use the theory of dimensionality reduction to transform multiple indicators into a few comprehensive indicators, which can reflect most of the information of the original variables, thus simplifying the problem, by screening the repeated economic impact indicators, making a more objective evaluation of the indicators affecting economic development through a few principal components, and combining with cluster analysis to more intuitively reflect the economic development of different cities The results were tested for reliability using KMO and bartlett's method, making the results more accurate, credible and scientific.

3. Case Study

3.1. Data Processing

This paper takes 21 cities and autonomous prefectures in Sichuan Province as the research objects. Based on the existing economic evaluation system, a total of 19 economic impact indicators are selected based on the actual development situation of each city and prefecture, respectively as follows: primary industry (X1), secondary industry (X2), tertiary industry (X3), population size (X4), employed persons (X5), total wages of employed persons (X6), urban land area (X7), fixed asset investment (X8), total retail sales of social goods (X9), public budget income (X10), public budget expenditure (X11), per capita disposable income of urban residents (X12), disposable income of rural residents (X13), value added of the private economy (X14), consumption expenditure of urban residents (X15), consumption expenditure of rural residents (X16), energy consumption of gross regional product (X17), total highway mileage (X18), and higher education (X19). Data standardization. First eliminate the dimension and directionality of indicators. Standard deviation standardization was used to standardize the sample data, and the final standardized data is obtained by MATLAB.

3.2. Principal Component Analysis

1) Correlation Matrix. From the original data processed to obtain the normalisation matrix Z_{ij}, the correlation coefficient matrix $R=\left(r_{ij}\right)$ was obtained using SPSS.

2) KMO and Bartlett’s Test. In this paper, KMO and Bartlett tests were conducted using SPSS software and the results are shown in Table 1. the KMO value was 0.767, indicating a weak bias and strong correlation between the variables, and the significance of the Bartlett test was less than 0.01, indicating suitability for principal component analysis.

Table 1: KMO and Bartlett’s Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.767
Bartlett’s Test of Sphericity	Approx. Chi-Square	947.261
	df	171
	Sig	0.000

3) Total Variance Interpretation. Table 2 is obtained by arranging eigenvalues, in which each eigenvalue corresponds to one component. The first three principal components have a cumulative contribution rate of 92.256%, including the information content of 92.256% of the original economic

ISSN 2616-5902 Vol. 4, Issue 5: 47-51, DOI: 10.25236/AJBM.2022.040510

index, indicating that the first three principal components can basically explain the principal variance, so the three principal components are selected.

Table 2: Variance Cumulative Contribution Rate

Principal Component	Initial Eigenvalues		
	Characteristic root	Contribution rate (%)	Cumulative contribution rate (%)
1	14.506	76.346	76.346
2	2.090	11.001	87.347
3	0.933	4.909	92.256

The scree plot of eigenvalues reflects the importance of each component. As can be seen from Fig 1, the principal component on the left has already contained most information of the original 19 economic indicators. So it is reasonable to choose three principal components.

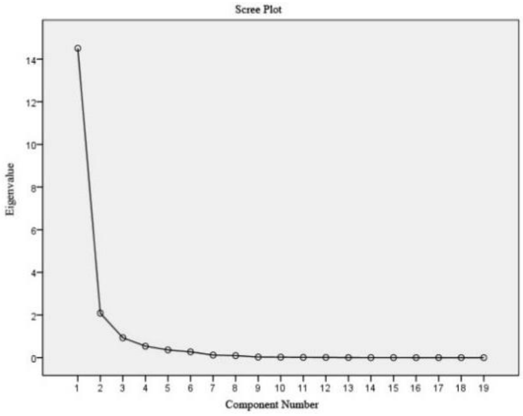

Figure 1: Scree Plot

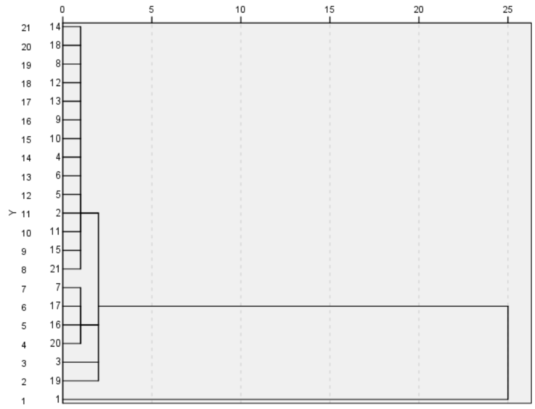

Figure 2: Systematic Clustering Pedigree

4) Rotating Component Matrix. Since the three principal component loadings in the component matrix did not clearly show the characteristics of the individual economic indicator variables, they were rotated by factor rotation to obtain a better explanation.

By rotating the component matrix it can be seen that the first component loadings are higher for the secondary industry, tertiary industry, employed persons, total wages of employed persons, fixed asset investment, total retail sales of social goods, public budget revenue, public budget expenditure, value added of the private economy and higher education, which are closely related to the total economy; the second component loadings are higher for urban residents' per capita disposable income, rural residents' The second component has higher loadings on urban disposable income per capita, rural disposable income per capita, urban consumption expenditure and rural consumption expenditure, all of which reflect the living standard of residents; the third component has higher loadings on the primary industry and total highway mileage.

Table 3: Component Score Coefficient Matrix

	Component				Component		
	1	2	3		1	2	3
X1	-0.012	-0.457	0.285	X11	0.223	-0.156	-0.121
X2	0.089	-0.053	-0.046	X12	-0.026	0.239	-0.150
X3	0.151	0.021	-0.102	X13	-0.202	0.347	0.153
X4	0.101	-0.047	0.082	X14	0.118	-0.022	-0.052
X5	0.076	-0.093	0.105	X15	-0.081	0.285	-0.094
X6	0.135	-0.037	-0.075	X16	-0.263	0.415	0.174
X7	-0.051	0.112	0.140	X17	0.211	-0.031	-0.572
X8	0.135	-0.050	-0.052	X18	-0.106	0.043	0.371
X9	0.127	-0.035	-0.055	X19	0.148	-0.039	-0.108
X10	0.163	-0.057	-0.117				

5) Comprehensive Score Component Matrix. The coefficient matrices in Table 3 were weighted and summed to calculate the component scores, denoted as F_1, F_2 , F_3 and F_i are all standardized scores, so anything greater than 0 is the upper middle level and anything less than 0 is the lower middle level. Finally, the total score of the principal components F is calculated by weighting the variance

ISSN 2616-5902 Vol. 4, Issue 5: 47-51, DOI: 10.25236/AJBM.2022.040510

contribution rate. Using SPSS to calculate the main component scores, the comprehensive scores and rankings are shown in Table 4.

$$F = 52.049\% \times F_1 + 26.032\% \times F_2 + 14.175\% \times F_3$$

Table 4: Main Component Scores, Comprehensive Scores and Rankings

City	F_1	F_2	F_3	F	Ranking
Chengdu	3.90575	1.69702	0.52249	2.548734	1
Zigong	-0.85759	0.86141	0.62144	-0.13403	11
Panzhihua	-0.61851	1.73096	-1.90796	-0.14177	14
Luzhou	-0.14795	0.21544	0.43044	0.040093	4
Deyang	-0.56609	1.07113	0.05265	-0.00834	7
Mianyang	-0.29343	0.61038	0.75976	0.113866	2
Guangyuan	0.04504	-1.3404	0.28228	-0.28548	17
Suining	-0.63294	0.20888	0.6271	-0.18617	15
Neijiang	-0.16795	-0.19862	0.02795	-0.13516	12
Leshan	-0.13738	0.21652	-0.53442	-0.09089	9
Nanchong	0.20951	-1.06768	1.52979	0.047953	3
Meishan	-0.88226	1.09781	0.39449	-0.1175	10
Yibin	-0.04629	0.05455	0.12236	0.007452	5
Guangan	-0.41785	0.1715	0.2268	-0.14069	13
Dazhou	0.39706	-1.10697	0.53005	-0.00637	6
Ya’an	-0.26229	-0.5119	-0.40163	-0.32671	20
Bazhong	0.0644	-0.99877	-0.15814	-0.2489	18
Ziyang	-0.93878	0.63816	0.83056	-0.20476	16
Ngawa	0.39307	-0.35177	-2.95401	-0.30572	19
Garzê	0.35895	-1.46262	-1.28469	-0.37603	21
Liangshan	0.59552	-1.53499	0.28268	-0.04956	8

3.3. Cluster Analysis

Based on the correlation matrix between the economic indicators, the 19 cities were clustered according to the economic indicators using the systematic clustering method (Fig 2), and the 19 cities were divided into 5 major categories according to the economic indicators by the actual situation.The degree of economic development varies widely among cities and states in Sichuan Province, and cities are roughly divided into five categories according to economic development indicators: the first category is Chengdu; the second category is Tibetan Qiang Autonomous Prefecture of Ngawa; the third category is Panzhihua; the fourth category is Tibetan Autonomous Prefecture of Garzê, Ya'an, Bazhong and Guangyuan; the remaining cities in the fifth category are Liangshan Yi Autonomous Prefecture, Dazhou, Nanchong, Zigong, Deyang, Mianyang, Luzhou , Leshan, Neijiang, Yibin, Meishan, Suining, Ziyang and Guangyan. Luzhou, Leshan, Neijiang, Yibin, Meishan, Suining, Ziyang and Guang'an. Each of these cities has a similar economic structure and a comparable level of development.

4. Summary and Conclusion

This paper uses a principal component analysis model to quantitatively analyze 19 economic impact indicators of 21 cities and autonomous prefectures in Sichuan Province in 2016, while using a cluster analysis model to qualitatively classify cities into five categories of comprehensive economic strength. Based on the actual situation, three principal components were extracted from the 19 economic impact indicators and analysed according to the sample data, and the conclusions were corroborated by KMO and Bartlett's spherical test, combined with cluster analysis. The two results are thus compared and corroborated to obtain a more accurate ranking of the overall economic strength of cities, as shown in Table 4.

As shown in Table 5, different scholars have adopted a variety of evaluation methods to study the regional economic development level of Sichuan Province. In general, the above-mentioned scholars' research objects are similar to this paper, but the research methods are slightly different, and the results obtained are also slightly different, mainly in the middle ranked cities.

Table 5: The results of studying the same problem with different methods

Scholar	Method	Research Object	Results (Overall score Ranking)
Ren Ping et al.[8]	Factor Analysis and Cluster Analysis	economic development levels	Chengdu, Panzhihua, Mianyang, Deyang, Leshan, etc
Wang Na[9]	Data Envelopment Analysis	Sichuan province economic development efficiency	Chengdu, Zigong, Panzhihua, Deyang, Neijiang, etc
Shu Fuhua [10]	Principal Component Analysis	Economic development level of 10 cities in Sichuan Province	Luzhou, Mianyang, Nanchong, Yibin, Deyang, etc

The results reflect more objectively the economic development level of each city in Sichuan Province, with a wide gap between the economic development level of each city. Chengdu, as the central city of the province, having a clear lead in all economic development indicators. The imbalance in regional economic development is becoming more obvious, and coordinating and promoting the common development of regional economies is an important development task. Based on the above analysis, the following recommendations are made: firstly, to give full play to Chengdu's economic radiation role as the provincial centre and drive the economic development of neighbouring cities together. Secondly, the less economically developed regions should make full use of the advantages of local resources, prescribe the right medicine according to the economic impact indicators, strengthen economic construction and drive economic development with advantageous resources. Finally, the state should increase its support for the less economically developed regions, create sub-centres and encourage more talents to return to the less economically developed regions through preferential policies to support the construction of their hometowns.

References

[1] Huang Yuting. Evaluation of economic development level in the Yangtze River Economic Belt based on cluster and factor analysis [J]. The Farmers Consultant, 2020(04): 242-243.

[2] Zhang Dong, Li Zhixiao, Li Zhenzhen. Empirical analysis on comprehensive evaluation of new urbanization level in Cangzhou city [J]. Economic Research Guide, 2020(07): 112+115.

[3] YU Na, CHEN Jiang-hua, YANG Cheng-gang. The Economic Development Evaluation of Yangtze River Delta Urban Agglomeration Based on Principal Component Analysis [J]. Journal of Hefei Normal University, 2019, 36(02): 31-38.

[4] Sichuan Provincial Bureau of Statistics. Sichuan Statistical Yearbook 2017. Website of Sichuan Provincial Bureau of Statistics: http://tjj.sc.gov.cn/

[5] SEOK J, WARREN H S, CUENCA A G, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases[J]. Proceedings of the National Academy of Sciences, 2013, 110(9): 3507-3512.

[6] Bai Xue. Similarity Measures in Cluster Analysis and Its Applications [D]. Beijing Jiaotong University, 2012.

[7] Yang Fangwen, Yang Zihao. Evaluation of urban competitiveness in Anhui Province [J]. Co-Operative Economy & Science, 2018(11): 48-50.

[8] REN Ping, ZHOU Jieming. Research on the Comprehensive Appraisal for Regional Difference of Economic Development Level in Sichuan Province [J]. Journal of Sichuan Normal University (Natural Science), 2007(01): 102-105.

[9] WANG Na. Evaluation of economic development efficiency in Sichuan Province based on DEA model [J]. Times of Economy & Trade, 2018(14): 98-99.

[10] Shu Fuhua. Evaluation of Urban Economic Development Level in Sichuan Province Based on Principal Component Analysis [J]. Journal of Aba Teachers University, 2018, 35(04): 63-68.