Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/84/2025.21211

Research on Economic Development Problems in Guangdong Province Based on Principal Component Analysis and Cluster Analysis

Man Wa Tik Shi[1,a,*]

[1]*School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing, Jiangsu Province, 210094, China*
a. 921158700101@njust.edu.cn
**corresponding author*

Abstract: Located on the southeast coast of China, Guangdong Province has a favorable geographical position and is one of the more economically developed provinces. At the same time, it has been effective in solving the problem of unbalanced regional development, but at the same time, there are some remaining problems. In this paper, 13 indicators are selected from four aspects, namely, national economy, per capita economy, foreign investment economy, and basic economy, to compare the economic development of 21 cities in Guangdong in 2022. Through the method of principal component analysis and k-mean clustering, it is found that for all cities in Guangdong Province, the best economic development is in the cities of Guangzhou and Shenzhen, followed by Dongguan and Foshan, and finally other cities in Guangdong Province such as Zhuhai, Zhongshan, and so on. Based on analyzing the raw data and providing insights into the economic development of Guangdong Province, it is hoped that it will add to the future development of Guangdong Province and contribute to the future development of Guangdong Province.

Keywords: principal component analysis, K-mean clustering, regional development

1. Introduction

In recent years, as the global economy has shifted from globalization to anti-globalization dominated by Trumpism, economic issues have once again become the focus of attention.2022, Guangdong Province achieved good economic results against the backdrop of the global economic downturn. The annual GDP grew by 1.9% [1]. The primary industry increased the most, providing 5.2% [1]. Similarly, the contribution of the secondary sector to the economy was significant, exceeding 50% [1]. The total foreign trade imports and exports amounted to 831 billion yuan, a year-on-year increase of 0.5%, ranking first in the country for 37 consecutive years [1].

Nevertheless, there are just as many problems exposed, the most serious of which is the imbalance between regional economic development. In 2022 alone, the gross domestic product (GDP) of the Pearl River Delta (PRD) region will be 104681.81 billion yuan, while the GDP of the mountainous areas of Guangdong Province will be only 737.115 billion yuan [1]. The combined economic output of the three regions of eastern Guangdong, western Guangdong, and northern Guangdong is less than one-tenth of that of the PRD region, while the combined fiscal revenue is less than half of that of the

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/84/2025.21211

PRD region [2]. The huge gap between the two makes the study of regional development imbalance urgent.

China's regional economic research on Guangdong Province is more in-depth, for example, Zhao quantitatively analyzed the economic development level of counties and cities (districts) in Guangdong Province by constructing an indicator system and principal component analysis [3]. The study found that the Pearl River Delta region is economically developed, while the east and west wings and the northern mountainous areas are relatively underdeveloped. The results of the study pointed out the significant economic differences between regions and put forward suggestions such as optimizing the regional economic structure and establishing a linkage mechanism in order to promote coordinated economic development within Guangdong Province. As well as He also constructed an evaluation system for the quality of local fiscal revenues in Guangdong Province from 2000 to 2015 based on factor analysis [4]. Its study found that the quality of fiscal revenues is affected by factors such as economic aggregate, industrial structure, and operational efficiency of fiscal revenues, and proposed measures such as developing private economy and high-tech industries, optimizing industrial structure, adjusting the structure of tax types, and narrowing regional gaps in order to improve the quality of fiscal revenues in Guangdong Province.

The purpose of this paper is to analyze the economic development status of cities and municipalities in Guangdong Province through PCA and K-Means clustering methods, draw corresponding conclusions, and put forward suggestions according to the local conditions to help Guangdong Province's sustainable economic development.

2. Research indicators

This paper constructs an evaluation system of economic development indicators for cities in Guangdong Province from four aspects, namely, national economy, per capita economy, foreign trade economy, and basic economy, by drawing on the evaluation indicators about urban economic development at home and abroad. As shown in Table 1.

Table 1: Evaluation system of urban economic development indicators

Target level	Business layer	Indicator layer
Level of economic development	National economy	gross regional product X1 Gross domestic product X2 Gross regional product growth X3 Share of tertiary sector X4
	Per capita economy	Regional GDP per capita X5 General budget income per capita X6 full-time teacher X10
	Foreign economy	total export amount X9 overseas foreign direct investment (OFDI) X10 Per capita disposable income of urban residents X12
	Basic economy	Passenger traffic in local municipalitiesX7 Total postal and telecommunication operationsX8 total retail sales X11 Year-end resident population X13

The statistical data used in this study came from the Guangdong Provincial Statistical Yearbook, and descriptive statistics were first performed on the collected data, as shown in Table 2.

DOI: 10.54254/2753-8818/84/2025.21211

Table 2: Descriptive statistics

Norm	Average value	(statistics) standard deviation	Analyze n
X1	58220530.95	88529082.98	21
X2	2543028.95	1820479.32	21
X3	1.39	1.23	21
X4	48.37	7.53	21
X5	7.89	4.36	21
Continue Table 2.			
X6	5711.07	5457.43	21
X7	1172.14	1593.12	21
X8	237.51	367.04	21
X9	380.91	760.60	21
X10	866198.90	1894202.41	21
X11	21333181.90	28060183.86	21
X12	4.56	1.55	21
X13	602.70	462.21	21

3. Introduction to research methods

This paper uses principal component analysis to analyze the economic indicators of cities in Guangdong Province. Principal component analysis was proposed by Prof. Harold Hotelling in 1933.Principal component analysis (PCA) simplifies the complexity of high-dimensional data while preserving trends and patterns. It captures most of the variance in the data by transforming the data into fewer dimensions, which serve as a summary of features [5]. The suitability of the data for principal component analysis can be assessed using the Kaiser-Meyer-Olkin test and the Bartlett's test of sphericity, the results of which are shown in Table 3.

Table 3: KMO and Bartlett's test

Kaiser-meyer-olkin metrics		0.745
Bartlett sphericity check	Approximate chi-square	497.262
	Df	78
	Sig.	0.000

Table 2 shows that the KMO test was found to be 0.745 which is greater than the required 0.6, indicating that the data can be factor analyzed. Also the correlation is 0 less than 0.05 which allows for principal component test.

The main steps of principal component analysis are described next:

Assuming that there are a total of j research data, each with i features, the collected data xij are normalized, i.e.

$$x_{ij}^{*} = \left(x_{ij} - \bar{x}_{j}\right) / s_{j} \tag{1}$$

Where, is the mean and variance. Second, according to the above formula to get the standardized data and get the standardized matrix X, and establish the covariance matrix R. The covariance matrix reflects the correlation between the data after the data standardization of the individual indicators of the data, the larger the value the more it indicates that it should be subjected to principal component analysis. the formula for the calculation of R is:

$$R = X^{*T} X^{*} / \left(N - 1\right) \tag{2}$$

DOI: 10.54254/2753-8818/84/2025.21211

Where is the data matrix after normalization. Through this the eigenvalues of R can be obtained$\lambda_1 \geq \lambda_2 \geq L \geq \lambda_m$ and the corresponding eigenvectors μ_1, μ_2, L, μ_m.

Calculate the variance contribution ratio:

$$\eta_i = \lambda_i / \sum_i^p \lambda_i \left(i = 1,2,L,p\right) \tag{3}$$

Calculate the variance contribution ratio:

$$\sum_{i=1}^m \lambda\imath / \sum_{i=1}^p \lambda_i \left(i = 1,2,L,m,L,p\right) \tag{4}$$

For how many principal components should be selected depends on the variance contribution of each component, usually the first m principal components corresponding to a cumulative variance contribution greater than 80%, representing data with m principal components.

Finally, the city factor score is calculated, which is obtained through the following two formulas:

$$F_m = W_1 Z_1 + W_2 Z_2 + \mathbf{L} + W_i Z_i \tag{5}$$

$$G = W_1 Z_1 + W_2 Z_2 + L + W_i Z_i \tag{6}$$

Finally, the obtained scores are ranked in descending order to get the factor score rankings [6].

4. Results of data analysis

In this paper, SPSS software was used to analyze the economic indicators of cities in Guangdong Province. First of all, it can get the total variance plot of the explanation obtained after the data were analyzed by principal component analysis, as shown in Table 4.

Table 4: Total variance explained

Initial eigenvalue				Rotate the sum of squares to load		
Norm	Add up the total	Variance	Cumulative variance	Add up the total	Variance	Accrue
1	8.997	69.211	69.211	7.751	59.627	59.627
2	1.641	12.619	81.831	2.194	16.873	76.5
3	1.03	7.926	89.757	1.723	13.257	89.757
4	0.512	3.941	93.698			
5	0.471	3.625	97.323			
6	0.241	1.851	99.174			
7	0.037	0.288	99.462			
8	0.031	0.236	99.698			
9	0.022	0.17	99.868			
10	0.009	0.068	99.935			
11	0.005	0.038	99.973			
12	0.002	0.016	99.99			
13	0.001	0.01	100			

Table 4 illustrates that the three principal components explain 59.627%, 16.873%, and 13.257% respectively. The three principal components can add up to 89.757% greater than 80%.It means that the three principal components have been able to capture the vast majority of the variability in the original data and can represent the main features of the original data well. At the same time, it also lays the foundation for the next step of cluster analysis.

DOI: 10.54254/2753-8818/84/2025.21211

5. Analysis of PCA results

Through the above Tables 1 to 4, the eigenroots of the first principal component, the second principal component, and the third principal component can be obtained as 8.997, 1.641, and 1.030, respectively. and the scores of the cities and municipalities in Guangdong Province can also be obtained. As shown in Figure 1.

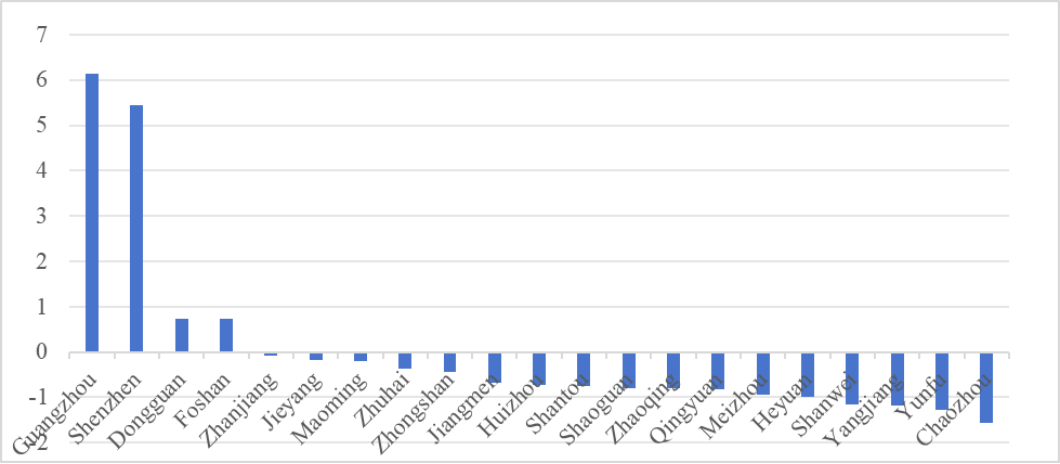


Figure 1: Scores by city (Photo/Picture credit: Original).

Figure 1 shows that Guangzhou City ranks first with a total score of 6.14 and Shenzhen City ranks second with a total score of 5.44. These two cities are far ahead of the others in terms of economic development. Next are Dongguan and Foshan, which rank third and fourth with total scores of 0.73 and 0.72 respectively. Zhanjiang and Jieyang are very close in terms of their total scores, at -0.09 and -0.18 respectively, ranking fifth and sixth. The scores of the cities after this are all lower, indicating that the economic development of these cities is lagging behind somewhat.

6. K-means cluster analysis results

For the original data, in order to facilitate the analysis, this paper also used k-mean cluster analysis, and the results are shown in Table 5.

Table 5: Clustering results

	Economic Transcendence Cities	Pillar cities of the economy	Economic development cities
Cities in Guangdong Province	Guangzhou, Shenzhen	Dongguan, Foshan	Zhuhai, Shantou, Shaoguan, Heyuan, Meizhou, Huizhou, Shanwei, Zhongshan, Jiangmen, Yangjiang, Zhanjiang, Maoming, Zhaoqing, Qingyuan, Chaozhou, Jieyang, Yunfu

Through Table 2, it can be found that Guangzhou and Shenzhen are the first type of cities, Foshan and Dongguan are the second type of cities, and Zhuhai, Shantou, Huizhou, Zhongshan, and other cities are the third type of cities.

DOI: 10.54254/2753-8818/84/2025.21211

7. Analysis of results

Based on PCA and K-mean clustering methods, this paper analyzes several economic indicators of cities and towns in Guangdong Province in 2022, and the study shows that:

Principal component analysis shows that Guangzhou and Shenzhen have the highest economic scores, followed by Dongguan and Foshan, indicating that economic development within Guangdong Province is uneven, with other cities lagging behind. Cluster analysis further confirms this disparity, placing Guangzhou and Shenzhen in one category, Dongguan and Foshan in another, and the remaining cities in a third category.

The reasons for such a huge gap between cities are both "external" and "internal". In terms of external factors, from the raw data found, the largest amount of foreign investment in several cities is gathered in the Pearl River Delta (PRD) region, led by Guangzhou and Shenzhen, while the cities in the interior of Guangdong Province, foreign investment is much less, a large number of FDI (Foreign Direct Investment) influx has contributed to the rapid development of the PRD region, more than 90% of the FDI is concentrated in this area, while the other regions receive less than 10% of FDI, this phenomenon exacerbates the disparity between the various municipalities so huge. While other regions received less than 10% of FDI, this phenomenon exacerbated the polarization effect of the PRD region on the surrounding areas [7]. From an internal point of view, the economic policies of the municipal governments are also "important pushers" of economic development. Take the Shenzhen Municipality, which is famous for its economic development, as an example, the Shenzhen Municipal Government has made a lot of very effective policies for economic development, and not only has it introduced the "Ten Major Programs" to promote high-quality economic and social development [7]. To promote high-quality economic and social development, but also issued the encouragement of multinational corporations to set up headquarters in Shenzhen, as well as 5G high-quality industry plans and so on [8, 9]. All of these policies have pushed the Shenzhen economy forward in a big way.

8. Conclusion

This paper analyzes the economic development of cities in Guangdong Province through principal component analysis and k-mean clustering method and finds that there is indeed a big gap between the economic development of cities in Guangdong Province, so each city should designate corresponding strategies according to local conditions. For the Pearl River Delta (PRD), where economic development is at the forefront of the country, regional integration should be strengthened, and the construction of the Guangdong-Hong Kong-Macao Greater Bay Area should be further promoted to create an internationalized level of business and enhance regional competitiveness; industrial upgrading should be promoted, and investment in new and high tech industries should be vigorously strengthened and the proportion of traditional manufacturing industries should be appropriately reduced. For the eastern Guangdong - western Guangdong - northern Guangdong urban agglomeration, where economic development is relatively weak, the situation should be adapted to the local conditions. Coastal eastern Guangdong should strengthen its maritime advantages and vigorously develop its port economy, while the poorer cities in western and northern Guangdong should protect their ecological environment, develop agriculture with Chinese characteristics and eco-agriculture, vigorously carry out the slogan of "green water and green mountains are gold and silver mountains", and develop eco-tourism and health and wellness industries. At the same time, the provincial government should appropriately tilt resources to provide more policy support and resource allocation to cities with weaker economic development in the province, strengthen the infrastructure construction of economically backward cities, especially the transportation network, to promote interconnection between regions, and encourage the flow of foreign capital to the East

DOI: 10.54254/2753-8818/84/2025.21211

Guangdong-West Guangdong-North urban agglomeration through policies, so as to optimize the economic deconstruction.

References

[1] *Guangdong Provincial Government Local Chronicles Office, & Guangdong Yearbook Editorial Committee. (2023). Guangdong Yearbook 2023 (Vol. 37). Guangdong Provincial Government Local Chronicles Office.*

[2] *Li, S. (2018). Review and outlook on 40 years of economic development in Guangdong. In Proceedings of the 2018 World Economic Special Zone Development (Shenzhen) Forum (pp. 65-66).*

[3] *Zhao, M., & Zhang, L. (2008). A comparative study of the economic and industrial development of cities in Guangdong Province based on factor analysis. In Proceedings of the 6th China Management Science and Engineering Forum (pp. 191-196).*

[4] *He, B. (2017). Evaluation of fiscal revenue quality in Guangdong Province under the new normal: A factor analysis approach. In Proceedings of the 12th Graduate Academic Forum of Harbin University of Commerce (pp. 98-108).*

[5] *Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. Nature Methods, 14, 641-642. https://doi.org/10.1038/nmeth.4346*

[6] *Yu, N., & Chen, J. (2019). Economic development evaluation of the Yangtze River Delta urban agglomeration based on principal component analysis. In Proceedings of the 2019 Yangtze River Delta Higher Quality Integrated Development Forum (pp. 153-164).*

[7] *Wang, S., Zhao, Y., & Wang, Y. (2014). Multi-scale and multi-mechanism research on regional economic differences in Guangdong Province. Geographical Research, 34(10), 1184-1192.*

[8] *Shenzhen Development and Reform Commission. (2024, February 19). Report on the implementation of Shenzhen's 2023 National Economic and Social Development Plan and the draft plan for 2024. Shenzhen Development and Reform Commission. https://fgw.sz.gov.cn/fzgggz/jjxs/content/post_11148964.html*

[9] *Shenzhen Bureau of Science and Technology Innovation. (2024, November 4). Notice from the General Office of the Shenzhen Municipal People's Government on Further Promoting the Industrialization of Scientific and Technological Achievements. Shenzhen Bureau of Science and Technology Innovation. https://stic.sz.gov.cn/xxgk/zcfg/content/post_11694568.html*