

The Analysis of Economic Differences among Prefecture-level Cities in Hebei Province Based on K-means Clustering

Liping Du, Jiayu Zhang, Sa Li

College of Science, North China University of Science and Technology, Tangshan, Hebei, 063000, China

Keywords: K-means clustering; Economic differences; Regional economy

Abstract: The K-means clustering algorithm is applied to perform cluster analysis of 11 prefecture level cities in Hebei Province under six economic indicators such as analysis regional GDP, annual average salary, fiscal revenue, disposable income of rural residents, disposable income of urban residents, total volume of imports and exports. After standardizing the data from different regions in 2014 and 2019, the optimal number of clusters obtained through elbow method is four, which means that the 11 prefecture level cities are divided into four classes: economically developed areas, relatively developed areas, moderately developed areas, and underdeveloped areas. It is shown that Tangshan, Langfang, and Baoding have risen from the second, third, and fourth class to the first, second, and third class, respectively. Finally, an analysis of economic differences is conducted for different classes of regions, and relevant suggestions is provided based on the classification of economic development characteristics in each region.

1. Introduction

Hebei Province is surrounded by the capital cities of Beijing and Tianjin, adjacent to the Bohai Sea to the east, Shanxi Province to the west, Inner Mongolia and Liaoning to the north, and Shandong and Henan provinces to the south. Despite its unique geographical location in the Beijing-Tianjin economic circle and the Bohai Sea economic circle, Hebei Province has been positioned as a strategic place that reduces the functions of Beijing and an indispensable supporting role in the Beijing-Tianjin-Hebei coordinated development and integration policy. In order to effectively promote the implementation of the "three reductions, one reduction, and one supplement" policy, heavy industry, one of the best industries in Hebei Province, has begun to focus more on quality rather than quantity^[1], resulting in a lack of coordination^[1] in the economic, environmental, and energy development of Hebei Province. This is mainly reflected in the huge economic development gap between Hebei Province and Beijing and Tianjin in the Beijing Tianjin Hebei region. Therefore, the study of regional economic differences^[2,3] among prefecture-level cities in Hebei Province is of great significance for vigorously promoting the rapid development of the less developed areas within the province, narrowing the economic development differences among various regions, and helping Hebei Province to gradually form a good interaction based on

its own development situation, taking advantage of its proximity to Beijing and Tianjin, and further promoting the coordinated development of Beijing, Tianjin and Hebei.

2. Data and indicator selection

It is crucial to establish an effective recruitment process and establish corresponding recruitment standards in real estate enterprises. By clarifying recruitment needs, selecting suitable candidates, and adopting scientific interview evaluation methods, the accuracy and efficiency of recruitment can be effectively improved. This helps to introduce high-quality employees and add core competitiveness to the enterprise^[4]. It is particularly important to find a suitable indicator system for measuring regional economic strength^[5]. We have chosen the following indicators as indicators to measure the economic strength of 11 prefecture level cities in Hebei Province, namely, regional GDP(x_1), annual average salary(x_2), fiscal revenue(x_3), disposable income of rural residents(x_4), disposable income of urban residents(x_5), total volume of imports and exports(x_6).

We select the data of prefecture level cities in 2014 and 2019 to compare the regional economic differences in Hebei Province. K-means clustering analysis^[6,7] is used to cluster 11 prefecture level cities, and objective analysis is conducted on the economic development of each region over the past five years. In order to visually observe and study the changes in various indicators, the growth rate of each indicator is conculated in 2019 relative to 2014.

3. K-means clustering analysis

The K-means clustering analysis is one of the commonly used methods in clusterin. K represents the K-class, and means represents the center. Essentially, it is to determine the center points of the K-class, find these center points, and complete the clustering. At present, this algorithm has been used in many research problems.

3.1 Class and distance

Assuming a given dataset $A = \{a_i | a_i \in R_m, i = 1, 2, \dots, n\}$, and $T_i (i = 1, 2, \dots, k)$ representing k classes and $c(T_1), c(T_2), \dots, c(T_k)$ is the cluster centers of class k .

Definition 1. If vector $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ and vector $y_j = (y_{j1}, y_{j2}, \dots, y_{jm})$ respectively represent two data objects, then the Euclidean distance between the two vectors is

$$d(x_i, y_j) = \sqrt{\sum_{t=1}^m (x_{it} - y_{jt})^2}.$$

Definition 2. Definition 2. The centroid of data objects of the same class is defined as

$$c(T_i) = \frac{1}{|T_i|} \sum_{y_j \in T_i} y_j.$$

Where $|T_i|$ is the number of data objects T_i .

3.2 K-means clustering algorithm

The goal of the K-means clustering method is to cluster n objects into a specified k clusters based on their similarity, with each object belonging to and only belonging to the cluster with the smallest distance to the center of the cluster. The specific steps are as follows.

- (1) Determine the number of clusters k ;
- (2) Randomly select K points from the dataset as the initial clustering centers, which do not necessarily coincide with the data object;
- (3) According to the Euclidean distance formula, calculate the distance between each data object and each cluster center, and assign the data object to the nearest cluster;
- (4) Recalculate the centroid in each class, and the centroid serves as the new center for that class;
- (5) Repeat steps 3 and 4 until the clustering center no longer changes or the maximum number of iterations has been set, stop and output the clustering result.

3.3 Identification of the optimal K value

The key to the K-means clustering analysis is to determine the K value, and the elbow method is used here. As we know, when the number of clusters K increases, sample partitioning becomes more refined, the degree of aggregation of each cluster gradually increases, and the sum of squared errors (SSE) gradually decreases. The relationship graph between SSE and K shows the shape of an elbow, and the K value corresponding to this elbow represents the true number of clusters in the data. The formula for calculating the sum of SSE is as follows

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 .$$

Where C_i is the i th cluster, p is the sample point in C_i , m_i is the mean of all samples in C_i . The SSE is the clustering error of all samples, representing the quality of clustering performance. Based on data from various prefecture level cities in Hebei Province in 2014 and 2019, the elbow diagram is as follows.

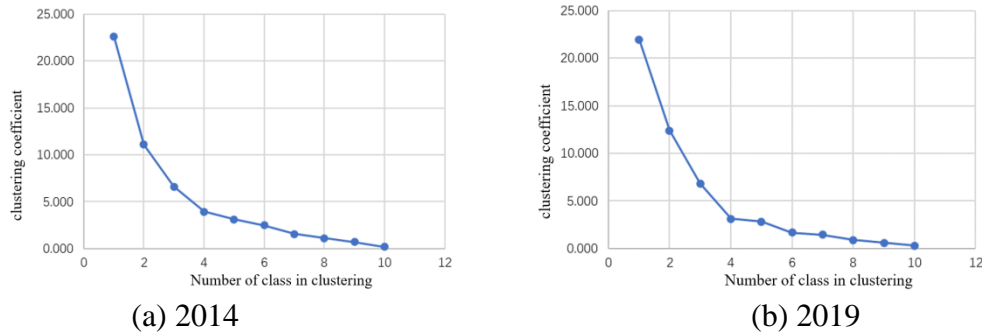


Figure 1: Elbow diagram in 2014 and 2019

From Fig 1, it can be seen that the K value corresponding to the elbow is 4, and the decrease in SSE decreases sharply. Then, as the k value continues to increase, it tends to flatten out. Therefore, the optimal number of clusters in this article should be 4.

3.4 Cluster results and analysis

According to the clustering results of various prefecture level cities, the clustering results in 2014 were divided into four classes. The first class was only Shijiazhuang City, the second class was Tangshan City, the third class included Qinhuangdao, Langfang, and Cangzhou, and the fourth class included Chengde, Zhangjiakou, Baoding, Hengshui, Xingtai, and Handan. And then, in 2019, except for Shijiazhuang, Tangshan has risen from the second class to the first class. The second class only leaves Langfang Originally belonging to the second class, Qinhuangdao and Cangzhou were downgraded to the third class. Therefore, the third class includes Qinhuangdao, Baoding, and

Cangzhou, while the fourth class includes Chengde, Zhangjiakou, Hengshui, Xingtai, and Handan.

As the capital of Hebei Province, Shijiazhuang is an important information center for politics, economy, science, finance, and technology. Due to the implementation of sustainable development policies, some industries in Shijiazhuang are subject to certain constraints, resulting in lower per capita GDP in 2019 compared to Tangshan and Langfang. However, due to its undeniable advantages in population, provincial capital, and policies, Shijiazhuang still has unlimited potential for future development. So, Tangshan has risen from the original second class to the first, alongside Shijiazhuang.

Langfang is located in the central eastern part of Hebei Province, also known as the Pearl of the Beijing Tianjin Corridor. As the name suggests, it forms a straight line with the two major cities of Beijing and Tianjin and is located in the middle. Its geographical location is superior, not only possessing abundant water resources, mineral resources, and plant resources, but also in a dominant position in scientific innovation and education in Hebei Province. From the analysis of the two clustering results, it can be seen that the annual average wage and per capita disposable income of urban residents in Langfang City have always been in a leading position, indicating that the quality of life of the people in Langfang is better than that of other regions.

The third class of city is called an economically developed area. In 2014, it was composed of Langfang, Qinhuangdao, and Cangzhou, but by 2019, Langfang had risen to the second class of economically developed areas. Qinhuangdao and Cangzhou are both located in the economic belt around the Bohai Sea. Among them, Qinhuangdao has unique historical resources and is the birthplace of China's modern tourism industry. It gathers rich tourism resources, with a mild climate and beautiful environment. It is a famous tourist city both at home and abroad, and is known as a paradise city. There are both advantages and disadvantages. The tourism cycle is too short, the number of high-tech enterprises introduced is relatively small, the real economy of industrial enterprises is limited, and there is no sufficient industrial production capacity, resulting in low local fiscal revenue. In addition, compared to 2014, the growth rate of Qinhuangdao's total import and export trade has significantly increased, indicating that Qinhuangdao has made good use of its coastal advantages to attract foreign investment and has achieved significant results in foreign trade. However, there is still a certain gap in the total amount compared to Shijiazhuang and Tangshan. Although Cangzhou and Tangshan are both coastal cities, there is a certain gap in economic strength and there is a risk of talent loss. However, Cangzhou, located in the core area of the Bohai Sea, has many advantages such as coastal port proximity, location transportation, and land resources, and is continuously optimistic about the future economic development of Cangzhou. Baoding is located in the central area of Hebei, in the hinterland of the triangle of Beijing, Tianjin, and Shijiazhuang. From the data, there are some similarities in the economic development of Baoding and Cangzhou, both of which are steadily rising. Baoding lacks a strong foreign investment economy and its urban construction is slightly lagging behind. However, the Xiong'an New Area established in 2017 is an opportunity for Baoding. More and more enterprises are choosing to survive here, and large enterprises will bring more employment opportunities and increase wage ratios. In summary, in 2019, cities located in economically developed areas are more suitable to be composed of Qinhuangdao, Cangzhou, and Baoding.

The fourth class of cities are areas with slow economic growth. Both clustering results show that Chengde, Zhangjiakou, Hengshui, Xingtai, and Handan are classified into one category. The per capita GDP of these five cities is relatively low compared to other cities, and the growth trend is relatively stable. In order to accelerate local economic development, some cities have taken different measures. Among them, Xingtai's foreign trade has increased by 42.50%, and local fiscal revenue is 62.73%, which is significantly higher than other cities. Due to the lower ranking of regional GDP, these five cities are still classified as cities with slow economic growth.

4. Conclusions and recommendations

From K-means clustering analysis, it can be clearly seen that Tangshan City, through its leading traditional heavy industry development, ranks among the top in various economic indicators compared to other prefecture level cities. As the provincial capital, Shijiazhuang has an unshakable political status and good development in import and export trade. Therefore, through comprehensive analysis, Shijiazhuang and Tangshan are collectively referred to as the first type of cities; Langfang's development is second only to the first type of cities, and some indicators are even higher than Shijiazhuang and Tangshan, making it known as the second type of city; Qinhuangdao and Cangzhou are steadily advancing, while Baoding's economy has improved by relying on the construction of Xiong'an New Area. The total value of these three regional indicators is smaller than that of the second category cities, but the difference is not significant, so they are called the third category cities; The remaining cities of Chengde, Zhangjiakou, Hengshui, Xingtai, and Handan are classified into one category and referred to as the fifth category.

On the basis of the economic boom of Hebei Province, to gradually form a good interaction, we will better promote the overall economic growth of Hebei Province, effectively promote rapid expansion of underdeveloped areas, narrow the economic development gap between different regions, and leverage the regional advantages of neighboring Beijing and Tianjin. We will further promote the policy of integrated coordinated development between Beijing, Tianjin, and Hebei, and propose the following suggestions for the four types of cities.

For the first class of cities, we will increase investment in fixed assets, cultivate talents for technological innovation, leverage a strong foundation of scientific research achievements, increase efforts to protect the ecological environment, optimize our own industrial structure, support vulnerable industries in the surrounding areas, strengthen infrastructure construction, and promote coordinated, stable, and sustainable development of regional economy, environment, and energy.

For the second class of cities, Langfang City should fully leverage its regional advantages as the pearl of the Beijing Tianjin Corridor and the resource advantages of the Bohai Rim economy, continuously strengthen economic exchanges and exchanges with the two major cities of Beijing and Tianjin, as well as the communication and exchange of excellent talents and new technologies, to promote the rapid development of regional economy.

For the third class of cities, we should continue to maintain stable economic development, increase employment rates, improve infrastructure construction, and enhance the quality of people's living standards. Coastal areas have always been important areas for economic development, and the vigorous development and construction of the Bohai Rim Economic Belt has promoted the rapid development of China's economy. Therefore, for regions such as Qinhuangdao and Cangzhou, it is more necessary to rely on their geographical advantages located in the Bohai Rim Economic Belt to promote high-quality regional economic development. At the same time, coastal areas should make full use of marine resources and engage in targeted foreign trade to achieve effective economic development and promote rapid development in underdeveloped areas.

For the fourth class of cities, they are areas with relatively backward economic development compared to the whole province. Among them, Handan is located in the heart of the Central Plains Economic Zone and is one of China's important transportation hubs. The data is relatively considerable compared to the other four provinces. It is necessary to fully utilize the historical natural resources of the region, improve the development system of the service industry, and maintain a level of green development. In the other regions, it is important to adopt some mutually reinforcing and progressive methods, including but not limited to increasing effective and high-quality investment, promoting the development of industrial entities with the aim of green and sustainable development, promoting supply side structural reform, and striving to achieve the joint

development of new and traditional industries.

Based on the above analysis, as an important regional location in North China, Hebei Province will steadily advance into the largest economic interaction circle in northern China, and the beautiful wish of coordinated development and integration of Beijing Tianjin Hebei will continue to be realized. In this development process, Hebei Province will become an important component of the largest urbanization region in northern China, and make significant contributions to the formation and development of this large urbanization region.

References

- [1] Li Jiena. *Analysis of Regional Economic Differences and Influencing Factors in Hebei Province* [D]. Hebei University of Economics and Business, 2022.
- [2] Xu Jin. *Analysis on the Relative Difference of the Relationship between Regional Population and Economic Growth* [J]. *Inquiry into Economic Issues*, 2021, (11): 81-90.
- [3] Yang Chengjia, Li Zhongxiang. *Development Level, Regional Differences and Distribution Dynamic Evolution of China's Digital Economy* [J]. *Statistics & Decision*, 2023, 39(09): 5-10.
- [4] Zhang Zengchen, Wang Pengzhen. *The Study on the Economic Strength of Various Districts and Cities in Hebei Province Based on Factor Analysis and Cluster Analysis* [J]. *Statistics and Management*, 2019(11):105-108.
- [5] Wei Yanhua, Ma Liping, Wang Bingcan. *Evaluation and Measurement of the Economic Development Differences among the Eight Comprehensive Economic Zones in China* [J]. *Journal of Quantitative & Technological Economics*, 2020, 37 (06): 89-108.
- [6] Pan Yurong, Jia Chaoyong, Rui Huaming. *Evaluation of Comprehensive Economic Strength of Zhejiang Cities Based on Principal Component Analysis and Cluster Analysis* [J]. *Journal of Baicheng Normal University*, 2022, 36(5): 64-70.
- [7] Huang Hengjun, Gao Haiyan, Zhang Menggao. *Functional Clustering Analysis: A Distance-Based One-step Framework* [J]. *Journal of Applied Statistics and Mangement*. 2019, 38(6):986-995.