

Different city economic evaluation models based on main analytic hierarchy Process model

Qian Sun*

School of Software Engineering, Harbin Engineering University, Harbin, 150001, China

*Corresponding author: sunqian20220323@163.com

Abstract. In this paper, 21 cities and autonomous prefectures in Sichuan Province are taken as samples, and the data extracted from 19 economic evaluation indexes are used to verify the correlation and partial correlation between index variables by KMO test and Bartlett sphericity test. Then, principal component analysis is performed, and the information is concentrated on several different factors through factor rotation. Finally, the ranking of its city's comprehensive economic strength is obtained. Combined with the cluster analysis model, the cities are divided into three categories by correlation according to the selected nineteen economic impact indicators. Based on quantitative principal component analysis and qualitative cluster analysis, the final comprehensive economic strength ranking of cities in Sichuan Province is given by comparing the results of the two rankings. The results show that Chengdu belongs to the first category of cities, with outstanding economic indicators, and other cities belong to the second or third category of cities.

Keywords: comprehensive economy; principal component analysis; cluster analysis; KMO and Bartlett test.

1. Introduction

Economic strength has always been an important indicator for evaluating a country and a city, and the level of economic development reflects residents' quality of life. As a result, people are eager to know the city's economic strength where they live, or the city they want to work, live, and ideally want in the future. Therefore, the ranking of urban economic strength has become a direction indicator and compass for people to choose a city. After 30 years' economic development in Sichuan Province, the economic growth rate of various cities has declined. In order to achieve rapid growth, it is necessary to carry out economic system reform. Therefore, this paper studies the factors that measure the economic development status, analyzes the contribution rate of the indicators to the economic development, and obtains the evaluation of urban economic status and ranks it.

At present, scholars at home and abroad have carried out research on urban economic evaluation. Liu Sifeng^[1] divided urban economic evaluation into four economic indicators, studied the correlation between sub-indicator and per capita GDP, and then used grey correlation clustering analysis. An indicator system that fully reflects the economic level is obtained. Zhao Licheng^[2] used the entropy weight method to calculate the final score, used DEA-Malmquist to measure the economic efficiency, and used the analysis of technical efficiency and technical progress index to formulate recommendations. Jia^[3] obtained a comprehensive evaluation of the economic development level of 47 countries through the combination of fuzzy set theory and Analytic Hierarchy Process (AHP), and used the superior and inferior solution distance method (TOPSIS). Yang^[4] conducted a comprehensive evaluation and research on Beijing's economic development level through AHP and weighted sum method. Luo and Tong^[5] comprehensively evaluated and compared the economic development capacity of 22 regions in China by using entropy weight method and factor analysis method. Although the above studies used a combination of different methods to optimize the model, their methods lacked data validation and could not well validate their own models.

This paper selects 19 comprehensive indicators of the comprehensive economic strength of cities, conducts principal component analysis (PCA dimensionality reduction) on 21 cities and autonomous regions, then uses the rotation matrix to make the dimensionality reduction results more meaningful, and finally obtains the city ranking. And through the cluster analysis to compare the two results, a

more reasonable ranking method is obtained comprehensively. Through qualitative cluster analysis and quantitative principal component analysis, the ranking of cities is more convincing.

2. Data acquisition and processing

This paper takes 21 cities and autonomous prefectures under the jurisdiction of Sichuan Province as the research objects, which is universal and random. First of all, it is necessary to determine the impact indicators of the comprehensive strength of the city's economy [2]. A total of 19 impact indicators are selected, and then the principal component analysis is carried out on the economic impact indicators of 21 cities and autonomous prefectures [6,7], and finally, new variables are obtained by combining the city indicators. By establishing a comprehensive economic strength evaluation system, we can reduce the dimension and analyze new variables, so as to get the city ranking. On this basis, in order to ensure the rationality of the city ranking, a cluster analysis [8] will be carried out to compare the two results, and a more reasonable ranking method will be obtained after careful consideration. The ranking of cities is more convincing through qualitative cluster analysis and quantitative principal component analysis.

The sample data in this paper are from "Sichuan Statistical Yearbook 2017" [9].

3. Model establishment

3.1 Model overview

In this paper, based on the main analytic hierarchy process, a large number of indicators indexes are reduced in dimension, and the weights of each index are calculated, so as to obtain the model results. The schematic diagram is shown in Figure 1.

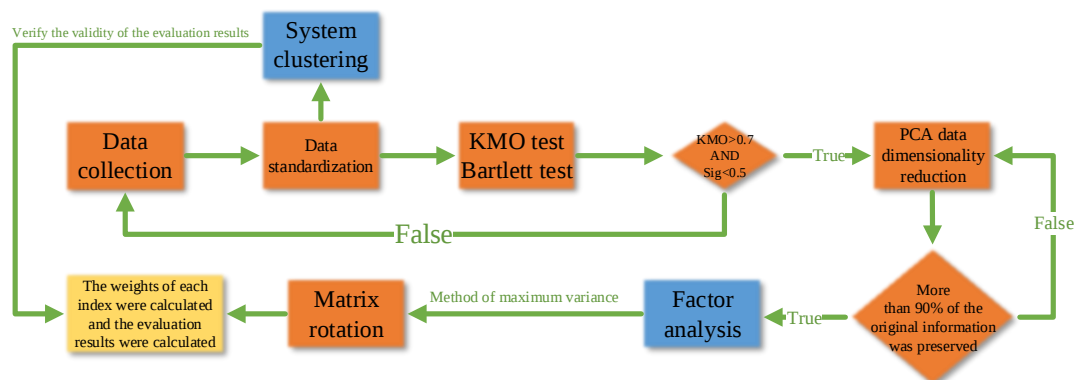


Figure 1 Schematic diagram of the economic evaluation model

After collecting the data of 19 indicators in 21 cities in Sichuan Province, the data matrix was standardized. Since the indicators are all positive indicators, the range method is not required for standardization. Suppose there are m samples, and each sample has n variables, which are respectively recorded as $x_1, x_2, x_3, \dots, x_n$. $m \times n$ constitutes a data matrix of order, and the standardized calculation formula is:

$$Y_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^n X_{ij}^2}}$$

Y_{ij} represents the normalized value corresponding to the element X_{ij} in the X .

3.2 PCA model establishment

3.2.1 KMO inspection

After data standardization, KMO (Kaiser-Meyer-Olkin) test is required, and the statistic is used to check the partial correlation between variables, and the value is between 0-1. The closer the KMO value is to 1, the stronger the partial correlation between variables, and the better the effect of factor analysis.

Partial correlation coefficient calculates the correlation coefficient between two variables under the influence of fixed remaining variables, which can reflect the degree of linear correlation between any two variables under the influence of fixed remaining variables. The formula for calculating the partial correlation coefficient is:

$$r_{xy.z_1 z_2 \dots z_h} = \frac{r_{xy.z_1 z_2 \dots z_{h-1}} - r_{xz_1 z_2 \dots z_{h-1}} r_{yz_1 z_2 \dots z_{h-1}}}{\sqrt{(1 - r_{xz_1 z_2 \dots z_{h-1}}^2)(1 - r_{yz_1 z_2 \dots z_{h-1}}^2)}}, (h \geq 2)$$

The formula h is the number of fixed indicators; $z_1, z_2, z_3 \dots z_h$ is the fixed indicator; r_{xy} is the correlation coefficient between the indicators x and y .

Suppose the sum of squares of correlation coefficients of variables is S , and the sum of squares of partial correlation coefficients is R , the calculation formula of KMO test statistic is:

$$K = \frac{S + R}{S}$$

This paper believes that when $KMO > 0.7$, it indicates that the partial correlation of the index data is strong, and factor analysis can be carried out.

3.2.2 Bartlett's test

Bartlett's test of sphericity is used to determine whether the correlation matrix is an identity matrix, that is, whether there is correlation between each variable[10]. It assumes that the correlation coefficient matrix is the unit matrix. When $Sig > 0.05$, it obeys the assumption of spherical test, and the variables are independent of each other, so factor analysis cannot be done. The statistic χ^2 of the Bartlett test is obtained according to the determinant of the correlation coefficient matrix P . The calculation formula of the statistic is:

$$\chi^2 = - \left[(n-1) - \frac{2k+5}{6} \right] \ln|p|$$

Query the chi-square distribution table according to the degrees of freedom and the observed value of the statistics, and the corresponding accompanying probability value can be obtained approximately. According to the relationship between the accompanying probability P and the significance level α , it is judged whether the variables are related and suitable for principal component analysis.

When $Sig < 0.05$, it is considered that each variable has a strong correlation, and factor analysis can be carried out.

3.2.3 Principal Component Analysis Correlation Coefficient Matrix Solving

Calculate the correlation coefficient matrix $R_{ij}(i, j = 1, 2, 3 \dots n)$ according to the standardized matrix Y_{ij} , where R_{ij} is the correlation coefficient between the index variables y_i and y_j

3.2.4 Solving for Eigenvalues and Eigenvectors of a Correlation Coefficient Matrix

According to the characteristic equation $|R - \lambda E| = 0$, the eigenvalue $(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$ is obtained, and the coefficient vector $(\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \dots, \alpha_{ni})$ is the eigenvector corresponding to the eigenvalue of the correlation coefficient matrix, where λ_i is the variance of the principal component y_i , and the larger the variance, the greater the contribution to the principal component.

3.2.5 Determine the principal components and calculate the variance contribution rate and the cumulative contribution rate of each principal component

The contribution rate of the i -th principal component is:

$$w_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} (i=1, 2, 3, \dots, n)$$

The contribution rate can reflect how much information the corresponding principal component indicators represent in the original n indicators. And indicate the influence of an influencing factor on the final data, similar to the weight rate of an indicator variable.

The cumulative contribution rate of the first n principal components is:

$$\sum_{i=1}^k w_i = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} (i=1, 2, 3, \dots, n)$$

It can reflect how much the corresponding top k principal components contain all the information content of the original variable. That is, the size of the cumulative weight and how much original information it contains. This paper believes that $\sum_{i=1}^k w_i \geq 90\%$ can better cover the original data information.

3.2.6 Computing the Elementary Loading Matrix

In order to make the actual meaning of each common factor more obvious, it is necessary to perform coordinate rotation again, which does not change the amount of information of each variable carried by the factor, but changes the amount of information carried by each factor, so that the actual meaning of each factor is obvious.

In this paper, the maximum variance orthogonal rotation method is used to make the variance of the factor loading as large as possible, that is, the information distribution is as uneven as possible, so that the information should be concentrated on several different factors.

where α_j is the normalized eigenvector and A is the factor loading matrix

$$\alpha_j = (\alpha_{1j}, \alpha_{2j}, \alpha_{3j}, \dots, \alpha_{nj})^T, j \in (1, n)$$

$$A = [\sqrt{\lambda_1} \alpha_1, \sqrt{\lambda_2} \alpha_2, \sqrt{\lambda_3} \alpha_3, \dots, \sqrt{\lambda_n} \alpha_n]^T$$

3.2.7 result solution

According to the steps of 3.2.5, we select the first K main factors, and then construct the factor model through the loading matrix A : where F represents the relationship expression matrix between the K main factors and all the indicator variables

$$F = YA$$

The final score is matrix Q , which represents the comprehensive evaluation results of all samples:

$$Q = \frac{\sum_{i=1}^k w_i \times F_i}{\sum_{i=1}^k w_i}$$

3.3 Cluster analysis model establishment

The data samples are grouped according to the similarity or dissimilarity of the features of the patterns to be classified, so that the data in the same group is as similar as possible, and the data in different groups is as different as possible. It is intended for data sample feature discovery rather than prediction^[8].

3.3.1 Variable correlation coefficient

Note the value $(x_{1j}, x_{2j}, \dots, x_{mj})^T \in R^n (j = 1, 2, \dots, m)$ of the variable x_j . Then the sample correlation coefficient of two variables x_j and x_k can be used as their similarity measure:

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\left[\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \right]^{\frac{1}{2}}}$$

3.3.2 Variable Maximum Coefficient Method

In the maximum coefficient method, the distance between two types of variables is defined as: $R(G_1, G_2) = \max(r_{jk})(x_j \in G_1, x_k \in G_2)$

At this time, $R(G_1, G_2)$ is equal to the similarity measure between the two variables with the least similarity in the two classes. The larger the value, the smaller the similarity.

4. Case Analysis and Model Solving

4.1 Data processing

This paper takes 21 cities and autonomous prefectures in Sichuan Province as the research objects. First, by consulting the existing complete and detailed economic evaluation system, referring to the selected economic impact indicators, and combining the actual situation of each city in Sichuan Province, the economic indicators are chosen comprehensively. A total of 19 economic impact indicators were selected.

The data indicators required by the two models used in this study are primary industry, secondary industry, tertiary industry, population, employed persons, total wages of employed persons, urban area, investment in fixed assets, social sales, total retail sales, public budget revenue, public budget expenditure, per capita disposable income of urban residents, disposable income of rural residents, added value of private economy, consumption expenditure of urban residents, consumption expenditure of rural residents, energy consumption of regional GDP, total expressway mileage, high schools. For the convenience of narration, these 19 indicators are mapped with numbers 1-19. The sample data involved in this article are all from the source "Sichuan Statistical Yearbook 2017".

First, in order to eliminate the influence of dimensions on the data, the indicator types should be treated uniformly. Standard deviation standardization was used to standardize the sample data. The standardized matrix $Y_{ij} \in [0, 1]$ is a dimensionless index observation, and the standardized data is obtained by SPSS, as shown in Figure 2.

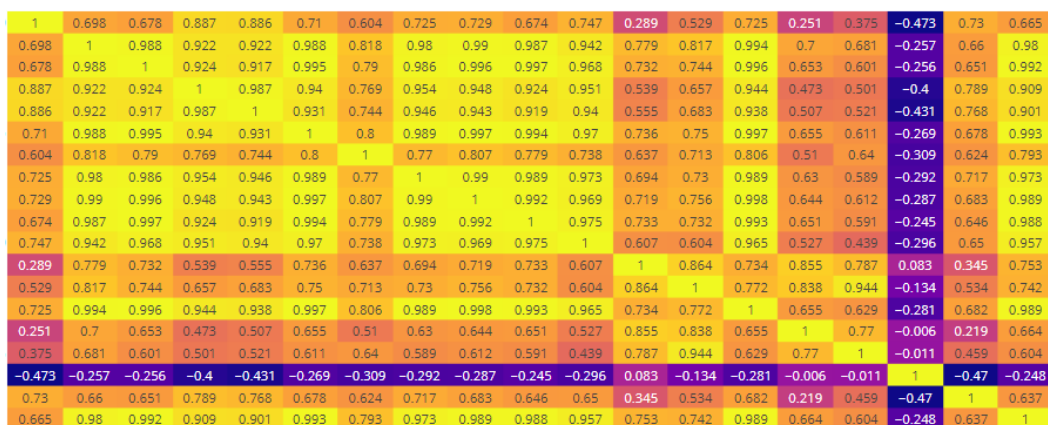


Figure 2 Normalized Numerical Matrix Heatmap

4.2 KMO and Bartlett test

Since there is a partial correlation between variables, KMO test and Bartlett test sphericity test are used to accurately test whether subjects are suitable for principal component analysis.

The statistical analysis software SPSS was used to KMO and bartlett test, and the results are as follows:

Table 1 KMO and Bartlett's test table

Kaiser-Meyer-Olkin measure		0.767
Bartlett test	Saliency	0.003

KMO=0.767>0.7 and Sig=0<0.05 in the Bartlett test, indicating that the data is suitable for factor analysis.

4.3 Variance interpretation table

According to the third step of the principal component analysis, using the correlation coefficient matrix R calculated above, according to the characteristic equation $|R - \lambda E| = 0$, the eigenvalue $(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$ is obtained, and the coefficient vector $(\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \dots, \alpha_{ni})$, $(i = 1, 2, 3, \dots, n)$ is the eigenvector corresponding to the eigenvalue of the covariance matrix, where λ_i is the principal component y_i . The greater the variance, the greater the contribution to the principal components. Table 2 is obtained by arranging the eigenvalues from large to small, which lists three eigenvalues, each of which corresponds to a component.

Table 2 Variance interpretation table

Total variance explained									
Element	initial eigenvalues			Extract the load sum of squares			Rotational load sum of squares		
	total	percent variance	Cumulaion %	total	percent variance	Cumulaion%	total	percent variance	Cumulaion %
1	14.506	76.346	76.346	14.506	76.346	76.346	9.889	52.049	52.049
2	2.090	11.001	87.347	2.090	11.001	87.347	4.946	26.032	78.081
3	.933	4.909	92.256	.933	4.909	92.256	2.693	14.175	92.256

First, look at the data in the "Total" column and the "Cumulative %" column under "Initial Eigenvalue". You can find that the eigenvalue corresponding to the first principal component is 14.506, and the contribution rate is 76.346%, which means that the first component. It contains 76.436% of the information of the original economic indicators. From the cumulative percentage of the first three components, it can be seen that the contribution rate is 92.256%, which is higher than the standard 90%. Therefore, extracting the first three components can well explain the information content contained in the original 19 economic indicators. Therefore, the principal component is selected. The number is determined to be 3.

"Sum of rotation squared loading" means that the variance contribution value, variance contribution rate and variance cumulative sum of the new three components are obtained after the factor rotation of the maximum variance method. It can be seen from the table that the variance percentage of the first three components after rotation are 52.049%, 26.032%, and 14.175%, respectively. It can be seen that the contribution rate distribution is more even.

4.4 Rotation Composition Matrix Table

Figure 3 is the component loading matrix after rotation. The load distribution of the corresponding economic indicators on each component after rotation is very clear, and it is easier to understand the meaning of each component than the unrotated component matrix.

As can be seen from Figure 3, the first component load is relatively high: primary industry, secondary industry, tertiary industry, employed persons, total wages of employed persons, investment in fixed assets, total retail sales of social sales, public budget Income, public budget expenditure, added value of private economy, and colleges and universities fully represent the financial operation of the government and society; The second component with a higher load is the per capita disposable income of urban residents, the per capita disposable income of rural residents, urban residents Consumption expenditure and rural residents' consumption expenditure fully reflect the income and expenditure level of the people, and reflect the quality of life of the people; The third component with a higher load is the total mileage of expressways, which reflects the urban traffic level.

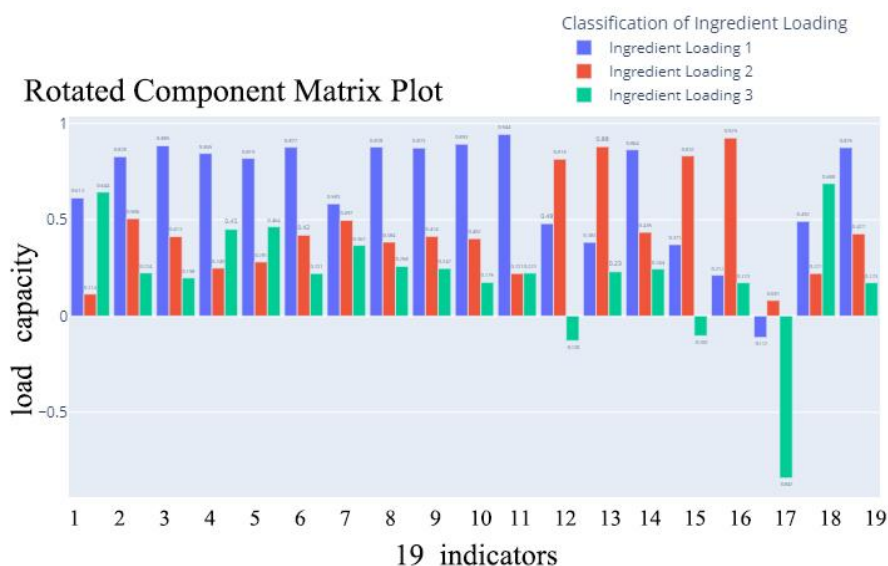


Figure 3. Histogram of load matrix after rotation

4.5 Calculate final score

Multiply the rotation matrix in 4.4 with the normalization matrix value to get the component score matrix $F_{21 \times 3}$ and the total score matrix $Q_{21 \times 1}$

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

Finally, the comprehensive score is obtained, as shown in Figure 4.

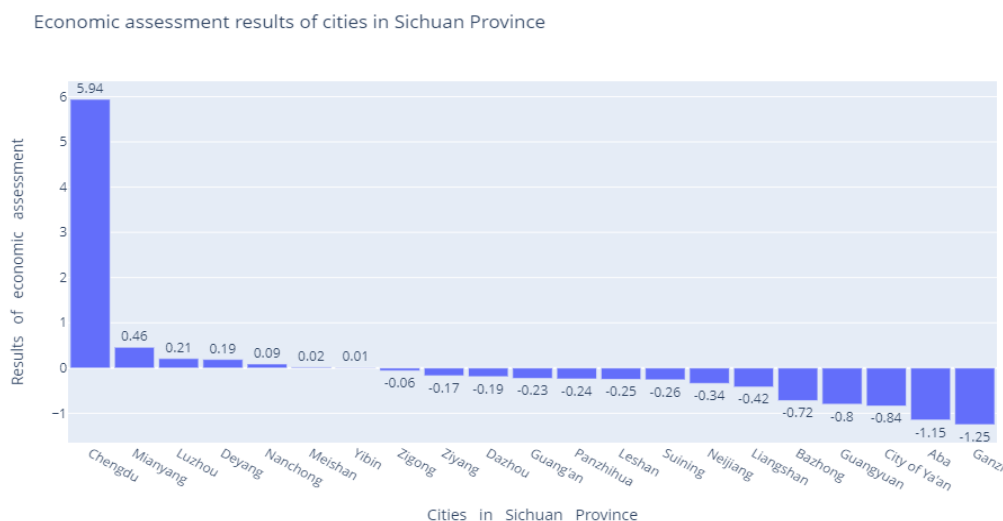


Figure 4 Economic Assessment Bar Chart

We have performed geographic visualization operations on the data in Figure 1, as shown in Figure 5. It can be seen that cities can be basically divided into three categories according to economic conditions: the upper left belongs to the poor areas, and the economic evaluation result is negative; The lower right belongs to the general type, the economic evaluation result is close to 0; While Chengdu in the middle belongs to the development and prosperity type, and the economic evaluation result is the maximum value.

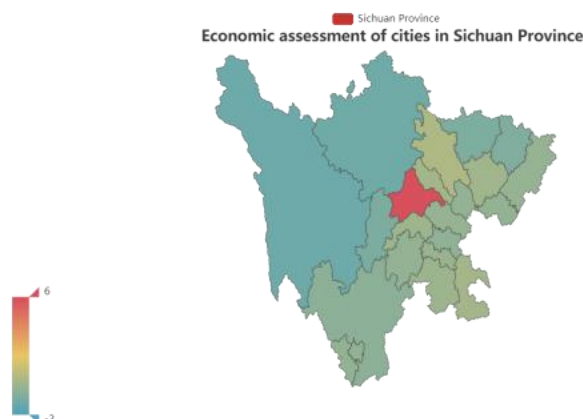


Figure 5 Geographical visualization of economic assessment in Sichuan Province

4.6 Cluster Analysis Model

This paper mainly adopts the method of systematic clustering analysis. The method is to take each sample as a class at the beginning, and then cluster the closest samples (that is, the group with the smallest distance) into subclasses first, and then group the aggregated subclasses according to the distance between the classes is merged again, and it continues, and finally, all subclasses are aggregated into one large class.

According to the standardized matrix Y_{ij} , the Wald method is used, the squared Euclidean distance is used to measure the interval, and finally, the pedigree map is obtained. It can be seen that Chengdu in Figure 6 is the first-class city; The second-class cities are Mianyang, Luzhou, Deyang, Nanchong, Meishan, Yibin, Zigong, Ziyang, Dazhou, Guang'an, Panzhihua, Leshan, Suining, Neijiang, Liangshan Yi Autonomous Prefecture; While Bazhong, Guangyuan, Ya'an, Aba Tibetan and Qiang Autonomous Prefecture, and Ganzi Tibetan Autonomous Prefecture are third.

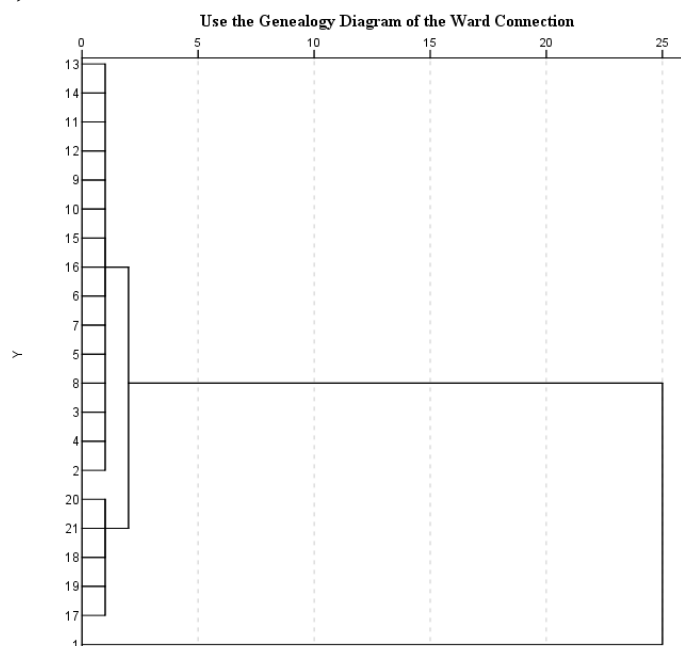


Figure 6 Pedigree diagram

According to the classification basis, we averaged the values of 19 indicators, and finally got the comparison of 19 indicators of the three types of cities in Figure 7. It can be seen that in the first category of cities, basically all indicators have the largest positive value. The energy consumption of the regional GDP is negative, indicating that the energy consumption of Chengdu is small, and the economic development is rapid, indicating that the economic structure transformation effect is remarkable; While the index value of the second type of city is basically around 0, and the development is average; Cities are poor cities, and the values of each index are basically the minimum values. It can confirm the results we got in the main AHP.

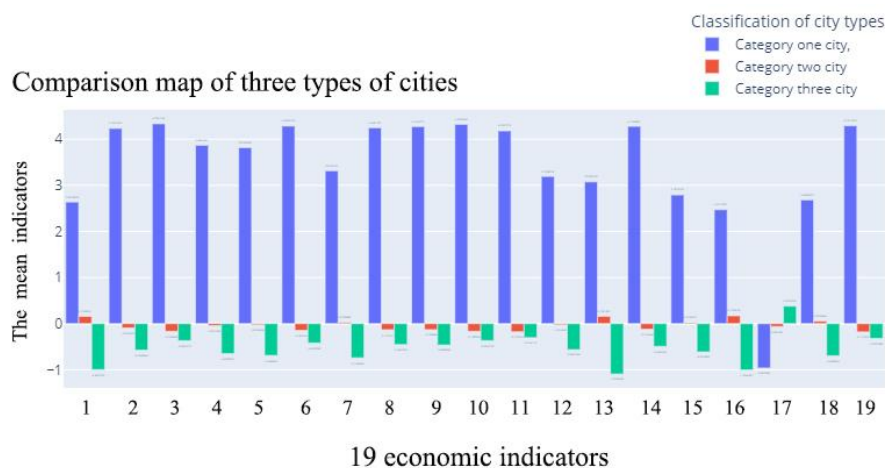


Figure 7 Histogram of three types of city indicators

5. Conclusion

In this paper, the principal component analysis model is used to quantitatively analyze the 19 economic impact indicators of 21 cities and autonomous prefectures in Sichuan Province in 2017. In addition, using the cluster analysis model, the comprehensive economic strength of cities can be qualitatively divided into three categories. These results are compared with each other to confirm that the most accurate ranking results of comprehensive strength of urban economy are obtained.

The strength ranking results are Chengdu, Mianyang, Luzhou, Deyang, Nanchong, Meishan, Yibin, Zigong, Ziyang, Dazhou, Guang'an, Panzhihua, Leshan, Suining, Neijiang, Liangshan Yi Autonomous Prefecture, Bazhong City, Guangyuan City, Ya'an City, Aba Tibetan and Qiang Autonomous Prefecture, and Ganzi Tibetan Autonomous Prefecture.

However, both the principal component analysis and cluster analysis models have their own shortcomings. The principal component analysis is that the meaning of the function is not clear and can only be comprehensively considered through research experience and actual situations. The disadvantage of cluster analysis is that it reflects the correlation between classes, but some indicators have a precise relationship in processing, but there is no intrinsic correlation in practice.

References

- [1] Liu Sifeng, Zhu Yongda. Regional Economic Evaluation Index and Triangular Membership Function Evaluation Model [J]. Chinese Journal of Agricultural Engineering, 1993(02):8-13..
- [2] Zhao Licheng, Ren Chengyu Reevaluation of economic efficiency of Bohai Rim Economic Circle from the perspective of green economy -- Based on entropy weight method and DEA Malmquist method [J] Resource development and market, 2012,28 (01): 45-47
- [3] Junsong Jia,Ying Fan,Xiaodan Guo. The low carbon development (LCD) levels' evaluation of the world's 47 countries (areas) by combining the FAHP with the TOPSIS method[J]. Expert Systems With Applications,2012,39(7):

- [4] Yanfang Yang,Xiling Li,Haixia Zheng. Analysis on Beijing's Low-Carbon City Evaluation Index System[C]//Education and Management(ISAEBD 2011 CCIS 210).,2011:177-183.
- [5] Zheng-qing Luo,Xiao-fang Tong. Evaluation on Development Capability of Low-carbon economy and Countermeasures in China[C]//Proceedings of the 2011 3rd International Conference on Environmental Science and Information Application Technology(ESIAT 2011 Part 1-2).,2011:437-442.
- [6] Zhang Peng, research on Comprehensive Evaluation Based on principal component analysis [D] Nanjing University of technology, 2004
- [7] Zhang Hui, Zhao Qiuhong Screening method of economic indicators based on the basic principle of principal component analysis [J] Journal of Shandong University of finance, 2013 (02): 52-61
- [8] Cao Kaidi, Xu Tingyu, Liu Yun Summary of cluster analysis [J] Smart health, 2016,2 (10): 50-53
- [9] Sichuan Bureau of statistics Sichuan statistical yearbook 2017 Website of Sichuan Bureau of Statistics: <http://tjj.sc.gov.cn/>
- [10] Xie K, ZHANG J F. Short-term wind speed prediction using PCA-WNN based on KMO-Bartlett Typical Wind Speed [J]. Power generation equipment,2017,31(02):86-91