Multivariate Statistical Analysis of Regional Economic Development in Sichuan Province

Liping Sun[1], Hongjun Yuan[1], Lingyun Hu[2,*]

[1] School of Statistics and Applied Mathematics , Anhui University of Finance and Economics, Bengbu, Anhui 233030, China

[2] School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China

*120081270@aufe.edu.cn

Abstract

The problem of uneven regional economic development is widespread in China, as is the economic development of cities within Sichuan Province. This paper selects 12 macroeconomic indicators from four aspects: economic development, education and human resources, transportation and medical security, and comprehensively analyzes the economic development of 21 prefecture-level cities in Sichuan Province in 2019. Firstly, the factor analysis method was used to reduce the dimensionality of 12 variables, and the scores and comprehensive scores of each factor were calculated. Secondly, cluster analysis is carried out based on the comprehensive scores of factors of various cities at the local level, so as to have a full understanding of the economic development of various regions in Sichuan Province.

Keywords

Regional economy; Sichuan Province; factor analysis; cluster analysis.

1. INTRODUCTION

With the deepening of reform and opening up, coupled with the vast territory of China, not only the social and economic phenomenon of uneven economic development between the east and west has been formed, but also the economic development between provinces. The issue of uneven regional economic development has attracted the attention of most scholars. Li Xuemei and Zhang Suqin (2009) used principal component analysis and factor analysis to comprehensively analyze the economic development of 28 provinces and municipalities (autonomous regions) in China, and the results showed that the economic development of the eastern region was relatively good, and its comprehensive score ranked relatively high, while the comprehensive score of the western provinces ranked relatively last[1]. Meng Ying, Xie Shouxiang and Peng Xiao (2010) used multiple statistical analysis methods to comprehensively evaluate the economic development of cities in Jiangsu Province through four systems: economic development, education and human resources, transportation and communication conditions, and health and environmental protection[2]. Xu Jiali and Qu Shili (2020) comprehensively analyzed the economic development of nine cities in Fujian Province based on multivariate statistical analysis [2]. Liu Xiuli, Guo Haizhen, Zhang Bo and Wu Pansheng (2018) used multivariate statistical analysis methods to explore the economic level and differences of cities in Shanxi Province from 2005 to 2014, and the results showed that the internal economy of Shanxi Province showed a downward trend over time[11]. Huang Shuyuan and Martin Ugly

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

(2020) comprehensively analyze the internal economic development of Gansu Province by using factor analysis and cluster analysis methods from the four subsystems of foundation, people's livelihood, economy and public advice [14]. Sichuan is located in the southwest region of China, and the geomorphological differences between cities are relatively large, and the terrain is complex and diverse, resulting in obvious differences in the economic development level of each city. Therefore, this paper selects 12 economic indicators from the perspectives of economic development, education and human resources, transportation, and medical security, and uses factor analysis and cluster analysis methods to comprehensively evaluate and analyze the economic development of 21 prefecture-level cities in Sichuan Province, so as to have a clearer understanding of the economic situation in the region.

2. THEORETICAL INTRODUCTION TO STATISTICAL METHODS

2.1. Data Standardization

Usually the units of measurement for each variable are different, so data preprocessing is done before building a mathematical model. The role of preprocessing is mainly twofold, one is to process the data without dimension, and the other is to homotropize the index[4]. Since the evaluation index system constructed in this paper all describes the economic development status of local cities from the same direction, only dimensionless processing is required. Based on the characteristics of simple calculation and wide application, this paper uses the standardized method for preprocessing, and the calculation formula is:

$$y_i = \frac{x_i - \overline{x}}{S} \tag{1}$$

where mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, variance $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

2.2. Principles of Factor Analysis

Factor analysis is a method of reducing dimensionality and simplifying data[10-13]. It explores the basic structure of the observed data by studying the internal dependencies between many variables, and represents its basic data structure with a few "abstract" variables[4]. Based on the different research objects, factor analysis can be divided into two types, one is R-type factor analysis, which mainly analyzes factors for variables, and the other is Q-type factor analysis, which mainly focuses on samples. Since this paper studies the economic development level of various cities in Sichuan Province and comprehensively evaluates the economic level of various regions by using different indicators, R-type factor analysis is used for research and analysis.

There are n samples, each sample has p indicators, and there is a strong correlation between p indicators, and the expression of the coefficient model is:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i \tag{2}$$

where X_i represents the observation of the i th indicator $(i = 1, 2, \cdots, p; m < p)$, $F_1, F_2, \cdots, F_m$ represents the common factor, ε_i represents the special factor of X_i, and a_{ij} represents the load of index i on the common factor j.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

Based on equation (2) can be deformed to obtain the matrix expression between each indicator and common and special factors, there is:

$$X = AF + \varepsilon \tag{3}$$

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} = (A_1, A_2, \cdots, A_m)$$
$$X=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad F=\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \quad \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \tag{4}$$

where X represents the observation matrix of each index, A represents the factor load matrix, F represents the common factor vector, and ε represents the special factor vector.

2.3. Principles of Clustering

2.3.1 Distance Formula for Clustering

"Things are clustered together, and people are grouped." People's understanding of things is no longer from the individual level, but from the level of parts or the whole. Therefore, cluster analysis is widely used in the economic field as a method of categorical grouping[5,6,9]. According to the different research objects, cluster analysis can be divided into two types, one is Q-type clustering, and the research object is aimed at samples; The other is R-type clustering, which studies variables for the subject. Since this paper uses Q-shaped cluster analysis based on factor analysis of the economic development level of cities in Sichuan Province and attempts to classify and study the comprehensive score of factors of each city and analyze the economic differences between prefecture-level cities belonging to different categories.

Before clustering, you first analyze the similarity between samples. The method of measuring the degree of similarity between samples mainly uses distance measures. The methods of distance measurement mainly include Minkowski distance, Lance and Williams distance, and Mahalanobis distance. At present, scholars commonly use Mahalanobis distance and the Euclidean distance in Minkowski distance. Let n samples be regarded as n points in p-dimensional space, then the distance between the points of samples X_i and X_j is expressed as d_{ij}.

(1) The formula for calculating Minkowski distance is:

$$d_{ij}(q) = \left(\sum_{k=1}^{p} \left| X_{ik} - X_{jk} \right|^q \right)^{\frac{1}{q}} \tag{5}$$

When $q = 1$, the Minkowski distance is simplified to an absolute distance whose expression is $d_{ij}(1) = \sum_{k=1}^{p} \left| X_{ik} - X_{jk} \right|$;

World Scientific Research Journal Volume 9 Issue 2, 2023
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

When $q=2$, the Minkowski distance is simplified to the Euclidean distance, which is expressed as $d_{ij}(2)=\left(\sum_{k=1}^{p}\left|X_{ik}-X_{jk}\right|^{2}\right)^{\frac{1}{2}}$;

When $q=\infty$, the Minkowski distance is simplified to the Chebyshev distance, which is expressed as $d_{ij}(\infty)=\max_{1\le k\le p}\left|X_{ik}-X_{jk}\right|$.

(2) The calculation formula of Mahalanobis distance is:

$$d_{ij}^{2}(M)=(X_i-X_j)'\sum^{-1}(X_i-X_j) \quad (6)$$

where X_i and X_j represent p-dimensional samples, and $\sum^{-1}$ represents the covariance matrix between samples.

2.3.2 Systematic Clustering Method

The basic idea of the systematic cluster analysis method is to cluster the samples first and then cluster the classes according to the principle of distance, and the process continues until each sample is clustered into the appropriate category. The methods for calculating the distance between classes mainly include the single linkage method , the centroid method, the group average method and the Ward method. Since the group average method has the characteristics of using sufficient information between samples, this method is selected to calculate the distance between classes.

3. EMPIRICAL ANALYSIS OF ECONOMIC DEVELOPMENT RESEARCH IN VARIOUS CITIES IN SICHUAN PROVINCE

3.1. Data Source and Evaluation Index Selection

The macro data used in this article are all from the Sichuan Statistical Yearbook 2020 released by the Sichuan Provincial Bureau of Statistics.

A city's comprehensive economic strength usually refers to the full strength, potential, and position and influence of a city in the domestic and foreign economy[7]. The economic development of each city and prefecture can be understood through the comprehensive economic strength of each city. Therefore, when comprehensively analyzing the economic development differences of various regions in Sichuan Province, the indicators are selected according to the principles of comprehensiveness, typicality and data availability, so as to comprehensively evaluate and analyze the economic differences of various regions in a more comprehensive way. According to the principle of index selection, this paper selects 12 economic indicators from four aspects: economic development, education and human resources, transportation and medical security, and uses factor analysis and cluster analysis methods to comprehensively analyze the economic conditions of cities in Sichuan Province, see Table 1.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

Table 1. Evaluation index system

Numbler	Name	Numbler	Name
X1	Per Capita GDP	X7	Quantity of Full-time Teachers
X2	Total Retail Sales of Consumer Goods	X8	Students Enrollment
X3	Total Value of Foreign Trade Imports and Exports	X9	Road mileage
X4	General Budgetary Revenue	X10	Turnover Volume of Passenger Traffic
X5	Number of employees at the end of the period	X11	Health Care Institutions
X6	Average Wages of Employees	X12	Urban Basic Pension Insurance

3.2. Factor Analysis of Economic Development Level of Cities in Sichuan Province

Firstly, equation (2.1) is used to standardize the original data; secondly, factor analysis is carried out based on the standardized data, and the KMO and Bartlett test results, the eigenvalues and variance contribution rate of common factors, the factor loading matrix and the factor score are extracted in turn. Finally, according to the rotated factor load matrix, the factor score expression and the final comprehensive score value of each local city are determined, see Table 2.

Table 2. KMO and Bartlett Tests

The number of KMO sampling cutness		0.766
Bartlett Tests	approximate chi-square	718.812
	df	66
	p-value	0.000

Before factor analysis, the correlation between variables is first tested, and when the correlation coefficient is greater than 0.7, it indicates that factor analysis is possible between the variables. From Table 2, it can be seen that the KMO test result is 0.766 and the Bartlett test has a significance of 0, rejecting the null hypothesis. Both KMO results and Bartlett results indicate that the indicator system can be analyzed for factor analysis and can proceed to the next step.

Table 3. Eigenvalues and variance contributions

Component	Eigenvalue	The contribution rate of each factor (%)
1	9.116	75.965
2	1.321	11.009
3	1.047	8.727
4	0.214	1.783
5	0.144	1.198
6	0.102	0.849
7	0.042	0.351
8	0.008	0.069
9	0.004	0.037
10	0.001	0.008
11	0.001	0.005
12	0.000	0.000

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

The results from Table 3 show that the common factors with characteristic values greater than 1 are mainly distributed in the first three factors. At the same time, considering the cumulative contribution rate of the first three factors, its value reached 95.701%, which indicates that the first three common factors explained 95.701% of the information in the original data, which was well representative. Therefore, the first three common factors are extracted in this paper, see Table 4.

Table 4. Rotated factor load matrix

var	F1	F2	F3
X1	0.690	-0.566	0.291
X2	0.982	0.024	0.163
X3	0.943	-0.035	0.268
X4	0.966	-0.006	0.229
X5	0.974	0.179	-0.021
X6	0.114	0.154	0.971
X7	0.975	0.194	0.055
X8	0.975	0.194	0.054
X9	0.319	0.853	0.325
X10	0.893	0.200	0.202
X11	0.862	0.431	-0.037
X12	0.960	-0.036	0.235

From Table 4, it can be obtained that the variables X2, X3, X4, X5, X7, X8, X10, X12 have a large load on the first common factor, and their variance contribution rate reaches 75.965%, so they are defined as economic strength factors; The variables X9 and X11 showed large loads on the second common factor, and their variance contribution rate reached 11.009%, so they were defined as economic environment factors. The variables X1 and X6 exhibit high loading on the third common factor, and their variance contribution rate reaches 8.727%, so they are defined as economic average factors.

Table 5. Factor score factor

var	F1	F2	F3
X1	0.099	-0.512	0.208
X2	0.120	-0.064	0.001
X3	0.103	-0.116	0.105
X4	0.110	-0.092	0.064
X5	0.135	0.072	-0.177
X6	-0.136	0.029	0.842
X7	0.123	0.077	-0.110
X8	0.123	0.077	-0.111
X9	-0.060	0.611	0.173
X10	0.089	0.075	0.034
X11	0.103	0.277	-0.197
X12	0.110	-0.116	0.073

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

Based on Table 5, the score function of each factor can be obtained, as follows:

$$F1 = 0.099X1 + 0.120X2 + 0.103X3 + 0.110X4 + 0.135X5 + 0.136X6 + 0.123X7 + 0.123X8 + 0.060X9 + 0.089X10 + 0.103X11 + 0.110X12$$

$$F2 = -0.512X1 - 0.064X2 - 0.116X3 - 0.092X4 + 0.072X5 + 0.029X6 + 0.077X7 + 0.077X8 + 0.611X9 + 0.075X10 + 0.277X11 - 0.116X12$$

$$F3 = 0.208X1 + 0.001X2 + 0.105X3 + 0.064X4 - 0.177X5 + 0.842X6 - 0.11X7 - 0.111X8 + 0.173X9 + 0.034X10 - 0.197X11 + 0.073X12$$

The comprehensive score function can be obtained by calculating the weight of each common factor according to the contribution rate after rotation, where the weight calculation formula:

$$\alpha_i = \frac{y_i}{\sum_{i=1}^{n} y_i} \quad , \quad i = 1,2,3$$

,

where y_i represents the variance contribution rate after the rotation of the common factor Fi. According to the formula $F = 0.7566F1 + 0.1225F2 + 0.1208F3$, The comprehensive ranking of the economic development level of cities at all levels in Sichuan Province can be calculated, see Table 6.

Table 6. The scores of each factor and the results of the composite score

region	F1	F2	F3	F	rank
Chengdu	4.081	-0.125	1.201	3.218	1
Nanchong	0.321	1.526	-0.970	0.312	2
Liangshan Yi Autonomous Prefecture	-0.047	1.649	0.304	0.203	3
Mianyang	0.213	0.179	-0.180	0.161	4
Dazhou	0.167	0.911	-0.718	0.151	5
Yibin	0.134	0.262	-0.026	0.130	6
Luzhou	0.165	0.080	-0.509	0.073	7
Deyang	0.038	-1.198	-0.014	-0.120	8
Neijiang	-0.126	0.002	-0.445	-0.149	9
Leshan	-0.120	-0.648	-0.122	-0.185	10
Guangan	-0.204	0.043	-0.504	-0.210	11
Bazhong	-0.256	0.822	-1.167	-0.234	12
Guangyuan	-0.422	0.572	-0.164	-0.269	13
Suining	-0.193	-0.460	-0.732	-0.291	14
Ganzi Tibetan Autonomous Prefecture	-1.041	1.697	2.204	-0.314	15
Meishan	-0.243	-0.847	-0.291	-0.323	16
Ziyang	-0.389	0.085	-0.472	-0.341	17
Zigong	-0.182	-1.023	-0.833	-0.364	18
Aba Tibetan and Qiang Autonomous Prefecture	-1.034	-0.207	2.777	-0.472	19
Panzhihua	-0.399	-2.260	0.846	-0.476	20
Ya‘an	-0.463	-1.062	-0.186	-0.503	21

By ranking the scores of each factor, it can be found that the top three cities in the common factor $F1$ are Chengdu, Nanchong and Mianyang, while the factor score of Ganzi Tibetan

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

Autonomous Prefecture ranks the lowest among the 21 prefecture-level cities. The top three regions in the common factor $F2$ are Nanchong, Ganzi Tibetan Autonomous Prefecture and Liangshan Yi Autonomous Prefecture, while the bottom three cities are Ya'an, Deyang and Panzhihua; The top three regions in the common factor $F3$ are Aba Tibetan and Qiang Autonomous Prefecture, Chengdu and Ganzi Tibetan Autonomous Prefecture, while the bottom three regions are Zigong, Nanchong and Bazhong. Based on the comprehensive score of each factor score and ranking, it can be obtained that the top three cities in Sichuan Province are Chengdu, Nanchong and Liangshan Yi Autonomous Prefecture, of which Chengdu's comprehensive economic score is 3.218, much higher than the second place Nanchong; the city with the lowest factor comprehensive score is Ya'an, whose factor comprehensive economic score is -0.503.

Based on the scores of the three common factors, the economic strength factor accounted for the largest proportion of the comprehensive score, reaching 75.66%, and the other two common factors accounted for a relatively low proportion. Therefore, the higher the score on the economic strength factor, the higher the score on the overall economy and the higher the overall ranking.

3.3. Cluster Analysis Based on the Results of Factor Analysis of Economic Development Level of Various Cities in Sichuan Province

The systematic cluster analysis method was used to group the comprehensive factor scores of each city in Sichuan Province. According to the results of the genealogy chart of cluster analysis, the economic development of 21 prefecture-level cities in Sichuan Province can be divided into three categories, the first is Chengdu; The second category is Aba Tibetan and Qiang Autonomous Prefecture and Ganzi Tibetan Autonomous Prefecture; The third category is Zigong, Panzhihua, Luzhou, Deyang, Mianyang, Guangyuan, Suining, Neijiang, Leshan, Nanchong, Meishan, Yibin, Guang'an, Dazhou, Ya'an, Bazhong, Ziyang and Liangshan Yi Autonomous Prefectures, see Figure 1.

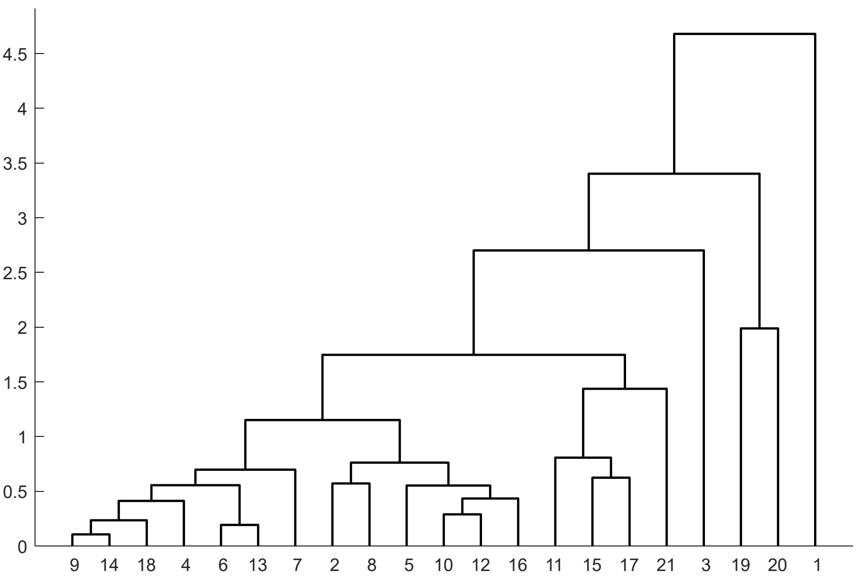

Figure 1. Genealogy chart

4. CONCLUSION

Using factor analysis and cluster analysis to comprehensively analyze the economic development of 21 prefecture-level cities in Sichuan Province, there are two outstanding

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

characteristics in the analysis results, one is that Chengdu is unique in both the comprehensive economic competitiveness score and the classification category, which is in sharp contrast with the second place. Second, the ranking of Panzhihua has changed significantly. Ten years ago, Xia Guoen found that Panzhihua's comprehensive economic development ranking was second[8]. The main reason is that Panzhihuanin is rich in metal resources and enjoys the title of "China's Vanadium and Titanium Capital", so its economic development is relatively good. Due to the destruction of the environment, people began to attach great importance to the construction of ecological civilization, which is the main reason why Panzhihua's economic development ranking is gradually lower. At the same time, under the background of the construction of the Chengdu-Chongqing Twin Cities Economic Circle, promote the development of Chengdu, Deyang, Meishan and Ziyang, accelerate the layout of productivity integration, promote the joint construction of infrastructure networks, the sharing of public services, and the opening up of government affairs, build a comprehensive pilot area for the integration of the capital and the Meiji capital, and build a central city-led, group-based multi-level network space structure, and promote the integrated development of northeast Sichuan, and explore the co-construction and sharing mechanism of resumes in the fields of industrial development and ecological environmental protection. For the western Sichuan region, vigorously develop the eco-tourism industry and build a national tourism demonstration area, so as to improve the level of economic development.

ACKNOWLEDGMENTS

This work is supported by Planning Project for Philosophy and Social Science in Anhui Province (Project number: AHSKY2020D42), Key Research Project for Natural Science in Colleges and Universities of Anhui Province(Project number: 2022AH050602), Major Research Fund Project of Anhui University of Finance and Economics (Project number: ACKYA21004), Key Project of Humanities and Social Sciences in Colleges and Universities of Anhui Province (Project number: SK2021A0228), Anhui University of Finance and Economics Graduate Research and Innovation Fund Project (Project number: ACYC2021400).

REFERENCES

[1] Li Xuemei, Zhang Suqin: Application of principal components analysis in regional economy analysis [J]. Computer Engineering and Applications,2009, 45(19): 204-206+238.

[2] Mengying, Xie Shouxiang, Peng Xiao: Multivariate Statistical Analysis of Regional Economic Differences in Jiangsu Province [J]. Special Zone Economy, 2010(4): 51–52.

[3] Xu Jiali,Qu Shili: Research on Regional Economic Development of Fujian Province Based on Multivariate Statistical Analysis [J]. China Circulation Economy, 2020(29): 85–88.

[4] Zhu Jianping: Applied Multivariate Statistical Analysis (Science Press, Beijing,2006).

[5] Zhang Xia,Xu Qi-fa: Measurement and Analysis of China's Provincial Economic High-quality Development in the New Era[J]. On Economic Problems,2021,499(03):16-25.

[6] Zhang Tonggong, Liu Jiangwei:The Regional Differences of Finance Supporting the Development of Real Economy in New Period [J]. Regional Economic Review,2018,33(03):84-95.

[7] Liu Xin, Wang Hong: The Factor Analysis Method was Used to Evaluate the Comprehensive Economic Strength of Cities in Sichuan Province[J]. Reform of Economic System,2002,(04):128-131.

[8] Xia Guo-en: Evaluation of the Comprehensive Economic Strength of Cities in Sichuan Province [J]. Journal of Zhengzhou Institute of Aeronautical Industry Management,2009,27(03):27-29.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202302_9(2).0001

[9] Wang Qing-feng, Dang Yao-guo, Wang Li-min:Study on County Economy Based on Factor and Cluster Analysis——with 18 counties in Henan Province as Examples[J]. Application of Statistics and Management, 2009, 28(03):495-501.

[10] Ren Sheng-gang,Peng Jian-hua:The Evaluation and Comparison of Regional Innovation Capacity Based on Factor Analysis[J]. Systems Engineering, 2007(02):87-92.

[11] Liu Xiu-li, Guo Hai-zhen, Zhang Bo, Wu Pan-sheng:Evaluation of regional economic development level in Shanxi Province based on factor analysis[J]. Journal of Northwest Normal University (Natural Science), 2018,54(02):102-107+120.

[12] Chen Wen-cheng:Region Economy Unbalanced Development Research Based on Factor Analysis[J].Journal of Applied Statistics and Management, 2010,29(03):490-501.

[13] Guo Li-ping , Chen Zhong-nuan:Analysis of Comprehensive Strength and Differentiation of Types on Regionally Tourist Economy in China[J]. Geography and Territorial Research,2001(03):88-91+96.

[14] Huang Shuyuan, Ma Dingchou: Multivariate Statistical Analysis of Regional Economic Development in Gansu Province [J]. Productivity Research, 2020, 340 (11): 80-84+145.