Research on the Urban Construction Status of Prefecture-Level Cities in Heilongjiang Province Based on SPSS Analysis

Shuguang Wang[a] and Haoyan Wang*[b]
School of Finance and Public Administration, Harbin University of Commerce, No. 1 Xuehai Street, Harbin, China

Keywords: SPSS Analysis, Software Cluster, Analysis Analysis of Factors, Construction of City.

Abstract: In the different analysis of regional construction and development, many experts and scholars believe that economic development is the main object of analysis, and use GDP index to evaluate the construction situation (Ning, 2001). Although the data are easy to obtain and can be used for vertical comparison of regional economic differences, the GDP indicators cannot fully and completely reflect the socio-economic development level of a region. Therefore, the multivariate statistical analysis method can be used to analyze the construction status of Heilongjiang Province (Wang and Zhao, 2018), which can reflect the real degree of its construction relatively completely and comprehensively.

1 INTRODUCTION

1.1 Cluster Analysis

Cluster analysis is a multivariate statistical analysis method to study individual classification according to the characteristics of things. The basic principle: the individuals in the same class have great similarities, and individuals in different classes differ greatly; There is a certain degree of similarity between the study variables. According to multiple observation indicators of the samples, the statistics that can measure the degree of similarity between the samples or variables are specifically found (Chen, 2015). Based on these statistics, some variables or samples with a greater degree of similarity are divided into two categories. The basic principle of this method is to directly compare the properties of all things in the sample, group those with similar properties into one category, and divide those with large property differences into different categories. That is, the difference in nature between the same kind of things is small, and the difference in nature between classes is large. Distance is the most commonly used way to describe the degree of kinship between samples, among which Euclidian distance is the most widely used in cluster analysis. Its expression is as follows (Tian and Zhai, 2018):

$$d_{ij} = \sqrt{\sum_{t=1}^{m.} |x_{it} - x_{jt}|^2}$$

Where, X_{it} represents the observed value of the K th index of the ith sample, X_{jt} represents the observed value of the K th index of the J th sample, and d is the Euclidean distance between the ith sample and the J th sample. If d is smaller, the properties of the two samples between i and j will be closer and closer. Samples with similar properties can be grouped together (He, 2015).

1.2 Analysis of Factors

Factor analysis is a data simplification technology. It explores the basic structure of observed data by studying the internal dependencies among many variables, and uses a few independent unobservable variables to represent its basic data structure.

The original variable is the explicit observable variable, while the imaginary variable is the unobservable latent variable and is called a factor. In order to analyze practical problems comprehensively and objectively, it is often necessary to consider the research objects in many aspects and collect multiple observation index data. If these indicators are

[a] https://orcid.org/0000-0001-6137-6915
[b] https://orcid.org/0000-0003-4023-4453

Wang, S. and Wang, H.
Research on the Urban Construction Status of Prefecture-Level Cities in Heilongjiang Province Based on SPSS Analysis.
DOI: 10.5220/0012071200003624
In *Proceedings of the 2nd International Conference on Public Management and Big Data Analysis (PMBDA 2022)*, pages 142-146
ISBN: 978-989-758-658-3
Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

analyzed one by one, it will undoubtedly lead to a one-sided understanding of the research object, and it will be difficult to reach a comprehensive and consistent conclusion. Factor analysis is to consider the mutual relationship between various indicators, using the idea of dimension reduction to convert multiple indicators into a few discrete comprehensive indicators so as to make the research relatively simple. The basic principle is to group variables according to the size of correlation so that the correlation between variables in the same group is higher and the correlation between variables in different groups is lower. Each group of variables represents a common factor, and each observed variable can be expressed as "the sum of the linear function of the common factor and the special factor." The factor load matrix and the common degree of analysis variables were studied by constructing a factor analysis model. The main common factors were selected according to the variance contribution of the common factors (Wang and Zhang, 2015). After the variance maximization of the factors was orthogonal rotation, the regression method was used to estimate the factor score, and the proportion of variance contribution of each factor was weighted and summarized to obtain a comprehensive evaluation. Its expression is:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)} - \mu\right)\left(x^{(i)} - \mu\right)^{T}$$

2 SPSS SOFTWARE WAS USED FOR DATA ANALYSIS

2.1 Selection of Indicators

Taking prefecture-level administrative units as regional analysis objects, the construction status of twelve prefecture-level cities in Heilongjiang Province was analyzed by multivariate statistical analysis method. The selection of indicators mainly follows the principles of representativeness, comprehensiveness, systematization, and accessibility, and fourteen indicators are selected as the basis for analysis. Its indicators are as follows(Wu and Li, 2018): Total population at the end of X1 (10,000), X2 local fiscal revenue (100 million yuan), X3 residential area (10,000 square meters), X4 health institutions (number), X5 per capita green park area (square kilometers), X6 urban water penetration rate (%), X7 urban gas penetration rate (%), X8 highways (kilometers), X9 number of middle schools (number), X10 social consumption Total retail sales of defective goods (RMB 100 million), the total output value of X11 (RMB 100 million), the total output value of agriculture, forestry, animal husbandry and fishery (RMB 100 million), number of industrial enterprises (units), urban construction area (square kilometers) of X14. See Table 1 for details.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 1: Basic (original) data of twelve prefecture-level cities in Heilongjiang Province.

Number	region	X1	X2	X3	X4	X5	X6
1	Harbin	1000.1	339.5674	5490.2888	4393	51.53	100
2	Qiqihar	403.7	74.5843	1095.2536	2735	9.71	100
3	Mudan Jiang	227.9	54.5433	1538.7163	2238	7.70	93.4
4	Jiamusi	214.9	46.5923	447.3445	1977	8.62	92
5	Daqing	278.1	152.8677	653.4401	1431	21.16	98.8
6	Yichun	87.3	15.2430	117.7372	684	16.52	89.7
7	Qitaihe	68.5	19.2536	112.2320	634	6.10	82.5
8	Jixi	149.4	33.6601	612.6107	1064	7.43	97.7
9	Heihe	127.7	42.5270	311.6536	1001	2.19	94.6
10	Suihua	371.7	67.4869	684.2342	2194	3.76	100
11	Shuangyashan	120.3	25.4147	141.4018	1098	6.32	93.5
12	Hegang	88.7	22.9843	41.5041	705	7.54	79.6

X7	X8	X9	X10	X11	X12	X13	X14
100	877.2	450	770.9157	5183.8	1168.6479	1196	473.0
99.1	600.4	247	93.5006	1200.4	752.7777	351	131.0

80	447.9	114	382.6405	831.7	360.4715	306	92.7
90.2	605.4	127	112.5432	811.8	701.1178	319	188.0
92.5	249.6	148	529.2844	2301.1	502.3565	490	327.2
95.5	132.8	53	8.4937	295.2	202.2127	63	121.9
91	98.7	48	21.0838	206.4	72.7146	102	67.6
92.6	361.7	96	51.5093	572.4	398.2630	208	80.4
91.9	526.4	88	30.0252	614.4	524.7133	115	27.9
97.9	437.6	250	79.0378	1150.2	1058.4713	369	92.8
92.1	163.7	78	18.2568	493.9	375.1901	152	118.0
87.2	10.3	50	22.7219	340.2	210.5500	139	85.0

2.2 Cluster Analysis Was Conducted Based on SPSS Software

Firstly, SPSS software was used for systematic cluster analysis of the data. The analysis process and results are as follows.

Table 2: Case Summary.

Effectivity		Deficiency		Aggregate	
Number of cases	Percentage	Number of cases	Percentage	Number of cases	Percentage
12	100.0	0	0.0	12	100.0

It can be seen from Table 2 that there are no lost or unparticipated samples in the clustering process of the selected twelve prefecture-level cities in Heilongjiang Province, which also indicates that the cluster analysis has carried out similar clustering on various indicators of the twelve samples, so the next step of analysis can be carried out.

Table 3: Clustering of samples.

Case	Two clusters	Three clusters
1: Harbin	1	1
2: Qiqihar	2	2
3: MudanJiang	2	3
4: Jiamusi	2	2
5: Daqing	2	2
6: Yichun	2	3
7: Qitaihe	2	3
8: Jixi	2	3
9: Heihe	2	3
10: Suihua	2	2
11: Shuangyashan	2	3
12: Hegang	2	3

According to the sample classification table in Table 3, we can see the sample clustering situation when the twelve samples are divided into two clusters and three clusters respectively. When the samples are clustered into three categories, the clustering results shown in Table 4 can be obtained.

Table 4: Result of clustering.

Region	Category
Harbin	1
Qiqihar, Jiamusi, Daqing, Suihua	2
MudanJiang, Yichun, Qitaihe, Jixi, Heihe, Shuangyashan, Hegang	3

Next, we use K-mean clustering method to analyze the samples: because there is no change or only a small change in the clustering center, convergence is realized. The maximum absolute coordinate change of any center is 0.000. The current iteration is two. The minimum distance between the initial centers is 6775.821. Table 5 can be obtained:

Table 5: Cluster member table.

The serial number	Region	Cluster	Distance
1	Harbin	1	0.000
2	Qiqihar	2	2018.111
3	MudanJiang	3	850.618
4	Jiamusi	2	1904.606
5	Daqing	2	2094.661
6	Yichun	3	789.013
7	Qitaihe	3	776.024
8	Jixi	3	982.068
9	Heihe	3	802.365
10	Suihua	2	1139.625
11	Shuangyashan	3	746.527
12	Hegang	3	580.750

According to the final K-means clustering member Table 5, we can see that twelve samples are

clustered into three categories. The first category is Harbin, the second category is Qiqihar, Jiamusi, Daqing, Suihua, and the third category is Mudanjiang, Yichun, Qitaihe, Jixi, Heihe, Shuangyashan and Hegang. It can be seen that the results obtained by the two clustering methods are consistent.

2.3 Factor Analysis Was Conducted Base on SPSS Software

Factor analysis method in SPSS software was used to process the data of fourteen indicators of urban construction in twelve prefecture-level cities of Heilongjiang Province in 2020, and the characteristic value, contribution rate and cumulative contribution rate of the principal factors were obtained. See Table 6:

Table 6: The main factor characteristic value, contribution rate and cumulative contribution rate of economic development level of twelve prefecture-level cities in Heilongjiang Province (%).

Principal factor	1	2
Value of characteristic	10.733	1.543
Contribution rate	76.667	11.024
Cumulative contribution rate	76.667	87.691

It can be seen from Table 6 that the eigenvalue of the variable correlation coefficient matrix is greater than the two main factors of one (Tang, 2007), and the cumulative contribution rate reaches 87.961%, which together explains 87.961% of the total variance of the original variable. Obviously, the information represented by the two principal factors can fully explain and provide the information expressed by the original data, and only 12.039% of the information is lost. Thus, the score of the two principal factors on each original variable is obtained Y_1,Y_2 (See Table 7),At the same time, to obtain a comprehensive index that can reflect the economic development level of prefecture-level cities($\sum Y$),Taking the contribution of two main factors as the weight, the comprehensive construction scores of twelve prefecture-level cities in Heilongjiang Province are defined as follows:

$$\sum Y=0.87429Y_1+0.12571Y_2$$

Table 7: Ranking table of comprehensive construction scores of 12 prefecture-level cities in Heilongjiang Province.

City	Y_1	Y_2	ΣY	Sort
Harbin	2.80925	-0.7965	2.36	1
Qiqihar	0.40901	1.51006	0.55	2
Suihua	0.27543	1.75749	0.46	3
Daqing	0.45881	-0.97039	0.28	4
Jiamusi	-0.03308	0.35229	0.02	5
Mudanjiang	-0.13117	-0.81468	-0.22	6
Jixi	-0.3674	0.49756	-0.26	7
Heihe	-0.44954	0.79631	-0.29	8
Shuangyashan	-0.54346	0.07777	-0.47	9
Yichun	-0.6483	-0.31156	-0.61	10
Qitaihe	-0.8901	-0.8428	-0.88	11
Hegang	-0.88946	-1.25007	-0.93	12

Generally speaking, the higher the comprehensive score, the better the regional economic development level; If the score is greater than zero, it means that the development level of this region is above the provincial average development level; otherwise, it is below the provincial average development level. Therefore, it is necessary to actively adjust the development ideas to promote the rapid and coordinated development of regional construction (Zhang, 2011).

From the comprehensive score, the economic development level of Harbin is obviously above the average level of provincial economic development ($\sum Y>0$); Qiqihar, Daqing, Suihua, Jiamusi close to the province's average level of economic development; Mudanjiang, Yichun, Qitaihe, Jixi, Heihe, Shuangyashan and Hegang are significantly lower than the average level of economic development of the whole province.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

3 CONCLUSIONS

The results of the above four regional types are basically consistent with the economic development status of the twelve prefecture-level cities in Heilongjiang Province. In the future, developed areas should use Harbin-Dalian expressway to drive the rapid development of the suburban economy, increase its economic radiation radius, and effectively promote the deep development of the second and third industries. More developed areas should take animal husbandry and agricultural and sideline products processing industry as the leading industries to develop green agriculture and take the road of "green industry." Less developed areas should vigorously develop a port and border economic cooperation, develop cross-border tourism between China and Russia, implement the strategy of sustainable development affected by tourism trade, accelerate the transformation of resource-based cities whose resources are close to exhaustion, and seek development directions according to local conditions. Backward areas should take forestry and tourism as the leading industries, optimize the land structure, improve the degree of industrialization, and develop township enterprises. At the same time, Harbin--Daqing--Qiqihar as the first-level development axis, Mudanjiang--Jiamusi--Shuangyashan as the second-level development axis, and radiate to Heihe, Yichun and other cities; We should strengthen the diffusion of economy, technology and capital from economically developed central cities to surrounding cities and promote the process of regional industrial integration.

REFERENCES

Chen Leilei. Research on K-Means Text Clustering with Different Distance Measures [J]. Software, 2015,36 (1) : 56-61.

Heilongjiang Provincial Bureau of Statistics. Heilongjiang Statistical Yearbook [Z]. Beijing: China Statistics Press.

He Xiaoqun. Multivariate Statistical Analysis [M]. Beijing: China Renmin University Press, 2015.

Ning Yue-min, TANG Li-zhi. The concept and index system of urban competitiveness [J]. Modern Urban Research, 2001, 10(3).

Tang Gongshuang. Analysis of principal component Analysis from factor analysis based on SPSS [J]. Statistical Education, 2007, (2).

Wu Rongqiang, Li Jinhong. Classification of anode Pressure drop in Aluminum electrolyzer based on Cluster analysis [J]. Journal of Software, 2018, 39(3): 166-169.

Wang Xin, Zhao Nan, et al. Using SPSS software to analyze the occurrence rules and prevention of school bullying incidents [J]. Software, 2018,39 (1) : 159-164.

Wang Xiuxiu, Zhang Wenyu, et al. Application of improved Principal component Cluster Analysis Method in Education Informatization [J]. Software, 2015,36 (7) : 10-16.

Zhai Xu, Tian Mingguang, et al. Prediction of Lead-acid Battery SOC Based on K-means Clustering and Gaussian Process Regression Integration [J]. Journal of Software, 2018, 39(1): 132-137.

Zhang Shijie. City Competitiveness Evaluation of Fuyang City based on Principal Component Analysis [J]. Journal of Anhui Radio and Television University, 2011, (4).

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS