Scientific and Social Research
2022, Volume 4, Issue 11

The Urban Competitiveness of Southwest Cities in China Based on Factor Analysis

Xia Sun[1]*, Yuan Yang[2]

[1]Department of Public and Basic Education, Yunnan Open University, Kunming 650500, China

[2]College of Communication and Information Engineering, Yunnan Open University, Kunming 650500, China

Corresponding author: Xia Sun, sunxia@ynou.edu.cn

Abstract: Factor analysis is an important way of data mining which can be performed using MATLAB, Python, and so on. We studied the urban competitiveness of Southwest cities in China using factor analysis. Two factors were extracted among 22 original variables. The factor score of cities was obtained using Python and were classified into three categories: Chongqing, Chengdu, Kunming and Guiyang are the first-tier cities; Baoshan, Pu'er, Lincang and Lijiang cities are of the lowest tier; the remaining cities belong to the second tier.

Keywords: Data mining; Factor analysis; Urban competitiveness

Online publication: November 21, 2022

1. Introduction

Factor analysis is an important way of data mining. The purpose of it is dimension reduction [1–3]. In this study, factor for urban competitiveness. Urban competitiveness refers to a city's competitive advantage compared to other cities, its ability to create wealth. It is a comprehensive reflection of a city's production capacity, quality of life, social progress and external influence in a certain period [4,5]. A city with a good competitiveness can not only improve the living standards of people in the region, but also promote the overall competitiveness of a country as well [6-8]. Therefore, the study of city competitiveness has been a hot topic for discussion.

We studied the urban competitiveness of 33 cities in southwest China using factor analysis in this paper. Among 22 original variables, we extracted two factors and obtained a comprehensive evaluation of the competitiveness about cities.

2. Factor analysis and original variables

2.1. Factor analysis

Factor analysis originated in the early 20th century, when scholars such as K. Pearson and C. Pearman made statistical analysis to define and measure intelligence [9]. Factor analysis is used to describe the covariance or correlation between the original variables with several potential and unobservable random variables (factors). It reflects most of the information of original variables with fewer independent factors, which can be expressed by mathematical models. P represents the original variables, the mean value of each variable (or after normalization) is 0, and the standard deviation is 1. Each of the original variables is expressed as a linear combination of factors as shown below.

$$\begin{cases} x_1 = a_{11} f_1 + a_{12} f_2 + a_{13} f_3 + \cdots + a_{1k} f_k + \varepsilon_1 \\ x_2 = a_{21} f_1 + a_{22} f_2 + a_{23} f_3 + \cdots + a_{2k} f_k + \varepsilon_2 \\ \quad\quad \cdots\cdots \\ x_p = a_{p1} f_1 + a_{p2} f_2 + a_{p3} f_3 + \cdots + a_{pk} f_k + \varepsilon_p \end{cases}$$

The above model is the mathematical model of factor analysis, which can also be expressed as $X = AF + \varepsilon$ in matrix form. F represents the common factor, since they appear in the linear expression of each original variables; A represents the factor loading matrix; $a_{ij}(i = 1,2, \ldots, p; j = 1,2, \ldots, k)$ represents factor loading, it is the load of the i-th original variable on the j-th factor; ε represents special factors; it means the part of the original variables that cannot be explained by common factors and its mean value is 0.

The goal of establishing the mathematical model of factor analysis is to obtain the load matrix. This can be done using SPSS, MATLAB, Python, and many more. Python is used in this paper.

2.2. Original variables

There are many factors that affect the competitiveness of cities, such as the GDP, the fiscal revenue, the educational development, the infrastructure condition, and the degree of opening. After analyzing a large number of indicators in “2020 China City Statistical Yearbook” [10], 22 original variables were selected and the following indicator system was constructed (**Table 1**).

Table 1. Urban comprehensive strength evaluation index system

Level indicators	The secondary indicators	Unit	Serial number
Economic development index	Gross regional product	100 million yuan	X_1
	GDP Per capita	100 million yuan	X_2
	Total retail sales of consumer goods	10,000 yuan	X_3
	Local general public revenue	10,000 yuan	X_4
	Local general public expenditure	10,000 yuan	X_5
	Annual savings balances of financial institutions	10,000 yuan	X_6
	Average wage	10,000 yuan	X_7
	Profits of industrial enterprises above designated size	10,000 yuan	X_8
	Ratio of tertiary industry to GDP	Percent (%)	X_9
Degree of openness	Actual utilization of foreign capital	10,000 dollars	X_{10}
	Total import and export	10,000 yuan	X_{11}
Technology and education	Number of students in institutions of higher learning	Person	X_{12}
	Number of students in secondary vocational schools	Person	X_{13}
	Books in the public library	10,000 volumes	X_{14}
	Number of patents granted	–	X_{15}

(Continued on next page)

(Continued from previous page)

Level indicators	The secondary indicators	Unit	Serial number
Infrastructure	Number of urban employees insured by basic medical insurance	Person	X_{16}
	Hospital beds	Piece	X_{17}
	Number of hospitals	–	X_{18}
	Per capita electricity consumption	Degree	X_{19}
	Green area	Hectare	X_{20}
	Drain length	Kilometers	X_{21}
	Investment in fixed assets of public facilities	10,000 yuan	X_{22}

3. Mathematical model and main results

3.1. Premise of factor analysis

One of the main tasks of factor analysis is to extract and synthesize the overlapping information in the original variables into factors, so as to finally achieve the purpose of dimension reduction. Therefore, it requires a strong correlation between the original variables. The correlation between the original variables can be determined using the following ways.

(i) Calculate the correlation coefficient matrix

Calculate the correlation coefficient matrix of the original variables and observe its correlation matrix. If most of the correlation values in the correlation coefficient matrix are less than 0.3, the variables are weakly correlated, then these variables are not suitable for factor analysis.

(ii) Calculate anti-Image correlation matrix

The reflection image matrix mainly includes negative partial covariance and negative partial correlation coefficient. If the absolute value of most other elements (except the main diagonal element) in the image correlation matrix is small, and the value of the diagonal element is close to 1, it means that these variables are strongly correlated and are suitable for factor analysis.

(iii) Bartlett's test of sphericity

Bartlett's test of sphericity takes the correlation coefficient matrix of the original variable as the starting point. As for the null hypothesis, H_0, the correlation coefficient matrix is the unit matrix. If the observed value of the statistic is large and the corresponding probability P is less than the given significance level α, the null hypothesis should be rejected and the original variables are considered suitable for factor analysis. Otherwise, the null hypothesis is accepted thus the original variables are considered not suitable for factor analysis.

(iv) Kaiser–Meyer–Olkin (KMO) test

The KMO test is used to compare the simple correlation coefficient and partial correlation coefficient between variables. The KMO value is between 0 and 1. The closer the KMO value is to 1, the stronger the correlation between the variables, and the more suitable the original variables are for factor analysis. The closer the KMO value is to 0, the weaker the correlation between variables, and the less suitable they are for factor analysis. A commonly used KMO metric will be as the following: > 0.9 means a good fit, 0.8 means suitable, 0.7 means normal, 0.6 means not suitable, < 0.5 means not suitable.

By using Python to obtain the correlation coefficient matrix and through the analysis of the correlation coefficient matrix, the variables we selected are suitable for factor analysis.

3.2. Extraction of common factors and comprehensive score

The characteristic roots from correlation coefficient matrix were obtained. And the cumulative sum of characteristic roots are as follows.

[16.96924052 18.7946781 19.90447637 20.62479234 21.21245723 21.50483717 21.74137291 21.84583375 21.92468486 21.94986647 21.96799909 21.97854387 21.98565612 21.99039744 21.99336265 21.99545513 21.99693689 21.99850156 21.99865236 21.99888999 21.99952717 22.]

The cumulative contribution rate of two factors is 0.854. There are usually two ways to determine the number of factors. One way is to select the characteristic roots with eigenvalues greater than 1. The other way is to select the number of characteristic roots with cumulative variance contribution rates greater than 0.85 as the number of factors. In this study, two factors were selected using the latter way. The factor analysis model is as follows:

$$
\begin{aligned}
X_1 &= 0.98942079f_1 - 0.12456659f_2 + 0.00552336 \\
X_2 &= 0.62563469f_1 + 0.21127558f_2 + 0.56421493 \\
X_3 &= 0.98741399f_1 - 0.12342195f_2 + 0.00977126 \\
X_4 &= 0.98923312f_1 - 0.13591814f_2 + 0.00294806 \\
X_5 &= 0.91271988f_1 - 0.37681267f_2 + 0.02499427 \\
X_6 &= 0.99374621f_1 + 0.06319377f_2 + 0.00849924 \\
X_7 &= 0.35457916f_1 + 0.06841186f_2 + 0.86963142 \\
X_8 &= 0.88227252f_1 - 0.21631166f_2 + 0.17482347 \\
X_9 &= 0.51198086f_1 + 0.27904327f_2 + 0.6603413 \\
X_{10} &= 0.96383675f_1 + 0.1889675f_2 + 0.03520754 \\
X_{11} &= 0.98565196f_1 + 0.06091338f_2 + 0.02473054 \\
X_{12} &= 0.9419085f_1 + 0.07658956f_2 + 0.1070938 \\
X_{13} &= 0.91119646f_1 - 0.34144132f_2 + 0.05313571 \\
X_{14} &= 0.93435233f_1 + 0.29142504f_2 + 0.04193414 \\
X_{15} &= 0.98963798f_1 + 0.13696161f_2 + 0.00189423 \\
X_{16} &= 0.97793352f_1 + 0.18735713f_2 + 0.00849015 \\
X_{17} &= 0.96868006f_1 - 0.14847978f_2 + 0.03961606 \\
X_{18} &= 0.83088756f_1 - 0.14221722f_2 + 0.28942364 \\
X_{19} &= 0.40645385f_1 + 0.2286361f_2 + 0.7828528 \\
X_{20} &= 0.97766424f_1 - 0.14560721f_2 + 0.02298138 \\
X_{21} &= 0.98141916f_1 - 0.15092655f_2 + 0.01403791 \\
X_{22} &= 0.83088756f_1 + 0.71916276f_2 + 0.07644893
\end{aligned}
$$

From the factor analysis model, f1 is named as economic factor because of its high load on multiple variables of economic development indicators. f2 has a high load on the level of public facilities and general public financial expenditure, thus it is named as infrastructure factor.

Furthermore, we obtained the score function of factors and calculated the score for each factor. The comprehensive score and the ranking of cities are shown in **Table 2**.

Table 2. Comprehensive score and ranking

	f1	**f2**	**Score**	**City**
1	3.57297	4.16697	3.63066	Chengdu
0	4.03169	-3.67407	3.28326	Chongqing
25	0.837016	-0.303642	0.726229	Kunming

(Continued on next page)

(Continued from previous page)

	f1	f2	Score	City
19	0.445344	0.423721	0.443244	Guiyang
6	-0.0283619	0.326737	0.00612715	Mianyang
21	-0.0163058	-0.311717	-0.0449976	Zunyi
5	-0.171641	0.21961	-0.13364	Deyang
13	-0.173891	-0.348925	-0.190891	Yibin
4	-0.2023	-0.30557	-0.21233	Luzhou
11	-0.208872	-0.362142	-0.223758	Nanchong
10	-0.259974	-0.0716109	-0.241679	Leshan
26	-0.224109	-0.409889	-0.242153	Qujing
2	-0.321599	0.255137	-0.265583	Zigong
15	-0.298853	0.0344467	-0.266465	Ya'an
27	-0.284108	-0.131855	-0.26932	Yuxi
12	-0.310757	0.0551049	-0.275222	Meishan
9	-0.344354	0.161885	-0.295186	Neijiang
23	-0.289756	-0.349647	-0.295573	Bijie
20	-0.335131	0.041507	-0.29855	Liupanshui
8	-0.331803	-0.00463521	-0.300026	Suining
3	-0.366606	0.214795	-0.310137	Panzhihua
24	-0.349972	0.000958733	-0.315887	Tongren
22	-0.363265	0.0963613	-0.318624	Anshun
14	-0.355813	-0.0114632	-0.322368	Guang'an
16	-0.366945	0.0153659	-0.329813	Dazhou
7	-0.381841	0.116953	-0.333396	Guangyuan
18	-0.405158	0.163164	-0.349959	Ziyang
29	-0.367155	-0.226113	-0.353456	Zhaotong
17	-0.408007	0.071547	-0.36143	Bazhong
28	-0.411991	-0.00839964	-0.372792	Baoshan
31	-0.419961	0.0332214	-0.375946	Pu'er
32	-0.435534	0.0105441	-0.392209	Lincang
30	-0.452976	0.111649	-0.398137	Lijiang

In terms of comprehensive scores, Chongqing, Chengdu, Kunming, Guiyang all scored more than 0, which means they belong to the top-tier. Chongqing has the highest score in terms of economic factor, which means that the economic development of Chongqing ranks first in southwest China, while Chengdu ranks second.

Except Mianyang, the comprehensive scores of the remaining 28 cities are all less than 0. Among them, Baoshan, Lincang, Pu'er and Lijiang are the minority communities in Yunnan Province, and they belong to the lowest tier of cities.

The remaining cities belong to the second-tier, in which most of the cities are located in Sichuan Province.

4. Conclusion

Factor analysis is an important way of data mining with the main purpose being dimension reduction. It is used to measure urban competitiveness in this study, which was performed by using Python. Among 22 original variables, two factors were extracted and scores were obtained for each city. The competitiveness of Southwest cities was evaluated through the scores obtained. In fact, factor analysis can be used for more original variables other than economy, such as education, psychology, and many more.

Funding

Science Research Foundation of Yunnan Education Bureau, No.2020J0492

Science Research Foundation of Yunnan Open University, No.21YNOU13

Web Front-End Development Research Center of Yunnan Open University

Disclosure statement

The author declares no conflict of interest.

References

[1] Si S, Sun X, (eds) 2020, Mathematical Experiment and Modeling in Python, Science Press, Beijing.

[2] Wu X, 2019, Multivariate Statistical Analysis with R and Python, Renmin University of China Press, Beijing.

[3] Xue W, 2007, SPSS Statistical Analysis Method and Application. Publishing House of Electronics Industry, Beijing.

[4] Ni P, 2004, China's Urban Competitiveness Report: Making Chinese Cities Win-Win, Social Science Literature Publishing House, Beijing.

[5] Sun L, Tan C, Hou S, 2007, Comprehensive Analysis of the Strength of Urban Development in Hebei Province, Journal of North China Institute of Aerospace Engineering, 17(3): 33–35.

[6] Zheng R, 2005, Factor Analysis Method for Performance Evaluation of Central City Competitiveness, University of Shanghai for Science and Technology, 27(6): 555-559.

[7] Chu L, Mu N, 2020, Research on The Competitiveness of Anhui Province Based on Factor Analysis. Economic Research Guide, 2020(13): 119–121.

[8] Sun H, Song R, The Application of Factor Analysis in City's Competitiveness, Green Economy, 29(7) 128–131.

[9] Wang X, 2017, Applied Multivariate Analysis to Statistical Analysis, Shanghai University of Finance and Economics Press, Shanghai.

[10] 2020 China City Statistical Yearbook, 2021, China Statistics Press, Beijing.