Principal Component Analysis and Cluster Analysis Application in Provincial Capital City of China Economic Research

Zhuoshi Li[1, a], Min Fu[2,b *], Yu Zhang[3,c] and Shang Gao[4,d]

[1,2,3,4] Jilin Agricultural University, Changchun, Jilin, 130118, China

[a]leezs643@sina.com, [b]lzs036@163.com,[c]499360221@qq.com, [d]leezs036@foxmail.com

Keywords: principal component analysis; cluster analysis; economic indicators; comprehensive evaluation

Abstract. In this paper, we use the principal component analysis and cluster analysis on the development of the provincial capital city of China economic research. We select seven economic indicators to related economic data of 2014 provincial capital cities around the country as the basis. The GDP, the public fiscal revenue, the public financial expenditure, the total amount of investment in fixed assets, Closing balance of savings deposits of urban and rural residents, respectively, in the staff and workers aspects of the average wage, the total retail sales of social consumer goods, the level of economic development of the country capital city comprehensive evaluation.

Introduction

The capital city of around the country the economic strength ranking by principal component analysis, the use of cluster analysis of the economic strength of the capital city of across the country are classified, the following conclusions and make relevant recommendations for the coordination of regional differences in economic development in all provinces. Principal component analysis is a multivariate statistical analysis[1]. It is a way to extract the number of variables into a few variables integrated multivariate statistical analysis method, which has many variables associated relationship, divided into several linear combinations of variance to measure the amount of information contained in the combination, most informative as the first principal component, using the same method to select multiple main ingredient[2].As cluster analysis, Different classes between samples cluster analysis is something with property classified as a place of thought, the data sample packet combining several classes or clusters, so that property in the same class between samples is substantially the same, and properties vary greatly[3].

Principal component analysis of the economic indicators

Using MATLAB 2013 programming, in view of the sensitivity of the principal component analysis of the data unit[6]. So first the raw data were normalized to eliminate the the dimensionless to make it comparable. Secondly ,we should standardized the correlation indicators matrix to obtained characteristic roots and eigenvector matrix again, and using the obtained characteristic root and characteristic vector cumulative variance contribution rate and the contribution rate, the finalization of the main ingredient, and generally variance contribution rate is greater than or equal to 85% variable for the final comprehensive evaluation.The variables are defined as follows[4].

x_1: GDP (billion yuan)

x_2: Public Revenue (billion yuan)

x_3: Public Financial Expenditure (billion yuan)

x_4: The Total Investment in Fixed Assets(billion yuan)

x_5: The Ending Balance of Savings Deposits of Urban and Rural Residents(billion yuan)

x_6: The Average Wage of Workers(yuan)

x_7: The Total Retail Sales of Social Consumer Goods(yuan)

Table 1 Standardization of Data

Provincial Capital Cities	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Beijing	2.4821	3.1282	3.0934	1.2226	3.6216	2.7852	2.8532
Tianjin	1.4953	1.3908	1.5192	2.4835	0.5008	0.8214	0.9018
Shijiazhuang	-0.3332	-0.5463	-0.4444	0.0882	-0.2042	-1.1399	-0.3467
Taiyuan	-0.8046	-0.6207	-0.6419	-0.9323	-0.3776	-0.5580	-0.6920
Hohhot	-0.7473	-0.6924	-0.6654	-0.9996	-0.7640	-0.7427	-0.7615
Shenyang	0.1082	-0.0127	-0.0971	0.9797	-0.0802	-0.4621	0.2599
Dalian	0.2029	0.0413	0.0989	1.0179	-0.1376	0.0501	-0.0697
Changchun	-0.3063	-0.4730	-0.3377	-0.2886	-0.4186	-0.4753	-0.3478
Harbin	-0.3037	-0.4505	-0.2633	0.3939	-0.3193	-0.8713	0.0311
Shanghai	2.8863	3.6206	3.4374	0.6796	3.2708	2.5894	2.6917
Nanjing	0.2723	0.0206	-0.1265	0.4563	-0.0561	0.6185	0.4326
Hangzhou	0.3361	0.1457	-0.1219	0.1197	0.2411	0.5197	0.4324
Ningbo	0.1025	-0.0217	-0.0403	-0.2215	-0.1216	0.3950	-0.0152
Hefei	-0.3699	-0.4106	-0.3398	0.2907	-0.5719	-0.3224	-0.5923
Fuzhou	-0.3688	-0.3938	-0.4338	-0.0546	-0.3799	-0.3816	0.0078
Xiamen	-0.6881	-0.3536	-0.4504	-1.0675	-0.6547	0.2694	-0.8454
Nanchang	-0.6270	-0.5717	-0.5447	-0.4349	-0.6341	-0.8915	-0.7663
Jinan	-0.2627	-0.3629	-0.4479	-0.5398	-0.3858	-0.1925	0.0386
Qingdao	0.2713	-0.0259	0.0317	0.4296	-0.2077	-0.2321	0.1603
Zhengzhou	-0.0758	-0.0977	-0.1602	0.1750	-0.1394	-1.0341	-0.0398
Wuhan	0.4723	1.0082	0.1370	0.8136	-0.0083	-0.3609	0.6250
Changsha	0.1072	-0.3030	-0.2710	0.2533	-0.3419	-0.1632	0.0679
Guangzhou	1.6972	0.3616	0.3921	0.1970	1.4976	1.2009	2.1074
Shenzhen	1.5203	1.0089	0.6874	-0.5955	0.8431	1.5167	0.8834
Nanning	-0.7294	-0.6109	-0.5457	-0.6232	-0.5790	-0.6639	-0.6073
Haikou	-1.0947	-0.7971	-0.8232	-1.3466	-0.8359	-1.0348	-1.0875
Chongqing	1.1657	0.9671	2.0164	2.8769	0.9110	-0.5736	0.9664
Chengdu	0.4833	0.0944	0.1747	1.0272	0.6109	2.0637	0.5432
Guiyang	-0.8676	-0.5879	-0.5697	-0.3807	-0.6798	-0.5976	-0.9398
Kunming	-0.6118	-0.3973	-0.3835	-0.4208	-0.3679	-0.5613	-0.4816
Lhasa	-1.2100	-0.8372	-0.8233	-1.4761	-0.2664	0.8669	-1.2604
Xi'an	-0.3292	-0.3411	-0.2439	0.4407	0.0406	-0.3060	-0.0590
Lanzhou	-0.9270	-0.7556	-0.7163	-1.0758	-0.6401	-0.6937	-0.9107
Xining	-1.0805	-0.8186	-0.7500	-1.2428	-0.8542	-1.1754	-1.1737
Yinchuan	-1.0207	-0.7445	-0.7374	-1.1508	-0.8455	-0.1359	-1.1584
Urumqi	-0.8450	-0.5608	-0.6089	-1.0942	-0.6657	-0.1269	-0.8476

Table 2   The correlation coefficient matrix of the seven economic indicators

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1.0000	0.9317	0.9214	0.6872	0.9189	0.7608	0.9725
x_2	0.9317	1.0000	0.9679	0.6001	0.9326	0.7550	0.9012
x_3	0.9214	0.9679	1.0000	0.6790	0.9289	0.7054	0.8846
x_4	0.6872	0.6001	0.6790	1.0000	0.5267	0.3055	0.6650
x_5	0.9189	0.9326	0.9289	0.5267	1.0000	0.8187	0.9333
x_6	0.7608	0.7550	0.7054	0.3055	0.8187	1.0000	0.7402
x_7	0.9725	0.9012	0.8846	0.6650	0.9333	0.7402	1.0000

As can be seen from Table 2 GDP occupy the first, which is an important link between the country's economic development.The following is the total fixed asset investment, public revenue, the correlation of total retail sales have strong relativity , the country's economic development to change the traditional social structure strong evidence.

Using the correlation coefficient matrix to calculate eigenvalues and its contribution rate.

Table 3 The Eigenvalues and its Contribution Rate

Number	Eigenvalues	Contribution Rate
1	5.7967	0.8281
2	0.7388	0.1055
3	0.2304	0.0329
4	0.1439	0.0206
5	0.0596	0.0085
6	0.0230	0.0033
7	0.0075	0.0011

Table 4 obtained cumulative contribution rate, the contribution rate can be seen in front two eigenvalues had reached just over 90%, indicating that the principal component analysis of the effect is very satisfactory. Select the following cumulative contribution rate of the previous four eigenvalues comprehensive evaluation.

Table 4 The Cumulative Contribution Rate

Number	The Cumulative Contribution Rate(%)
1	82.81
2	93.36
3	96.65
4	98.71
5	99.56
6	99.89
7	100

As can be seen in Table 3, the second eigenvalues trend has leveled off. Because we choose between the seven indicators related to economic development and dependence. After principal component analysis, the first eigenvalue with eigenvector corresponding roughly equal value of each component, so can be seen as the first principal component integrated variable this 7 economic development indicators, explaining the ratio of economic development information up 82.81%, similar to the second main component of economic development is pleased to explain the proportion of 10.55%, the third main component of economic development is pleased to explain the proportion of 3.29 %, the third main component of economic development is pleased to explain the proportion of 2.06%. The first four main components of the economic development of the information accumulated to explain the proportion of 98.71%. So we can make a comprehensive evaluation for the in front four economic development indicators to give the provinces the economic development of cities around the score and its ranking.

Cluster Analysis for the Comprehensive Evaluation[5]

Table 5 Cluster Analysis Result Classification

Category	Region	Quantity
The first category	Beijing,Shanghai	2
The Second Category	Guangzhou，Shenzhen	2
The Third Category	Tianjin，Chongqing	2
The Fourth Category	The rest of the other provincial capital cities	29

The first category is the rapid economic development of the region, the second is more rapid economic development of the region, but lower than the first class area which is a more obvious difference. The third category, although the total economy is improving, but on the whole there is still some degree of deficiency, for these areas, certain economic stimulus measures is a good choice and do work to attract foreign investment in order to bring the whole regional economic development. The fourth category overall development momentum is good, but there is still a lot of space available.

Acknowledgements

References

[1] Yuanwen Luo and Shixing Zhang,The key problems and countermeasures of medical insurance for urban workers.Guangxi Economic Management Institute Journal. Vol.1(2010), p.59-61.

[2] Zhenhai Yang,Applied Mathematical Statistics,first ed., Beijing University of Technology Press. Beijing, 2005.

[3] Zhuohua Zhou and Wennan Chen, Cluster Analysis in Investment in Securities. Chongqing University Journal. Macmillan.Vol.25(2002),p.14-16.

[4] Xiaoyan Li．Application of principal component analysis in public health indicators in the evaluation of the rural,China Health Economics. Vol.5(2007),p.41-46.

[5]Tao Jiang,Ming Gao,The main ingredient of Chongqing mountain yellow nutrient characteristics and cluster analysis,．Southwest Agricultural Sciences Journal.Vol.21(2008),p.36-39.

[6] Yanke Bao and Na Li,Matlab Mathematical Statistics and Data Processing, third ed., Northeastern University Press,Shenyang,2008.