

Cluster Analysis and Principal Component Analysis Application in Primary Health Care Institutions

Zhuoshi Li^{1, a}, Mei Wang^{2, b}, Fugui Zhu^{3, c}, Hang Chen^{4, d}

^{1,2,3} Jilin Agricultural University, Changchun, Jilin, 130018, China

⁴ Institute of Scientific and Technical Information of Jilin, Jilin, 130033, China

^aleezs643@sina.com, ^b 499360221@qq.com, ^c leezs036@foxmail.com

Keywords: cluster analysis; principal component analysis; primary health care institutions

Abstract. In this paper, we use the principal component analysis and cluster analysis on the development of a combination of medical institutions in the region to analyze. In order to analyze the number of organizations from all regions, the number of beds and so on, At first the cluster analysis method for our various regions medical and health institutions system classification. Finally using the principal component analysis to explore the various regions of the recommendations put forward some development problems of health care institutions.

Introduction

Primary health care institutions mainly refers to the smallest administrative division level medical institutions. With economic development, the level of grass-roots medical services have made great progress. Which is closely linked to its development of China's medical and health institutions. The health authorities in most areas of high prevalence, but the level of medical and health institutions as well as small parts can not achieve the requirements of people, so we want our medical institutions in various regions of the grass-roots research, select the need to increase development efforts in the region. To improve the level of primary health care institutions develop a force [1].

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, as cluster analysis, one can distinguish between densely and sparsely populated areas, the overall situation of mutual understanding and classification mode data attributes[2].

Principal component analysis is a multivariate statistical analysis. 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[3].

Cluster Analysis Applications in the Primary Medical Health Institutions

Firstly, cluster analysis[4] is used for hierarchical clustering and principal component extracted data using principal component analysis, in accordance with the cluster analysis results reflect the primary component for the development of provincial and base area municipal health care institutions for analysis as shown in table 1.

For a given data clustering software application Matlab[6], first input data to calculate the distance between samples using Euclidean distance method, then the distance between the shortest distance defined class. in accordance with hierarchical clustering, the distance between the class with the shortest distance method, the results of the most long-distance method, the average distance obtained by the same method as shown in table 2.

Table 1 Medical and health organization data of various provinces and cities (unit: ten thousand)

Provinces and Cities	Number of institution	Number of bed	Number of staff	number of treatment	Number of Inpatient
1. Beijing	0.8837	0.4745	5.7084	5902.4	3.6
2. Tianjin	0.4095	0.7023	2.2438	3573.9	12.1
3. Hebei	7.7177	6.9143	19.0636	26630.9	167.7
4. Shanxi	3.8443	3.9956	10.6332	7431.4	62
5. Neimenggu	2.2009	2.4543	6.7255	5524.8	38.7
6. Liaoning	3.4249	3.6464	10.0016	9361.8	73
7. Jilin	1.8804	2.1577	6.6326	5295.6	33.9
8. Heilongjiang	1.9470	2.8422	8.1287	5882.4	62.2
9. Shanghai	0.4379	1.7389	4.7289	9259.8	10.5
10. Jiangsu	2.8888	6.9516	19.2076	24596.8	168.3
11. Zhejiang	2.8939	2.6852	12.7410	22478.7	28.8
12. Anhui	2.1812	5.7826	13.5525	15529.4	179
13. Fujian	2.6374	2.9910	9.3236	10051.9	124.9
14. Jiangxi	3.8369	4.9312	11.4113	13091.9	245.3
15. Shandong	6.6462	12.5877	32.5600	41339.1	345.7
16. Henan	6.7252	10.1176	28.0477	35193.7	331.1
17. Hubei	3.4063	6.8293	15.6935	20008.4	238.1
18. Hunan	5.7177	8.6069	16.4844	14862.2	348.4
19. Guangzhou	4.4585	5.9457	21.4410	35910.7	196.1
20. Guangxi	3.3257	5.0264	12.5210	14490.5	247.3
21. Hainan	0.4839	0.5885	1.8920	2285.3	10
22. Chongqing	1.7310	4.1107	7.8560	8352	177.2
23. Sichuan	7.4215	12.2678	23.5351	28478.3	520
24. Guizhou	2.6264	3.7815	7.8283	7788	220.3
25. Yunnan	2.1887	4.4427	8.6135	12300.3	144.9
26. Tibet	0.6412	0.2583	1.3832	677.3	2.9
27. Shanxi	4.5889	3.3677	10.5034	9623.1	80.9
28. Gansu	2.5631	3.1891	6.9915	8349.4	68.4
29. Qinghai	0.5658	0.5103	1.6162	1199.2	18.7
30. Ningxia	0.3904	0.2832	1.2405	1566.1	5.9

Table 2 The expression data of principal component coefficients

Classification	Provinces and Cities
First Class	Shandong, Henan, Sichuan
Second Class	Hebei, Hunan, Guangzhou
Third Class	Shanxi, Neimenggu, Liaoning, Jilin, Heilongjiang, Anhui, Fujian, Jiangxi, Hubei, Guangxi, Chongqing, Guizhou, Yunnan, Shaanxi, Gansu, Jiangsu, Zhejiang, Xinjiang, Shanxi
Fourth Class	Shanghai, Beijing, Tianjin, Ningxia, Shanghai, Tibet

Principal component analysis in a healthcare analysis application

Principal component analysis of the data, obtained characteristic value, contribution rate and cumulative contribution rate [5] as shown in table 3.

Table 3 Characteristic value and contribution rate

Characteristic	Contribution rate	Cumulative contribution rate
4.4263	88.5261	88.5261
0.3331	6.6618	95.1879
0.1890	3.7800	98.9679
0.0416	0.8321	99.0800
0.0100	0.0200	100.1000

In the table, the contribution rate of the first principal component reached 88.5261% cumulative contribution rate of the first two principal components exceeds 95%, so I chose the first two principal

components analysis on it. The first two the primary components of the coefficient data variables shown in table 4.

Table 4 The expression data of principal component coefficients

the standard variation	principal component 1	principal component 2
institution	0.4388	-0.6003
bed	0.4661	0.2023
staff	0.4640	-0.3206
treatment	0.4385	0.0329
inpatient	0.4274	0.7031

According to table 4 to write the first two principal components of the expression:

$$y_1 = 0.4388x_1^* + 0.4661x_2^* + 0.4640x_3^* + 0.4385x_4^* + 0.4274x_5^* \quad (1)$$

$$y_2 = -0.6003x_1^* + 0.2023x_2^* - 0.3206x_3^* - 0.0329x_4^* + 0.7031x_5^* \quad (2)$$

The first principal component y_1 from the expression, it has a similar loads, a description of each standard y_1 on the importance of variation are similar, can be considered the first principal component reflects the primary health care sector in each respective standard variable overall levels.

The second the primary component y_2 from the expression, it has higher levels of negative load variation on the standard x_3^* , x_1^* has a higher positive load. Description second component reflects the contrast of two aspects, one clinic visits reflects, on the one hand reflects hospitalized people.

Summary

Difference of provinces grassroots medical institutions in the region, can be roughly divided into four different types. According to the results of the main components, the first principal component reflects the consolidation of staffing levels, the first two categories, Shandong, Henan, Sichuan, Hebei, Hunan, Guangzhou is the most populated province, level of economic development is also in the upper reaches, appropriate medical organization, staffing, clinic visits on at a high level. More than a dozen provinces in the third category of the population size in the middle level, the scale of the development of medical and health institutions is relatively small compared to some of the first class. The fourth category is more complicated, both big cities such as Beijing, Shanghai, there are Qinghai, Tibet, so the relative economic conditions are not good province. Beijing, Shanghai is divided into the fourth category is actually very easy to understand, and their grass-roots population ratio compared to other provinces is relatively small, demand so people are generally accepted by non-primary care, so was in the fourth class. Like Hunan, Hubei, Anhui, Guangxi, so more of the population of the provinces are divided in the fourth category a bit unreasonable, their economic situation is relatively less prominent, relatively scarce health care institutions, will appear this situation.

Acknowledgements

The Youth Foundation of Jilin Agricultural University (201328)
 Jilin Province Innovation and Entrepreneurship Students Training Program Project
 Jilin Agricultural University Undergraduate Science and Technology Innovation Fund Project

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

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