



Analysis of the Development Level of Chinese Public Libraries

Based on Factor Analysis and Cluster Analysis

Yiwei Sun*

South China Normal University, SCNU, Guangzhou, China

*Jessica_sun0212@163.com

Abstract. Public libraries are an important part of the construction of national public facilities, and to a certain extent reflect the level of economic and cultural development of the region. Therefore, a well-established analysis of the development level of Chinese public libraries has an important impact on the future improvement and development of Chinese public libraries. This paper uses factor and cluster analysis to analyze the index related to the development level of Chinese public libraries in 2021 and a comprehensive scoring system for the development level of Chinese public libraries is established. At the same time, dividing the development level of public libraries in 31 provinces of China into 4 categories. Finally, Base on the scoring system and the classification obtained, the paper analysis the development status and the defects of those 4 categories and puts forward policy suggestions for the development of public libraries in China.

Keywords: factor analysis, cluster analysis, Chinese public library development level

1 Introduction

Public libraries are a very important part of the national public cultural system and the public information resource system, and are closely related to local economic and cultural development. Also, public libraries play a role in alleviating social conflicts, narrowing social gaps, maintaining information equity, and improving educational standards [1]. So it's important to analysis the development level of Chinese public libraries and have a clear understanding of the current situation and Insufficient defects.

Scholars from internal and external have analyzed and studied the development level of public libraries around the world. There are already systematic research about public libraries in European and American countries. Orr proposed in 1973 that the evaluation of the quality of a library should focus on the evaluation of the service quality provided by the library rather than the evaluation of its hardware facilities [2]. F.W.Lancaster proposed in 1977 that performance indicators should be combined when evaluating the development level of a library [3].

In recent years, the empirical analysis and research on the development level of public libraries in China has gradually increased. Xia Guodong (1985) made a theoretical analysis of the situation of the Heilongjiang public library at the time and made a development plan [4]. Tan Xiangji(1999) pointed out the problems of insufficient funding, outdated equipment, unstable teams and institutional planning in China's public libraries. He believes that index, service and management are the main problems affecting the current development of public libraries [5]. Wang Ganhua and Huang Jian (1988) chose the service population as the evaluation index, and used the regression method to analyze the development level of public libraries in China [6]. Li Jianxia (2011) conducted a factor analysis and made an overview on the sustainable development capacity of Chinese public libraries, and obtained the conclusion that the overall sustainable development capacity of 31 regional public libraries China is not strong. She also conducted that the main factors that are influencing Chinese library development are the investment of funds, the level of informatization, and the growth of literature resources [7]. Zhang Qiang and Huang Yusen (2019) conducted a comprehensive performance evaluation of public libraries in China through factor analysis. Through the analysis of 8 indicators, it is concluded that the city's public library with better economic development has higher performance level [8]. Cheng Huiping and others(2015) constructed the evaluation index system from four dimensions, including the basic conditions of public libraries, and introduced the PLS structural equation model to explore the internal structure and correlation of the evaluation index system. [9]

By analyzing the scientific research achievements of scholars from around the world, this paper believes that factor analysis is a suitable model for constructing an evaluation system for the development level of Chinese public libraries. Factor analysis is a statistical technique that studies the extraction of common factors from variable groups. It can deeply excavate the relationship between influencing indicators and use fewer factors to describe the correlation between variables. At the same time, it avoids the problems of collinearity and the heteroscedasticity that may occur in regression analysis. There are few studies on the classification of the development level of Chinese public libraries, which makes it difficult for government to come out with targeted development suggestions for multiple regions at the same time. Therefore, this paper believes that it is necessary to conduct a cluster analysis on the development level of Chinese public libraries in order to have a hierarchical understanding of the development levels of public libraries in various places in China. In this way, an evaluation system for the development level of public libraries across China can be better constructed, and policy opinions that are different in different regions and suitable for local areas can be put forward.

2 Research methods and theoretical analysis

This chapter is divided into two sections, which mainly introduce the methodological principles used in analyzing the data. The first section mainly introduces the basic idea, principle and analysis steps of factor analysis. The second session mainly introduces the basic idea, principle and analysis steps of cluster analysis method.

2.1 Factor analysis

The basic idea of factor analysis is to find out a few random variables that can control all variables (samples) to describe the relationship between multiple variables (samples) through the study of the internal structure of the correlation coefficient matrix (similar coefficient matrix) of variables (samples).. This paper uses the “orthogonal factor analysis model” to analyze the data.

Mathematical model of R - type orthogonal factor analysis :

$$\begin{cases} X_1 = a_{11}F_{11} + a_{12}F_{12} + \dots + a_{1m}F_{1m} + \varepsilon_1 \\ X_2 = a_{21}F_{21} + a_{22}F_{22} + \dots + a_{2m}F_{2m} + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_{p1} + a_{p2}F_{p2} + \dots + a_{pm}F_{pm} + \varepsilon_p \end{cases}$$

Representation by matrix:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pm} \end{pmatrix} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Briefly represented as:

$$X = AF + \varepsilon$$

2.2 Cluster analysis

Cluster analysis is a multivariate statistical method for studying classification problems. Clustering is divided into R-type clustering (classification of variables) and Q-type clustering (classification of samples) according to the classification objects. In this paper, the Q-type clustering method is used. The basic idea is to divide the samples into several categories according to the distance between samples. In this paper, the “shortest distance method” is used for clustering.

The shortest distance clustering model defines the distance between type G_p and type G_q as the distance between two nearest samples in two type, which is:

$$D_{pd} = \min_{x(i) \in G_p, x(j) \in G_q} d_{ij}$$

Combined G_p and G_q into a new type named G_r , then the distance between type G_k and type G_r is :

$$\begin{aligned} D_{kr} &= \min_{x(i) \in G_k, x(j) \in G_r} d_{ij} \\ &= \min \min_{x(i) \in G_k, x(j) \in G_p} d_{ij}, \min_{x(i) \in G_k, x(j) \in G_q} d_{ij} \\ &= \min\{G_{kp}, G_{kq}\} \end{aligned}$$

By analogy, the distances between each two categories can be obtained.

3 Empirical analysis

This chapter is the main body of the full paper. It is divided into two sections and mainly records the process of modeling. The first section is the factor analysis and the overall merit of the development level of Chinese public libraries. The second section is a cluster analysis of the development level of public libraries in 31 different provinces and cities in China.

3.1 Factor Analysis and Comprehensive Evaluation of the Development Level of Chinese Public Libraries

Through literature review, in order to study the development level of public libraries in China, this paper believes that it should be analyzed from four dimensions: literature resource support capability, circulation service support capability, infrastructure support capability and informatization support capability. Based on these four dimensions, this paper selects 10 indexes from 31 provinces, municipalities and autonomous regions across the China in 2021 to conduct factor analysis on the development level of public libraries in China and build the comprehensive evaluation system.

The ten indexes are: X_1 -The full number of collection of books of the city, X_2 - Number of collections in public libraries per capita of the city, X_3 - Actual number of licensed readers of the city, X_4 - total transient man-time of public libraries of the city, X_5 - Number of people who borrowed books in public libraries of the city, X_6 - Times and number of books and periodicals being borrowed, X_7 - Number of seats in the reading room, X_8 - Public library building area per 10,000 people owned, X_9 - Times of public libraries organizing various lectures and X_{10} - Number of people attending lectures in public libraries.

To verify whether the data are suitable for factor analysis or not, KMO and Bartlett spherical test are performed on the data before starting the factor analysis. The result is as follows:

Table 1. KMO and Bartlett spherical test result[Owner-draw]

Kaiser-Meyer-Olkin		.690
Bartlett's sphericity test	Pseudo Chi-square	439.180
	degree of freedom(df)	45
	significance	.000

From the result of KMO and Bartlett spherical test, it can infer that Kaiser-Meyer-Olkin is greater than 0.5 and the significance is smaller than 0.05($P < 0.05$). This explain that the data are suitable for factor analysis.

We first calculate the eigenvalue of the index's correlation matrices and its accumulative contribution rate. Meanwhile, we calculate the unit eigenvector and the variance contribution rate of indexes. Select the first two factors as the initial factors according to the variance contribution rate for their accumulative contribution rate have reached 0.8171 which is higher than 0.8. This means these two factors can describe more than 80% information of the original index data. However, the absolute value of the elements

in the factor loading matrix differentiated towards 0 and 1 in a non-significant way, so try factor rotation to solve this problem.

The accumulative contribution rate of the first two factors after rotation is 0.87043. So selecting the first two factors can better represent the information contained in the original index data and reflect most information of the original data. Meanwhile, the absolute value of the elements in the factor loading matrix differentiated towards 0 and 1 in a significant way. So take the rotation factor as the final common factor. Therefore, based on the rotation factor, this paper obtains the final factor scoring system. Shown in below:

Table 2. Standardized scoring coefficient[Owner-draw]

	Factor1	Factor2
X_1	0.14103	0.11113
X_2	0.00071	0.46503
X_3	0.12820	0.13364
X_4	0.15011	0.04515
X_5	0.14897	0.00195
X_6	0.15279	0.03270
X_7	0.14637	-0.12607
X_8	-0.01432	0.46066
X_9	0.15093	-0.10293
X_{10}	0.10273	-0.21931

From table 2 we come out with factor-scoring function(where Y_i is the variable X_i after normalization):

$$F1 = 0.14103Y_1 + 0.00071Y_2 + 0.1282Y_3 + 0.15011Y_4 + 0.14897Y_5 + 0.15279Y_6 + 0.14637Y_7 - 0.01432Y_8 + 0.15093Y_9 + 0.10273Y_{10}$$

$$F2 = 0.11113Y_1 + 0.46503Y_2 + 0.13364Y_3 + 0.04515Y_4 + 0.00195Y_5 + 0.0327Y_6 - 0.12607Y_7 + 0.46066Y_8 - 0.10293Y_9 - 0.21931Y_{10}$$

Factor1(F1) is greatly positively related to X_1 -The full number of collection of books of the city, X_3 - Actual number of licensed readers of the city, X_4 - total transient man-time of public libraries of the city, X_5 - Number of people who borrowed books in public libraries of the city, X_6 - Times and number of books and periodicals being borrowed, X_7 - Number of seats in the reading room, X_9 - Times of public libraries organizing various lectures and X_{10} - Number of people attending lectures in public libraries. As a result, factor1 was named as the “service level factor”. Factor2(F2) is greatly positively related to X_2 - Number of collections in public libraries per capita of the city and X_8 - Public library building area per 10,000 people owned. As a result, factor2 was named as the “scale level factor”.

According to the variance contribution rate after rotation, we construct a comprehensive factor score function, shown as below:

$$F_{comprehensive} = (6.299F1 + 1.872F2)/10$$

In this way, the factor scores of the 31 region in china can be calculated. The scores are shown below:

Table 3. factor scores of the 31 region in China[Owner-draw]

Provinces and places	service level factor	scale level factor	comprehensive factor score	Rank
Jiangsu	3.105296284	1.477344357	2.234124304	1
Zhejiang	2.017136628	1.239401164	1.503313631	2
Guangdong	2.492075401	-0.621065297	1.456645276	3
Shandong	1.241608123	-0.533444763	0.684035156	4
Shanghai	-0.184417855	2.89424555	0.422385642	5
Hunan	0.715751808	-1.397152781	0.191501797	6
Henan	0.815667427	-1.785059285	0.182333644	7
Fujian	0.066481553	0.43351284	0.122637292	8
Anhui	0.233869272	-0.62146676	0.031868301	9
Liaoning	-0.074113516	0.30962195	0.010874813	10
Tianjin	-0.692352534	2.336773727	-0.001838152	11
Sichuan	0.21231547	-0.786995945	-0.012541595	12
Hubei	-0.022393726	-0.177201898	-0.047112563	13
Jiangxi	-0.127545423	-0.373177691	-0.149931703	14
Shanxi	-0.289891991	-0.037393263	-0.189853239	15
Beijing	-0.572762063	0.565443511	-0.256103931	16
Chongqing	-0.322472063	-0.32427881	-0.263808882	17
Hebei	-0.163862466	-0.90145802	-0.271178226	18
Inner Mongolia	-0.556413025	0.426066951	-0.271732875	19
Guangxi	-0.328650144	-0.581197228	-0.315529428	20
Yunnan	-0.276753984	-0.855324534	-0.333814197	21
Gansu	-0.553091308	-0.035500048	-0.355553285	22
Jilin	-0.634575444	0.064482484	-0.388350878	23
Shaanxi	-0.450915186	-0.583585082	-0.393110918	24
Ningxia	-0.871597261	0.819726379	-0.397306844	25
Heilongjiang	-0.638746562	-0.281794253	-0.455438388	26
Xinjiang	-0.689544661	-0.215006052	-0.475054953	27
Qinghai	-0.971595894	0.669875609	-0.488289204	28
Guizhou	-0.536153321	-0.912002118	-0.508019204	29
Tibet	-1.037519663	0.238797003	-0.610121481	30
Hainan	-0.904833877	-0.452187696	-0.655029911	31

From the result of factor analysis and the comprehensive factor score table(table3), The paper comes out with the following conclusion:

The comprehensive score of Jiangsu Province is 2.234, ranking first, and the two factors are ranked first and third respectively, so the comprehensive score of Jiangsu Province is much higher than that of other provinces. It shows that the development level of public libraries in Jiangsu Province is at the forefront of the China in these aspects. The development of both dimensions is relatively good, and the overall development is balanced. Zhejiang Province ranks second and Guangdong Province ranks

third, with little difference in the comprehensive factor scores between the two provinces. The service level factor score of Guangdong Province is 2.49, ranking second in the country. But its scale level factor score is only -0.621, ranking 25th. Therefore, it can be seen that the Guangdong Provincial Public Library has developed relatively well in terms of book collection, reader card issuance, and the number of lectures organized. However, due to the large population, the per capita public library collection, and the per capita construction area of public libraries is relatively backward. It can be seen from this that the public libraries of Guangdong Province have not been developed in a balanced way.

Shandong, Shanghai, Hunan, Henan, Fujian and Anhui ranked in the top ten in terms of comprehensive scores. Among them, Shanghai's scale level factor score is 2.894, ranking first in the country, indicating that Shanghai's per capita book collection and library size are in good condition. Shandong's scale level factor score ranks only 22nd in the country, indicating that its public library development is unbalanced.

Xinjiang, Qinghai, Tibet, Yunnan, Gansu and other provinces, cities and autonomous regions, because of the vast area and sparse population, while the service level lags behind, the per capita collections level ranks high in the country. The economies of these regions are less developed, and the development level of public libraries is also relatively low, so this paper believes that the economic development level of the region will affect the public library development level.

Through the above factor analysis, this paper makes a primary ranking of the development level of public libraries in 31 provinces in China. This article will further classify and stratify the development level of 31 province's public libraries, and conduct further analysis.

3.2 Cluster Analysis on the Development Level of Chinese Public Libraries

In this chapter, this paper will classify the development level of public libraries in the 31 provinces mentioned in the above article through Q-type cluster analysis. In this paper, the "longest distance method" is used to calculate the distance between categories. The clustering results are as follows:

The samples are divided into the following four categories: Category 1: Jiangsu, Category 2: Shanghai, Category 3: Guangdong, Zhejiang, Category 4: Shandong, Hunan, Henan, Fujian, Anhui, Liaoning, Tianjin, Sichuan, Hubei, Jiangxi, Beijing, Chongqing, Hebei, Inner Mongolia, Guangxi, Yunnan, Gansu, Jilin, Shaanxi, Ningxia, Heilongjiang, Xinjiang, Qinghai, Guizhou, Tibet, Hainan.

The first category is the category with the highest comprehensive development level of public libraries in China. Combined with factor analysis, it can be seen that Jiangsu's comprehensive score of its comprehensive scoring system far exceeds that of all other cities in China, and the development of each dimension is balanced.

The second category of province is the category with the best scale development level. Although the overall development level of Shanghai Public Library is relatively good, it is not as good as Guangdong and Zhejiang, but its scale level ranks first in the country, and it has the best development in terms of per capita collections and library building area.

The third category is the category with better comprehensive development level of public libraries: Guangdong and Zhejiang, whose comprehensive public library development level ranks among the top three in the country, but there is still a certain gap with Jiangsu. And the service level of public libraries in these two cities is higher than the scale level. The two cities have developed rapidly in economy, and the government management team is young, and they are relatively advanced in the offering of public cultural facilities.

The fourth category of provinces with the relatively backward development level of public libraries. Their development levels in all dimensions are relatively lacking. Most of the cities are also relatively backward in economy, resulting in a lack of public resources.

4 Conclusion

From the results of factor analysis and cluster analysis, it can be seen that the two-dimensional development level of Guangdong Province is extremely unbalanced, and the per capita collection and library area are relatively backward. Therefore, Guangdong Province should expand the scale of public libraries. The provincial government should allocate funds to build public libraries in economically backward cities to improve public cultural resources in these areas.

For the first type of area, Jiangsu, they should maintain the current level of development of public libraries and continue to develop and seek innovative forms to maintain the vitality of the development of public libraries.

For areas where the development level of public libraries is poor, the economic construction of the area should be strengthened first. The government should first develop the local economy. At the same time, the government or organizations should organize should take the lead in paying attention to the construction and development of the library in these places.

The country should attach importance to the construction of public libraries in various regions. The state should appropriately support the construction of public libraries in remote areas, such as Tibet, Xinjiang and Qinghai, to help them improve the overall development level of public library.

References

1. Wu Weici. The Role of Public Libraries in Constructing Harmonious Society[J]. *Library*, 2006(01):1-2+10.
2. Orr, R. M. (1973). Measuring the goodness of library services:A general framework for considering quantitative measures. *Journal of Documentation* 29(3), 315-332.
3. Baker, S. L. & Lancaster, F. W. (1991). *The measurement and evaluation of library services* (2nd ed.). Arlington, VA: Information Resources.
4. Wang Xuelin, Yin Fanmin, Xia Guodong. Strategic Plan for Modernization Development of Public Libraries in Heilongjiang Province[J]. *Library Development*, 1958(S2), 25-32.

5. Tan Xiangjin. Four Issues Affecting the Current Development of Chinese Libraries[J]. Journal of Library Science In China, 1999(03):32-37
6. Wang Ganhua, Huang Jian. Prediction and Analysis of the Development Level of Public Library in China[J]. Library World, 1988(03), 1-4.
7. Li Jianxia. Library Service Quality Evaluation and Cases Study[J]. Document, Information & Knowledge, 2011(05), 12-13.
8. Zhang Qiang, Huang Yusen, Wang Bo. Comprehensive Evaluation of Public Library Performance in China Based on Factor Analysis[J]. Budget Management & Accounting, 2019(03), 30-35.
9. Cheng Huiping, Wan Li, Zhang Yi. A study on the development level of regional public libraries in China based on partial least squares structural equation [J]. Library And Information Service, 2015, 59(12): 19-23. DOI: 10.13266/j.issn.0252-3116.2015.12.003.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

