Analysis of Higher Education Health Index -- Based on Principal Component Analysis Model

Jiabin Wang [a], Yueli Tang [b], Jiancheng Feng, Xiaobing Yang, Xiuyan Zheng, Shangyu Meng

College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, China

[a] 1176425275@qq.com, [b] 15273121007@163.com

Abstract. To evaluate the health index of education systems in different countries, the is paper establishes a higher education system health index evaluation model through principal component analysis. Five relevant indicators of six countries were selected for principal component analysis. Through annual relevant data of universities in six different countries in the past ten years, we obtained three indicators that mainly affect the health index of higher education, Cumulative contribution rate reached 93.67%, namely, the proportion of education expenditure, the proportion of international students and the rate of higher education enrollment. Then based on the contribution rate of the three main indexes, we constructed a principal component comprehensive evaluation health index model and obtained the health index of the higher education system in six countries. Among them, India and Brazil have the lowest health index scores, respectively 0.4106 and 0.4183. Therefore, the higher education systems of India and Brazil have a lot of room for improvement. Based on the empirical results, we put forward targeted improvement plans and systems.

Keywords: Principal Component Analysis; Higher Education Health Index; Cumulative Contribution Rate; Improvement Program.

1. Introduction

A high-quality education system is an education system for all people, taking into account the principles of fairness and quality, and giving full play to everyone's potential and talents [1]. Therefore, it is important for a country to have a healthy and sustainable higher education system strategic significance. For any country in the world, how to evaluate the health index of the higher education system, so as to reflect on which policies are effective and which can be better based on the strengths and weaknesses of each country's higher education system. The higher education system is an important part of a country's further education of citizens beyond the prescribed primary and secondary education. It is not only the value of an industry itself, but also a source of trained and educated citizens for the national economy [2]. Therefore, the establishment of a country's higher education system health index evaluation can help the country implement effective policies based on the evaluation situation.

The establishment of an evaluation model for evaluating the health index of the higher education system directly affects the health index of the higher education system. Therefore, this paper studies the method of establishing the evaluation model. To evaluate the health index of a country's higher education system more accurately.

The rest of the paper is organized as follows. In the second part, the methodology of the project is introduced in detail, including the construction of a health index model based on principal component analysis. The third part describes the principal component comprehensive evaluation health index model, based on the principal component comprehensive evaluation health index, and analyzes the evaluation results. Finally, the fourth part proposes corresponding policies for the corresponding countries.

2. Establishment and Solution of Health Index Model

2.1 Introduction to Principal Component Analysis

Principal Component Analysis[3] (PCA) is known as one of the most valuable results of applying linear algebra, and it is an unsupervised learning algorithm. It is a simple, non-parametric method of extracting relevant information from chaotic data sets. Its goal is to map high-dimensional data to a low-dimensional space through some kind of linear projection, and expects that the variance of the data in the projected dimension is the largest, so as to use fewer data dimensions in order to minimize the projection The square of the error, while retaining the characteristics of more original data points.

2.2 Data Preprocessing

We choose the indicators of the proportion of international students, the economic expenditure of education exhibition, the enrollment rate of higher education, the quantification of poverty and education, and the health index of education expenditure [4]. We found some data on developed and developing countries on the international website [5], that is, the corresponding data for the six countries of Australia, Japan, France, India, Brazil, and Mexico in the past ten years. The average value of an indicator in the past ten years is shown in Table 1 below.

Table 1. Statistics date

	Proportion of international students	The proportion of education in economic expenditure	Enrollment rate of Higher Education	Poverty and education quantification	Education funds
Australia	1.77211467	13.76196054	73.76213582	20451133.33	1.14463501
Japan	1.354943347	9.682112694	54.07881393	128109200	0.618772002
France	2.475995302	10.40405407	55.36039429	61014400	1.158034013
India	0.945940712	13.25490475	14.22917713	1228597989	0.912611991
Brazil	0.49438733	12.86846304	30.59704892	195168106.9	0.859071539
Mexico	0.885384373	19.52763995	23.5390469	113977526.3	0.873612493

Standardize various indicators. In practical applications, the dimensions of the indicators are often different. Therefore, the influence of the dimensions should be eliminated before the calculation of the principal components, and the original data should be standardized, that is, the following data transformations:

$$\tilde{a}_{ij} = \frac{a_{ij} - u_j}{s_j}, (i = 1, 2, \cdots, n; j = 1, 2, \cdots, m) \quad (1)$$

Where

$$u_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij}, s_j = \frac{1}{n-1}\sum_{i=1}^{n}(a_{ij} - u_j)^2, (j = 1, 2, \cdots, m) \quad (2)$$

2.3 Calculation of Correlation Coefficient

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{13} \\ r_{21} & r_{22} & \cdots & r_{23} \\ \cdots & \cdots & & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{np} \end{bmatrix}$$

Where $r_{ij}(i=1,2,\cdots,n;j=1,2,\cdots,p)$ is the correlation coefficient between the original variables x_i and y_j, and its calculation formula is

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)^2\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)^2}} \tag{3}$$

Because R is a real symmetric matrix ($r_{ij}=r_{ji}$), we only need to calculate the upper triangular elements or the lower triangular elements. The correlation coefficient matrix obtained by MATLAB is shown in Table 2.

Table 2. Correlation coefficient matrix

Variable	X1	X2	X3	X4	X5
X1	1.0000	-0.4500	0.7160	-0.3558	0.6098
X2	-0.4500	1.0000	-0.4467	0.0034	0.0696
X3	0.7160	-0.4467	1.0000	-0.6720	0.3686
X4	-0.3558	0.0034	-0.6720	1.0000	-0.1271
X5	0.6098	0.0696	0.3686	-0.1271	1.0000

2.4 Calculation of Eigenvalue and Eigenvector

When solving the characteristic equation $|\lambda I-R|=0$, the Jacobi method is used to get the eigenvalue $\lambda_i(1,2,\cdots,p)$, which is sorted in order of size ($\lambda_1\geq\lambda_2\geq\cdots\lambda_p\geq 0$), and then the eigenvectors $e_i(i=1,2,\cdots,p)$ corresponding to the eigenvalue λ_1 are obtained respectively. Here we find $|e_i||=1$, that is $\sum_{j=1}^{p}e_{ij}^2=1$, where e_{ij} is the j component of the vector e_i.

2.5 Calculation of Principal Component Contribution Rate and Cumulative Contribution Rate

The contribution rate of principal component z_i is

$$\frac{\lambda_i}{\sum_{k=1}^{p}\lambda_i}(i=1,2,\cdots,p) \tag{4}$$

The cumulative contribution rate is

$$\frac{\sum_{k=1}^{i}\lambda_i}{\sum_{k=1}^{p}\lambda_i}(i=1,2,\cdots,p) \tag{5}$$

Generally, the principal component corresponding to the eigenvalue $\lambda_i(i=1,2,\cdots,m)$ whose cumulative contribution rate reaches 85% - 95% is selected.

F. Solution of principal component

The calculated eigenvalues, principal component contribution rate and cumulative contribution rate are shown in Table 3.

Table 3. Results of principal component analysis

Principal component	Eigenvalue	Contribution rate	Cumulative contribution rate
1	2.61735	0.5235	0.5235
2	1.09243	0.2185	0.7420
3	0.973686	0.1947	0.9367
4	0.180132	0.0360	0.9727
5	0.136409	0.0273	1

According to the above table, it can be seen that the cumulative contribution rate of the first three characteristic roots has reached more than 90%, and the effect of principal component analysis is very good. The following will select the first three principal components for analysis. The feature vectors corresponding to the first three feature roots are shown in Table 4 below.

Table 4. The eigenvectors corresponding to the first 3 principal components of standardized variables

	X1	X2	X3	X4	X5
First eigenvector	-0.0829	0.7734	0.2944	0.0113	0.5552
Second eigenvector	-0.4167	0.2455	-0.2315	0.7899	-0.2975
Third eigenvector	-0.6784	-0.3954	-0.2248	-0.0861	0.5705

The three principal components are:

$$y_1 = -0.0829x_1 + 0.7524x_2 + 0.2944x_3 + 0.0113x_4 + 0.5552x_5$$
$$y_2 = -0.4167x_1 + 0.2455x_2 - 0.2315x_3 + 0.7899x_4 - 0.2975x_5$$
$$y_3 = -0.6784x_1 - 0.3954x_2 - 0.2248x_3 - 0.0861x_4 + 0.5705x_5$$

It can be seen from the coefficients of the principal components that the first principal component mainly reflects poverty and education quantification, the second principal component mainly reflects the proportion of international students and education funds, and the third principal component mainly reflects poverty and education quantification and education funds. Substituting the standardized data of the original five indicators of each country into the expression of the three principal components, the values of the three principal components of each country can be obtained.

3. Model Results and Analysis

The contribution rates of the three principal components are respectively used as weights to construct a principal component comprehensive evaluation health index model:

$$Z = 0.5235y_1 + 0.2185y_2 + 0.1947y_3$$

Substituting the three principal component values of each country into the above formula, the health index value of each country's higher education system and the ranking result can be obtained as shown in Table 5.

Table 5. Ranking and health index value results

Area	Australia	France	Japan	Mexico	Brazil	India
Ranking	1	2	3	4	5	6
Health index value	0.8135	0.8724	0.5316	0.4787	0.4183	0.4106

According to the above comprehensive ranking and health index value, it can be seen that the health index of the higher education system in India and Brazil is low, that is, there is more room for improvement in the higher education system in India and Brazil, so we choose to analyze the higher education system in India and Brazil.

Looking to the future, the higher education systems of India and Brazil should follow the path of sustainable development on the basis of ensuring their scale. First, the proportion of education expenditure must always be used as the basis for the health index of higher education, because this is the competition between countries in the era of knowledge economy. The second is to increase the enrollment rate of higher education. The enrollment rate of higher education in India and Brazil is far lower than that of developed countries Australia, Japan, and France. The third is to increase the proportion of international students. At present, the proportion of international students in India and Brazil is lower than that in developed countries, France. This is due to the competition between countries for innovation capabilities and high-quality human resources.

4. Recommendations for the Higher Education System

The higher education health system index evaluation model established in this paper can be used to evaluate the higher education health index of any country. Any country can predict its own higher education health index based on this model, thereby improving the quality of its own higher education. According to the results of the above analysis, when formulating the higher education system, the proportion of education expenditure should first be increased and investment in higher education should be increased. Secondly, we should consider the proportion of international students, encourage college students to study abroad and financially support college students to study abroad. Consider again increasing the enrollment rate of higher education. Finally, in combination with the country’s situation, other influencing factors are comprehensively considered to formulate corresponding policies.

References

[1] Hubei Provincial Commission for Discipline Inspection and Supervision Department. Constructing a punishment and prevention system in colleges and universities to promote the healthy development of higher education [N]. China Discipline Inspection and Supervision News, 2011-08-24 (003).

[2] Zhentian Liu, Li Sen, Mingkai Zhang, Wang Jian, Hongbiao Yin. Written talk: Systematic thinking and classified promotion of high-quality development of higher education[J/OL]. University Education Science: 1-16[2021-10-29].

[3] Shoukui Si, Mathematical Modeling Algorithms and Programs, Naval Aeronautical Engineering Institute, 603-605.

[4] Jiahui Qin, Jiuhao Chen, Zhihao Yang. Evaluation and prediction of the health and sustainability of the higher education system [J]. Science and Technology Wind, 2021(26): 139-141.

[5] Information on: www.ourworldindata.org.