International Journal of Social Science and Education Research Volume 4 Issue 8, 2021
ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

Checking the Pulse and Temperature of Higher Education

Weidong Liu[1, *]

[1]Business School, China University of Political Science and Law, Beijing, China

Abstract

This paper mainly models the health and sustainable status of higher education system, and then puts forward corresponding measures to promote the development of higher education. Firstly, the data of 14 variables that affect the health and sustainability of the national higher education system are processed, and principal component analysis and variable clustering method are used to reduce the dimension of variables, which is convenient for subsequent mod-eling. Then, the health and sustainable state model of national higher education system is estab-lished by using analytic hierarchy process and time series method. Then, through the assess-ment of the health and sustainability of higher education system in each country, the countries with room for improvement in higher education system are identified, the vision is put forward, the targeted policies are formulated, and the effectiveness of the policies is verified. Finally, the actual impact of policy implementation is discussed.

Keywords

Higher education system; Health status; sustainability; Assessment.

1. Introduction

The higher education system is an important part of a country's efforts to further its citi-zens beyond compulsory primary and secondary education. Therefore, it has the value of both the industry itself and the value of providing trained citizens for the country's economy. From Germany, the United States, Japan to Australia, and looking around the world, we have seen higher education methods in various countries. Each country not only educates its own students, but also attracts a large number of international students every year. The higher education sys-tems of these countries have their advantages and disadvantages. Only after adjustments are needed in the current pandemic, will countries have the opportunity to think about what works and what can be done better. The institutional changes required to advance any system require policy implementation over a longer period of time to achieve a healthier and more sustainable system.

Therefore, establish a model to assess the health of the national higher education system, determine the health and sustainability of the higher education system in the relevant country. Then put forward corresponding measures to transform the national higher education system from the current state to a healthier and more sustainable state. This is of great significance for clarifying the development direction of higher education in the country and promoting the de-velopment of higher education in the country.

2. Data Processing and Dimensionality Reduction

2.1. Variable Selection

In order to establish a model for evaluating the health and sustainability of the national higher education system, it is necessary to select factors that affect the health and sustainability of the national higher education system.

ISSN: 2637-6067 | DOI: 10.6918/IJOSSER.202108_4(8).0063

We start from the six aspects of the country's educational resources, scientific research re-sults, educational scale, educational quality and so on, fourteen factors affecting the health and sustainability of the national higher education system are initially selected for subsequent analysis and modeling, such as the proportion of higher education expenditure in GDP, per student expenditure, teacher-student ratio, number of Nobel Prizes, number of papers , the conversion rate of scientific and technological achievements, the penetration rate of higher education, the number of schools, the number of universities, the proportion of doctoral students, the graduation rate, the proportion of foreign students, etc. By searching for relevant information, we collected data of the above 14 influencing factors from 2015 to 2020 in Germany, the United States, Japan, Australia and China.

2.2. Data Processing

In order to make the collected data more reasonable, we need to process the collected receipts. The processing steps are as follows:

•If a variable has more data missing or all are empty, delete the variable;

•Eliminate variable data that does not meet the actual situation;

•When the data of a certain variable in a certain year is null, the average value of the two years before and after the variable is used to replace the null value;

•Remove outliers according to Pauta criterion (3σ criterion).

3σ criterion: Suppose the measured variable is measured with equal precision, and the variable $x_1,\ x_2, \dots \dots, x_n$ is obtained, calculate its arithmetic mean x and residual error $v_i = x_i - x(i = 1,2,\dots,n)$, and calculate the standard error σ according to the Bessel formula. If the residual error $v_b(1 \le b \le n)$ of a certain measured value x_b, satisfying a formula $|v_b| = |x_b - x| > 3\sigma$, it is considered that x_b is a bad value with a gross error value and should be eliminated. The Bessel formula is as follows:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} v_i^2} = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

After removing the relevant variables and data through the above methods, and adding the corresponding null values, we use 3σ to verify all the data and make the calculated value distribution diagram shown in Figure 1. It can be seen from Figure 1 that all the processed data satisfies $|v_b - 3\sigma| < 0$ (the difference is negative), that is, it satisfies the 3σ criterion.

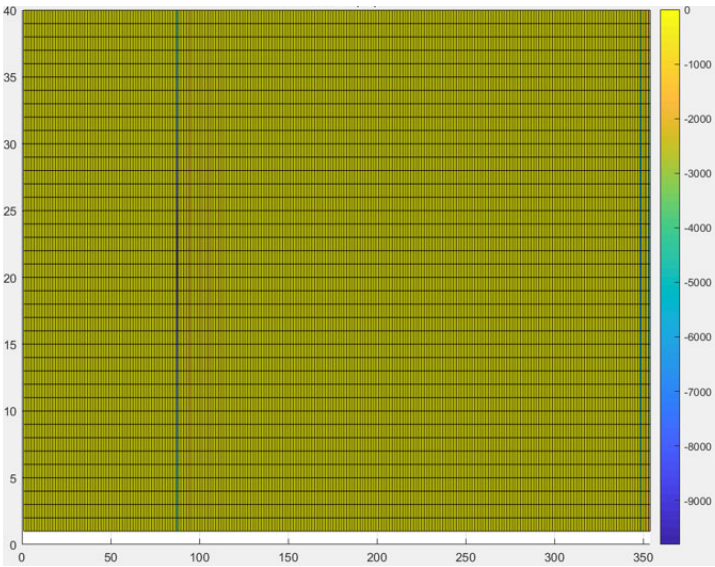

Figure 1. Using the Circular Cone to Simulating the Real Geography

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

2.3. Data Processing

To build a model to assess the health and sustainability of the national higher education system, involving 14 variables. However, many variables increase the complexity of data processing and reduce the efficiency and feasibility of modeling.

Among the variables involved in this question, many variables have a certain degree of correlation to a large extent. Therefore, this article first uses principal component analysis to reduce the dimensions of all variables, and initially selects a number of representative variables to represent all the information contained in all variables. Then, the method of variable clustering is further adopted to aggregate the preliminary screened variables into several categories according to the similar relationship of the variables, and finally the main modeling variables are screened out. This makes the selected variables more representative, independent and reasonable.

2.3.1. Principal Component Analysis

1. Standardize the raw data. Write the sample observation data of the variables in matrix form, as follows:

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Among them, the element $a_{ij}(i = 1,2,3, \dots, m; j = 1,2,3, \dots, n)$ in the matrix represents the observation data of the jth sample of the ith variable.

Because the sample observation data of each variable has the problem of inconsistent dimensions, it will cause the inaccuracy of the matrix solution. Therefore, it is necessary to standardize the raw data through the following methods.

$$\chi_{ij} = a_{ij} - \overline{a_j} / s_j$$

$$\overline{a_j} = \frac{1}{m}\sum_{i=1}^{m} a_{ij}$$

$$s_j^2 = \frac{1}{m-1}\sum_{i=1}^{m}(a_{ij} - \overline{a_j})$$

After the matrix A is processed above, the standardized matrix X can be obtained.

$$X=\begin{bmatrix} \chi_{11} & \chi_{12} & \cdots & \chi_{1n} \\ \chi_{21} & \chi_{22} & \cdots & \chi_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \chi_{m1} & \chi_{m2} & \cdots & \chi_{mn} \end{bmatrix}$$

2. Calculate the covariance and correlation coefficient matrix.

Formula for calculating covariance

$$\operatorname{cov}(x,y) = \operatorname{cov}(y,x) = \frac{1}{m-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_x^2 = \operatorname{cov}(x,x)$$

Can get the covariance matrix

$$S=\begin{bmatrix} s_1^2 & \operatorname{cov}(1,2) & \cdots & \operatorname{cov}(1,n) \\ \operatorname{cov}(2,1) & s_2^2 & \cdots & \operatorname{cov}(2,n) \\ \vdots & \vdots & \vdots & \vdots \\ \operatorname{cov}(n,1) & \operatorname{cov}(n,2) & \cdots & s_n^2 \end{bmatrix}$$

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

Formula for calculation formula of correlation coefficient

$$r_{xy} = \frac{\text{cov}(x, y)}{s_x s_y}$$

Can get the correlation coefficient matrix

$$R_{m,n} = \frac{1}{m-1} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

3. Calculate the eigenvalues and eigenvectors of the correlation coefficient matrix. Sort the eigenvalues from largest to smallest:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \cdots \geq \lambda_n$$

The feature vector is as follows

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix}$$

4. Determine the number of principal components according to the contribution rate of each component. The contribution rate can be calculated by the following formula:

Contribution rate = $\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$

The contribution rate represents the proportion of the variance of a certain principal component to the total variance, that is the proportion of a certain eigenvalue in the sum of all eigenvalues in the above formula. The greater the contribution rate, the more abundant information that a certain principal component contains the original variable.

Through principal component analysis of the sample observation data of 14 variables, the characteristic value, characteristic value contribution rate and cumulative contribution rate of each variable corresponding pair are obtained. As shown in Table 1.

Table 1. Principal Component Analysis Results Table

Number	Contribution rate (%)	Cumulative contribution rate (%)	Number	Contribution rate (%)	Cumulative contribution rate (%)
1	21.36	21.36	8	3.67	91.93
2	18.08	39.44	9	2.54	94.46
3	15.15	54.58	10	2.33	96.80
4	13.20	67.79	11	0.91	97.70
5	8.74	76.53	12	0.85	98.55
6	6.89	83.42	13	0.74	99.30
7	4.84	88.26	14	0.70	100

Generally speaking, if the sum of the contribution rates of the first n eigenvalues is greater than 85%, it can be considered that the matrix composed of principal component variables can

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

contain most of the information of the original variable matrix. It can be seen from the above table that the cumulative contribution rate of the first 10 variables has reached 96.8%, so it can be considered that the first 10 variables basically cover all the characteristic information of the original data, so the first 10 variables are taken as the preliminary dimensionality reduction results for the next analysis .

2.3.2. Variable Clustering Method

1. Variable similarity measure. Use the angle cosine to define the similarity between the 10 variables that were initially selected. The angle cosine calculation formula is as follows

$$c_{ij}=\frac{\sum_{i=1}^{n}x_{ij}x_{ik}}{(\sum_{i=1}^{n}x_{ij}^2\sum_{i=1}^{n}x_{ik}^2)}$$

2. Maximum coefficient method. Similar to the shortest distance method and the longest distance method that are most commonly used in sample cluster analysis, the most commonly used methods in variable clustering are the maximum coefficient method and the minimum coefficient method. Here we use the maximum coefficient method, the maximum coefficient calculation formula is as follows

$$D(G_1,G_2)=\max_{\substack{x_j\in G_1\\ x_k\in G_2}}$$

3. The cluster analysis steps are as follows

•Calculate the maximum coefficient $\{d_{ij}\}$ between n sample points and record it as matrix $D=(d_{ij})_{n,n}$;

• First, construct n classes, each class contains only one sample point, and the platform height of each class is zero;

• Combine the two closest categories with the largest coefficient as the new category, and use the largest coefficient value between the two categories as the platform height in the clustering graph;

• Calculate the maximum coefficient of the new class and the current class. If the number of classes is equal to 1, go to step 5), otherwise, go back to step 3);

•Draw a cluster map;

• Determine the number and class of classes

By further performing variable cluster analysis on the 10 variables obtained from the preliminary dimension reduction, the cluster analysis diagram shown in Figure 2 is obtained.

Table 2. Main variable table after data dimensionality reduction

1	2	3	4
Per student expenditure /$	Number of Nobel Prizes	Number of papers	Higher education penetration rate /%
5	6	7	8
Number of students	Proportion of PhD students /%	Proportion of international students /%	Teaching reputation

According to the cluster analysis diagram shown in Figure 2, the final dimensionality reduction result can be obtained, as shown in Table 2. That is, all the feature information in the original data is represented by the 8 variables shown in Table 2.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

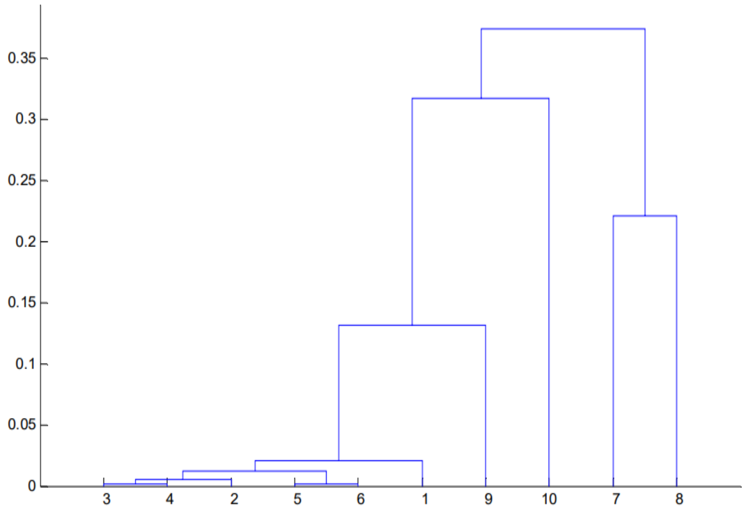

Figure 2. Cluster analysis result graph

3. Higher Education System Evaluation Model

According to the results of data processing and dimensionality reduction above, from six aspects of national educational resources, scientific research achievements, educational scale and educational quality and so on, eight factors affecting the health and sustainability of the national higher education system, including the average birth rate, the number of Nobel Prizes, and the penetration rate of higher education, are selected for the modeling of this chapter.

3.1. Health Assessment

Taking the health of higher education as the target level, regarding the country's education resources, scientific research results, education scale, and education quality as the standard layer, eight factors such as per-student expenditure, number of Nobel Prize, and higher education penetration rate are used as sub-criteria layers to establish an analytic hierarchy process model to evaluate the health status of higher education systems in various countries.

3.1.1. Variable Clustering Method

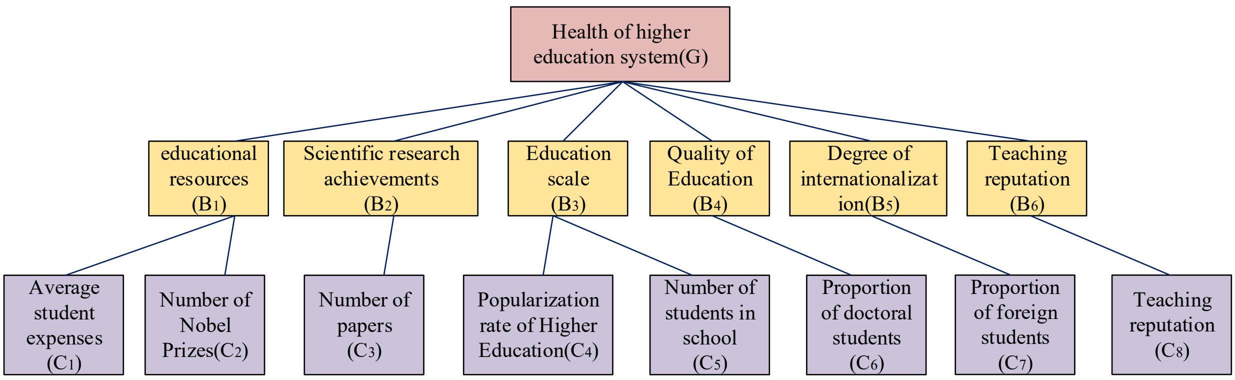

Figure 3. Analysis diagram of hierarchical structure model

3.1.2. Construction of Judgment Matrix and Consistency Test

(1) Judgment matrix

According to the element importance judgment value calibration method, we obtain the judgment matrix of the criterion layer and the sub-criteria layer as shown in the following table.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

Table 3. Judgment matrix at criterion level

G	B1	B2	B3	B4	B5	B6
B1	1	1/3	3	1/4	4	5
B2	3	1	2	1/2	3	4
B3	1/3	1/2	1	1/3	1/2	3
B4	4	2	3	1	3	4
B5	1/4	1/3	2	1/3	1	1/2
B6	1/5	1/4	1/3	1/4	2	1

Table 4. Sub-criteria level 1 judgment matrix

B1	C1	C2
C1	1	3
C2	1/3	1

Table 5. Sub-criteria level 2 judgment matrix

B3	C4	C5
C4	1	4
C5	1/4	1

(2) Eigenvalues and weights

The square root method can be used to obtain the eigenvalues and weights of the criterion layer and the sub-criteria layer

$$\lambda_{G\max} = 6.397, \omega_G = [0.305;0.182;0.223;0.128;0.097;0.065];$$

$$\lambda_{B1\max} = 2, \omega_{B1} = [0.22875;0.07625];$$

$$\lambda_{B3\max} = 2, \omega_{B3} = [0.1784;0.0446];$$

(3) Consistency check

$$CI_G = \frac{\lambda_{G\max} - 6}{5} = 0.1594,\ CR_G = \frac{CI_G}{RI_G} = 0.063<0.1$$

$$CI_{B1} = \lambda_{B1\max} - 2=0,\ CR_{B1} = \frac{CI_{B1}}{RI_{B1}} =0<0.1$$

$$CI_{B3} = \lambda_{B3\max} - 2=0,\ CR_{B3} = \frac{CI_{B3}}{RI_{B3}} =0<0.1$$

Among them, RI is the average random consistency index, as shown in Table 6. Therefore, they all pass the consistency test.

Table 6. Average random consistency index

The matrix order	1	2	3	4	5	6	7	8
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41

3.1.3. Health

Through the results of the analytic hierarchy process, we can get the weight of each influencing factor for health, as shown in Table 7. The influencing factor is represented by $x_i (i = 1,2, \dots ,8)$, and the corresponding weight of the influencing factor is represented by $\omega_i (i = 1,2, \dots ,8)$.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

Table 7. Weights of influencing factors

Factor(x_i)	1	2	3	4	5	6	7	8
Weights(ω_i)	0.22875	0.07625	0.182	0.1784	0.0446	0.128	0.097	0.065

The values of all influencing factors are normalized, and then multiplied by the corresponding weights and summed as the health, namely

$$\rho_n = \sum_{i=1}^{8} \omega_i \times \gamma_{ni} = 0.22875\gamma_{n1} + 0.07625\gamma_{n2} + 0.182\gamma_{n3} + 0.1784\gamma_{n4} + 0.0446\gamma_{n5} + 0.128\gamma_{n6} + 0.097\gamma_{n7} + 0.065\gamma_{n8}$$

Among them, ρ_n is the health of the nth year, and γ_{ni} is the normalized value of the factor x_i in the nth year.

3.2. Sustainability Assessment

Over time, the health of higher education may increase or decrease. In order to better describe this trend of change, we decided to use a time series model for analysis. In the predicted time range, assuming that the health is neither rapidly increasing nor rapidly decreasing, the moving average method can be used to predict future health through the past few years and the current health.

Suppose the health of the past t years is $y_1, y_2, \ldots y_t$, and take the number of moving average items $N < t$. The formula for calculating a moving average is

$$M_n^{(1)} = \frac{1}{N}(y_n + y_{n-1} + \cdots + y_{n-N+1})$$

When the basic trend of health to be predicted fluctuates at a certain level, a moving average method can be used to predict. Therefore, the establishment of a national higher education health prediction model is as follows:

$$\hat{y}_{n+1} = M_n^{(1)} = \frac{1}{N}(\hat{y}_n + \cdots + \hat{y}_{n-N+1}), n = N, N+1, \cdots$$

Its prediction standard error is $S = \sqrt{\frac{\sum_{n=N+1}^{t}\left(\hat{y}_n - y_n\right)^2}{t-N}}$.

4. Conclusion

In order to realize the vision proposed above, we use the following compound interest formula to calculate the annual growth rate of the health of China's higher education system compared to the previous year.

$$\rho_n = \rho_m(1+\delta)^5$$

Among them, ρ_n is the health of China's higher education system in 2025, ρ_m is the health of China's higher education system in 2020, and δ is the annual growth rate of the health of China's higher education system compared to the previous year.

Bring in the data to get δ=0.0676. That is, starting from 2020, when the annual growth rate of the health of China's higher education system compared with the previous year is guaranteed

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202108_4(8).0063

to be 0.0676, the vision of the health of China's higher education system not less than 0.50 by 2025 can be realized.

In order to make the annual growth rate of the health of China's higher education system 0.0676 compared with the previous year, it is necessary to impose corresponding policies on related influencing factors to make corresponding changes. Because the four factors of the number of Nobel Prizes, the penetration rate of higher education, the number of students in schools, and the reputation of teaching cannot change much in the short term, they can only change steadily according to the existing laws. The per-student funding, the number of papers, the proportion of doctoral students and the proportion of foreign students can have a certain impact on the existing law of change by slightly adjusting corresponding policies. Therefore, we propose the following suggestions:

✧ Maintain the number of Nobel Prizes, the penetration rate of higher education, the number of students in schools and the reputation of teaching steadily in accordance with the current laws;

✧ Increase funding for higher education and increase the per-student funding, so that the per-student funding will reach 2.5 times the current stage by 2025, that is, 12,500 yuan per person;

✧ Encourage teachers and students of higher education institutions to actively publish their scientific research results in high-level journals. By 2025, high-quality papers published by Chinese scholars will account for 32% of the world's high-quality papers;

✧ Appropriately increase the number of doctoral students in colleges and universities in order to increase the proportion of doctoral students. By 2025, the number of doctoral students in Chinese universities will increase by 30% compared with the current stage, which will reach about 600,000;

✧ Appropriately increase the number of students studying in China to increase the proportion of international students. By 2025, the number of students studying in China will reach 1.35 times of the current stage, or about 700,000.

References

[1] Duan Lizhong, Liu Sifeng, Lu Qi. Determination of moving steps in moving average method [J] Journal of Beijing University of Technology, 2004, 30(3):378-381.

[2] Si Shoukui, Sun Zhaoliang. Mathematical Modeling Algorithms and Applications [M]. Version 2. Beijing : National Defense Industry Press, 2016 :168-169.

[3] Wang Moran. MATLAB and Scientific Computing [M]. Version 2. Beijing: Electronic Industry Press, 2010 :99-168.

[4] https://www.oecd-ilibrary.org/education/education-at-aglance_19991487.

[5] https://cn.weforum.org/reports/the-global-competitivenessreport-2017-2018.

[6] Zhang Jing, Liao Xiangyang. Establishing the quality evaluation and monitoring system of higher education development guided by student development [J]. China higher education, 2014 (01): 32-34 + 40.

[7] Ma Xing, Wang Nan. Construction of university teaching quality evaluation system based on big data [J]. Education Research of Tsinghua University, 2018, 39 (02): 38-43.