J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 4135--4151

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Optimizing Physical Education Resource Allocation Using Principal Component Analysis

Tao Li[1], Weijie Liu[1,✉], Jingqi Chen[1]

[1] Hunan Sports Vocational College, Changsha, Hunan, 410019, China

ABSTRACT
Physical education teaching resources are an important part of teaching resources, and it is necessary to adopt a sustainable development approach to ensure the rational utilization of resources. In this paper, firstly, the factors affecting the allocation of physical education teaching resources in colleges and universities are analyzed by using principal component analysis and systematic cluster analysis, and the validity of the method is verified. Secondly, it constructs the influential element model of regional physical education teaching resources allocation efficiency level based on Tobit regression, and explores the locational factors affecting the distribution of physical education teaching resources. Finally, relevant countermeasure suggestions were put forward based on the analysis results. Using principal component analysis to downscale the 17 indicators of the influencing elements of physical education teaching resource allocation in the statistical data, four principal components were obtained, whose cumulative contribution rate was as high as 90.22%, which was greater than 85%, i.e., it had a 90.22% degree of explanation for the original data. Then, the dimensionality-decreased data were clustered and realized to evaluate and rank the allocation of physical education teaching resources in 23 sample universities. In addition, the results of Tobit multiple regression analysis showed that factors such as regional geographic location, regional population density, regional economic development and the scale of investment in physical education teaching resources all have different degrees of influence on the allocation efficiency of regional physical education teaching resources.

Keywords: Principal component analysis, systematic clustering method, Tobit multiple regression model, physical education teaching resources allocation

1. Introduction

Education as a cause to enhance the level of social development and realize the development of human beings, to enhance the balance of educational resources, to obtain educational fairness, has become one of the goals pursued by the cause of education, and one of the inevitable ways to build a people-

✉ Corresponding author.
E-mail address: lwj2475@163.com (W. Liu).
Received 01 January 2024; Revised 12 March 2024; Accepted 20 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-228

centered society [1-2]. School physical education, is to ensure that students have a good body and exercise habits, as to enhance students' physical and mental health, promote students' healthy development of educational practices, how to achieve a balanced allocation of school physical education resources, is one of the key elements to drive the balanced allocation of the entire educational resources [3-4]. The importance of school sports in educational practice requires us to continuously strengthen the research on the balanced allocation of school sports resources in China, to create good conditions for sports learning and exercise for every student in school, which is the goal of the balanced allocation of school sports resources [5-6].

Healthy body is guaranteed by adequate physical exercise, actively participate in physical exercise, give people a healthy body, in order to ensure better learning and work [7]. School sports is precisely to teach students relevant sports knowledge and learn how to exercise their bodies. Qualified socialist builders and successors must be moral, intellectual, physical and aesthetic all-round development, which requires that every student constantly participate in physical exercise, for every sports worker, it is responsible for balanced physical education resources, to give every young person a harmonious physical education environment, a fair opportunity to exercise [8-10].

School is a place to “cultivate socialist builders and successors with all-round development of morality, intelligence, physical fitness and aesthetics”, and most of the time for everyone's physical growth is spent in school, so what is the status of physical education resources at all levels and in all types of schools? So what is the situation of physical education resources at all levels and in all types of schools? Is every flower of the motherland able to enjoy fair physical education resources [11-12]? However, the distribution of physical education resources is widespread, this situation is mainly manifested in the inter-school, first of all, in the hardware facilities, the good condition of the school has a standard athletic field, gymnasium, a variety of sports venues with adequate sports equipment, etc., while the poor condition of the school can not even guarantee the most basic sports venues, can not talk about adequate sports equipment, etc. [13-16]. Secondly, in terms of software, key schools and model schools have more middle and senior teachers and higher teacher income, while schools with poor conditions do not have many excellent teachers or even a serious lack of physical education teachers. Further, in terms of student source, schools with good conditions have a lot of preferential conditions for special students in sports, which can attract a large number of students with excellent sports [17-18]. And it is difficult for schools with poor sports conditions to do so. The last manifestation of inequity within the school, in order to strengthen the requirements for promotion, there is a clear difference between the graduating class and the non-graduating class in the time of physical education teaching and exercise, and the graduating class spends very little time on sports [19-20]. The existence of these unfair physical education factors brings inequality to the students in terms of the opportunity to advance to higher education, and the existence of inequality will affect the harmonious development of society, and at the same time does not conform to the requirement that educational resources, as public resources, should be enjoyed fairly by people [21-22].

In this paper, the key factors affecting the efficiency level of physical education teaching resources allocation are explored by using the principal component analysis method, systematic clustering method and Tobit multiple regression model, and relevant strategy suggestions are put forward. Firstly, based on the principal component analysis method, we extracted the principal components of the factors affecting the allocation of physical education teaching resources of 23 sample colleges and universities, and combined with the systematic cluster analysis method, we ranked the level of allocation of physical education teaching resources of the colleges and universities, and at the same time, compared with the traditional principal component analysis method, we verified the validity of the method of this paper. Then, based on Tobit multiple regression model, it explores the influencing factors affecting the distribution of regional physical education teaching resources and their effects. Finally, relevant suggestions are put forward based on the analysis results.

2. Optimizing the allocation of resources for physical education based on cluster-principal component analysis

This paper combines principal component analysis and systematic cluster analysis to analyze the key factors affecting the allocation of physical education teaching resources, so as to construct the influence factor model, explore the influence mechanism of the allocation of physical education teaching resources, so as to put forward the corresponding strategic recommendations.

2.1. Principal component analysis

Principal Component Analysis (PCA) is one of the most fundamental models in multivariate statistical analysis and is widely used as an efficient simplification of multidimensional data sets. Principal component analysis is a method that transforms a multidimensional spatial problem into a low-dimensional spatial problem, and its essential idea is the dimensionality reduction of the data, by reducing the number of indicators in the original data set and forming fewer new indicators, as a way to reduce the complexity of data analysis and make the problem simple and easy to deal with, and the mutual irrelevance of the new indicators ensures the completeness of the original information to the greatest extent possible [23].

2.1.1. Principles of principal component analysis method. In practice, we usually need to collect different aspects of an object in various domains to assess an object. In the available data, these data often have overlapping information, and we cannot accurately distinguish which data have more information about the object to be assessed, and the amount and complexity of the data overwhelm the analysis process. Principal component analysis utilizes the idea of dimensionality reduction to recombine interrelated data and overlapping information into a new dataset in accordance with orthogonal transformations, forming fewer unrelated variables, reducing the dimensionality of the data space and ensuring that the new variables retain the original data, and in the newly formed indicator data most of the information is simplified and the accuracy remains the same, so that the results of the final assessment remain complete and convincing.

From a mathematical point of view it is the use of orthogonal transformations to convert the original random variables with correlated components into fewer new variables with uncorrelated components, the orthogonal system is designed to point in the most open orthogonal direction of the sample points, and then the multidimensional system of variables is downscaled. Assuming that there is n sample, each with p variables, then a data matrix of order $n \times p$ can be constructed, where the observed variable p is denoted by $x_1, x_2, ..., x_p$ and $x_j (j = 1, 2, ..., p)$ represents the j th observed variable. For the sake of discussion, assume x_j that the mean is 0 and the variance is 1, so the data matrix X for the n samples is:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = (x_1, x_2, ..., x_p) \tag{1}$$

Principal component analysis is the process of synthesizing p observed variable into p new variables (composite variables), i.e:

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ F_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ \dots \\ F_p = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \end{cases} \tag{2}$$

Abbreviated as:

$$F_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + \ldots + \alpha_{jp}x_p (j = 1, 2, \ldots, p) \tag{3}$$

The above model can also be represented by a matrix as: $F = AX$, where:

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \tag{4}$$

A is the matrix of principal component coefficients.

The model is required to satisfy the following conditions:

1) F_i, F_j are not correlated with each other $(i \neq j, i, j = 1, 2, \ldots, p)$.

2) The variance of F_1 is greater than the variance of F_2 is greater than the variance of F_3, and so on.

3) The first principal component $a_{k1}^2 + a_{k2}^2 + \ldots + a_{kp}^2 = 1, k = 1, 2, \ldots, p$.

Thus, F_1 is said to be the first principal component, F_2 is the second principal component, and so on, with a pth principal component. In the formula for the process of calculating the principal component is the weight of the eigenvalues a_{ij}, which is called the principal component coefficient in this paper.

2.1.2. Calculation process of principal component analysis method. The specific calculation steps of principal component analysis are as follows:

First, let the original data matrix of the observation sample be:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{5}$$

Where, n is the number of samples and p is the number of variables.

1) Raw data standardization. Data standardization is to eliminate the influence of the outline of each variable, to facilitate the different outline of the indicator data put together for comparison. The data standardization formula is:

$$X' = \frac{X - X_{\max}}{X_{\max} - X_{\min}} \tag{6}$$

In the formula $X_{\max}$ is the maximum value in the original data series and $X_{\min}$ is the minimum value in the original data series.

2) Calculate the correlation coefficient matrix of the sample matrix:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \tag{7}$$

3) Calculate the eigenvalues and eigenvectors. For calculating the matrix sample correlation coefficient matrix R, the p non-negative eigenvalues of the eigenequation value solution eigenequation $|\lambda I - R| = 0$ and the characteristic polynomials shown in Eq. (8) can be found by Jacobi's method and made to be in order of magnitude $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$:

$$r_n\lambda^p + r_{n-1}\lambda^{p-1} + \ldots + r_1\lambda + r_0 = 0 \tag{8}$$

The eigenvector $e_i(i=1,2,\cdots,p)$ corresponding to the eigenvalue λ is required to be $\| e_i \| = 1$, i.e., $\sum_i =1^p e_i^2 = 1$, respectively.

where e_{ij} denotes the jth component of vector e_i. The corresponding eigenvector corresponding to eigenvalue λ is:

$$C^{(i)} = (C_1^{(i)}, C_2^{(i)}, \ldots, C_p^{(i)}), i = 1, 2, \ldots, p \tag{9}$$

And it needs to be met at the same time:

$$C^{(i)}C^{(j)} = \sum_{k=1}^{p} C_k^{(i)} C_k^{(i)} = \begin{cases} 1(i=j) \\ 0(i \neq j) \end{cases} \tag{10}$$

4) Calculate the contribution rate and cumulative contribution rate. The formula for the contribution rate is:

$$CR = \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} (i = 1, 2, \cdots, p) \tag{11}$$

The cumulative contribution rate is calculated by the formula:

$$CRA = \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{p} \lambda_k} (\, i = 1, 2, \ldots, p) \tag{12}$$

The basic idea of principal component analysis is to carry out comprehensive analysis and evaluation by selecting as few principal components as possible, and at the same time, the loss of information should be as little as possible, so it is undoubtedly the best to select fewer eigenvalues to reach a higher cumulative contribution rate under the premise of meeting the above conditions at the same time.

5) Calculate the principal component loadings. The principal component loadings are the correlation coefficients between principal component z_i and variable x_j, calculated by the formula:

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1, 2, \cdots, p) \tag{13}$$

6) Calculate principal component scores and construct individual principal component expressions:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} \tag{14}$$

2.2. Systematic cluster analysis

Cluster analysis is a common tool in mathematical statistics, usually used to classify existing samples into categories in order to simplify the analysis process. Cluster analysis as an important tool in mathematical statistics, it is to "clustering" as the main purpose, the principle of its analysis is to unknown basic characteristics of the sample clustering, using the relevant mathematical methods to calculate the vector mode of the sample matrix and the dispersion of the sample matrix, clustering is based on the calculated distance between the samples.

However, the traditional cluster analysis method always divides the elements in the research samples into a certain category, and the standard of this category division is very absolute and has great limitations. While in real engineering, in most cases, there is no clear classification criteria, but rather the analysis and discussion between multiple categories. Therefore, it is necessary to adopt a correlation division of the elements in the study sample, i.e., fuzzy clustering. The samples obtained by this clustering method can reflect the actual situation more objectively because they contain multiple influencing factors.

Among the various clustering methods, systematic clustering and K-means clustering are the two most frequently used methods. Systematic clustering method is a fuzzy clustering approach, which is built on fuzzy equivalence relations to cluster samples for discussion [24]. For the classical cluster analysis method, if the sample set M is clustered, the classical equivalence relation can be established. For two elements m and n in M, there exists R as a relevant categorization criterion on M. If $(m,n) \in R$, then m and n can be categorized as one class and vice versa.

The basic idea of the systematic clustering method is to consider the elements in the sample as a class, and then determine the correlation between the elements according to the method of defining the distance between classes, aggregate the two or more classes with the highest correlation into a new class, and then according to the distance between the classes, cluster the new class with the other classes, and then aggregate the two or more groups with the highest correlation into a new class. The final clustering result has high correlation between elements of the same class and low correlation between elements of different classes.

When performing cluster analysis, the sample distance and correlation coefficient are determined first. There are various ways of clustering, first calculate the sample distance, and then take different distance criteria to classify, the way is the most frequently used one. Initially there are as many samples are divided into several classes, that is, each sample is a class, and then each time the clustering will be the highest similarity of the two class integration, integration and re-calculation of the similarity between the class and the class, until all the samples are integrated into a class, and the clustering process is plotted as a graph, according to the results shown in the graph can be related to the clustering. Because the clustering flow chart can clearly represent the clustering results, with a strong systematic nature, the systematic clustering method is thus named.

According to the analysis results of systematic clustering, the obtained inter-sample distance is not enough to meet the actual needs, how to determine the distance between classes for further clustering has become the key to solve the problem. For the different ways of defining the distance between classes, the systematic clustering method can be divided into six kinds: the shortest distance method, the longest distance method, the class average method, the center of gravity method, the intermediate distance method and the sum of squares of deviations method.

Here the shortest distance method is used:

Define the distance between classes as the distance between the nearest samples of two classes, i.e.:

$$D_{KL} = \min\nolimits_{i \in G_k, j \in G_L} d_{ij} \tag{15}$$

where d_{ij} is used to denote the distance between the ind sample and the jrd sample, $G_1, G_2, \ldots$ denotes the class, and D_{KL} denotes the distance between G_K and G_L.

When a step class G_K and G_L are merged into G_M, the interclass distance between the new class G_M and the other class G_J is calculated by the shortest distance method with the recursive formula:

$$D_{Mj} = \min\nolimits_{i \in G_M, j \in G_j} d_{ij} = \min\left\{\min\nolimits_{i \in G_K, j \in G_j} d_{ij}, \min\nolimits_{i \in G_L, j \in G_j} d_{ij}\right\} = \min\{D_{KL}, D_{LJ}\} \tag{16}$$

3. Analysis of the factors affecting the allocation of resources for physical education and sport

3.1. Analysis of Influencing Factors on the Allocation of Physical Education Teaching Resources in Colleges and Universities

3.1.1. Principal component analysis. Since the purpose of principal component analysis is to explore the interrelationships of the factors influencing the level of efficiency of physical education teaching resources allocation and the weights of the influences, each factor was treated as a variable and subjected to principal component analysis using MATLAB R2023b software.

Seventeen indicators that mainly respond to the distribution of physical education teaching resources were selected: teachers' education and title (X1), teachers' establishment (X2), teachers' gender (X3), implementation of policies (X4), leaders' emphasis (X5), national policies (X6), physical education capital investment (X7), schools' emphasis (X8), school funding (X9), local economic level (X10), curriculum resources development (X11), classroom teaching format (X12), curriculum construction efforts (X13), informatization level (X14), physical education class status (X15), physical education teachers' information literacy (X16), and students' needs (X17).

Raw data from the allocation of physical education teaching resources in 23 colleges and universities were entered into the MATLAB editor, and the data were subjected to principal component analysis to select the principal components with a cumulative contribution rate greater than 85% of the corresponding principal components, and the list of principal components is shown in Table 1.

Table 1. Principal component list

Principal component sequence number	Characteristic root	Contribution rate	Cumulative contribution rate
1	13.516	0.5025	0.5025
2	2.624	0.2017	0.7042
3	1.735	0.1246	0.8288
4	1.081	0.0734	0.9022

The contribution rate obtained from the principal component analysis is shown in Figure 1. There are obvious inflection points in the change curves of each principal component, and the cumulative contribution rate of the first four principal components reaches more than 85%, which basically represents the information of the 17 indicators, with the first principal component including more than 50% of the information.

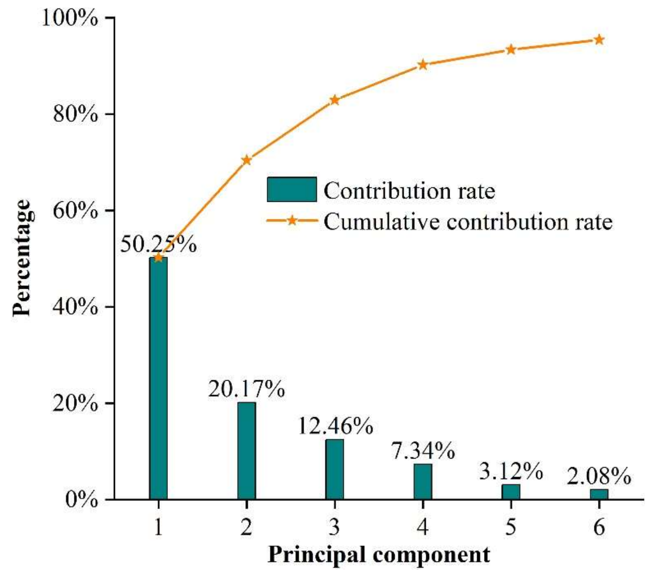

Fig. 1. Contribution rate

The eigenvectors of the first four principal components are shown in Table 2. From Table 2, it can be seen that the original indicators of the sample can be reduced to those that represent most of the sample information.

Table 2. Eigenvectors of the first 4 principal components

Primary index	Principal component 1	Principal component 2	Principal component 3	Principal component 4
X1	0.0546	-0.5334	-0.2971	0.2717
X2	-0.0315	0.0171	-0.6930	0.3395
X3	0.2152	-0.1253	0.3520	0.4591
X4	0.2216	0.0906	0.2500	0.1689
X5	0.2303	0.1490	-0.1392	-0.3397
X6	0.2748	0.0093	0.0826	0.1535
X7	0.2586	0.1555	0.1283	0.1347
X8	0.2749	0.1024	0.0744	0.1716
X9	0.2365	0.1722	-0.1969	-0.2034
X10	0.2608	0.0980	0.0693	-0.0284
X11	0.2749	0.0564	0.0051	0.0593
X12	0.2281	0.1337	-0.2959	0.2705
X13	0.2630	0.1693	-0.1197	-0.0075
X14	0.1506	-0.4590	0.1364	0.1432
X15	0.2607	0.0888	-0.0618	-0.0772
X16	0.2385	-0.2959	-0.0025	-0.3153
X17	0.2450	-0.2503	-0.0318	-0.2414

According to Table 2, the expressions for the four principal components can be obtained, such as the expression for principal component 1:

$$\begin{aligned} Y1 &= 0.0546X1 - 0.0315X2 + 0.2152X3 + 0.2216X4 + 0.2303X5 \\ &+0.2748X6 + 0.2586X7 + 0.2749X8 + 0.2365X9 + 0.2608X10 \\ &+0.2749X11 + 0.2281X12 + 0.2630X13 + 0.1506X14 + 0.2607X15 \\ &+0.2385X16 + 0.2450X17 \end{aligned} \tag{17}$$

The contribution rate and cumulative contribution rate of the four principal components were used as weights to find the weighted average, and the comprehensive evaluation model was constructed:

$$Y = \frac{0.5025Y1 + 0.2017Y2 + 0.1246Y3}{0.9022} \tag{18}$$

The SAS software is used to find out the comprehensive score of 23 sample colleges and universities and rank them, and the results of the comprehensive score ranking are shown in Table 3.

According to the comprehensive score ranking in Table 3, with 0 as the critical value, it can be determined that there are a total of 14 colleges and universities belonging to the balanced distribution of physical education teaching resources, i.e., U1~U14, of which, U1, U2 and U3 belong to the high-efficiency distribution of physical education teaching resources, U4~U10 belong to the distribution of the medium-efficiency level, and U11, U12, U13 and U14 belong to the low level of balance. There are 9 colleges and universities belonging to the type of unbalanced development, i.e., U15~U23.

Table 3. Principal component scores and comprehensive scores

	University number	Y1	Y2	Y3	Y4	Y
1	U1	8.3623	-1.6452	-2.6786	4.7347	3.9198
2	U2	6.1254	0.8574	-1.0651	4.0415	3.4563
3	U3	2.9024	2.2837	0.6796	2.4943	2.2210
4	U4	1.8952	2.0098	2.3773	2.0103	1.8332
5	U5	2.9885	-2.5939	1.3728	1.5437	1.2742
6	U6	0.5915	1.9470	1.0490	0.9438	0.9096
7	U7	1.7639	-2.1522	1.6488	1.3234	0.7290
8	U8	0.6394	1.3453	0.5053	0.7186	0.7267
9	U9	0.0370	1.9215	1.4247	0.5392	0.6470
10	U10	0.8850	-0.7322	1.7244	0.6978	0.5674
11	U11	0.5984	-0.4521	1.1528	0.4643	0.3914
12	U12	-0.1630	1.4876	-0.0608	0.2182	0.2334
13	U13	-0.1777	1.5076	-0.7598	0.0138	0.1331
14	U14	0.4264	-0.9159	0.6556	0.1989	0.1233
15	U15	-0.8001	1.9759	-0.9561	-0.2507	-0.1359
16	U16	-0.2714	-0.2565	0.4086	-0.1765	-0.1521
17	U17	-0.2916	0.03657	-2.8714	-0.4972	-0.5508
18	U18	-0.9457	-0.5059	0.5440	-0.5664	-0.5647
19	U19	-1.1232	-0.3790	-0.1157	-0.7904	-0.7263
20	U20	-0.9803	-0.4953	-0.5165	-0.7551	-0.7281
21	U21	-1.6874	0.5701	0.2784	-1.0253	-0.7739
22	U22	-1.5355	0.1939	-0.6702	-1.0322	-0.9044
23	U23	-1.1786	-2.9555	0.6915	-1.3338	-1.2217

3.1.2. Cluster analysis. The data on the scores of the first four principal components on each sample obtained from the principal component analysis process described above were extracted, and then the extracted data matrix was subjected to cluster analysis using the CLUSTER process in SAS.

From the cluster analysis process, the ordering between classes was determined by the average of the first principal component scores of the samples in each class, and the first principal component scores of the samples in the classes were used to determine the ordering of the samples in each class, and the 23 observation samples (colleges U1 to U23) were divided into three classes: {U1, U2}, {U3, U4, U5, U7, U8, U13, U6, U9, U10, U11, U12, U14, U16}, {U15, U17, U18, U19, U20, U21, U22, U23}. It is thus concluded that universities U1 and U2 are high effectiveness level physical education teaching resource allocation, U3, U4, U5, U7, U8, U13, U6, U9, U10, U11, U12, U14, U16 are medium effectiveness level physical education teaching resource allocation, while U15, U17, U18, U19, U20, U21, U22, U23 are low effectiveness level physical education teaching resource allocation, which is approximately the same as the conclusion obtained from the principal component analysis.

A comparison of the results obtained from the comprehensive evaluation of the various methods of sorting is given as shown in Table 4.

Observing the raw data, we can find that U7 is higher than U6 in all the main indexes, only slightly lower than U6 in the second principal component, but it is lower than U6 in the ranking of traditional principal components, which is obviously not very reasonable, and U7 is higher than U6 in the ranking of "principal components - cluster analysis". Similarly, comparing the ranking of U13 in several methods, U13 is ranked lower in traditional principal component analysis, but observing the raw data, we find that U10 has higher scores than U9 in all the main indexes, so its ranking should be higher than that of U9.

Thus, it can be seen that the "principal components - cluster analysis" in the comprehensive evaluation ranking is more reasonable than the traditional principal components analysis method of comprehensive evaluation ranking, more in line with the actual situation.

Table 4. Various methods sort and compare

	University number	Principal component-Cluster ranking	First principal component ranking	Principal component composite score ranking
1	U3	3	3	3
2	U2	2	2	2
3	U10	9	11	10
4	U18	18	18	18
5	U17	17	16	17
6	U16	16	17	16
7	U15	15	15	15
8	U8	7	9	8
9	U5	5	7	5
10	U9	10	8	9
11	U7	6	6	7
12	U1	1	1	1
13	U6	8	5	6
14	U4	4	4	4
15	U14	14	14	14
16	U22	22	20	22
17	U19	19	19	19
18	U21	21	22	21
19	U13	13	13	13
20	U12	12	12	12
21	U23	23	23	23
22	U11	11	10	11
23	U20	20	21	20

3.2. Analysis of factors affecting the efficiency of regional physical education resource allocation

In order to study the optimization method of physical education teaching resource allocation, this paper further explores the influencing factors of the efficiency level of regional physical education teaching resource allocation on the basis of the previous physical education teaching resource allocation in colleges and universities.

3.2.1. Modeling of Impact Factors. In order to practically test the influence of regional economic development level, the scale of investment in physical education resources, the number of resident population in the region, regional geographic location factors and regional policies on the efficiency of regional physical education resource allocation, this study takes the efficiency of physical education resource allocation as an explanatory variable. The above five influencing factors as explanatory variables, construct the restricted dependent variable Tobit (Hausman test rejects the fixed effect model and accepts the random effect model) random effect model [25] (the value of the variable of the efficiency of the allocation of physical education resources between [0,1]) of the factors affecting the efficiency of the allocation of regional physical education resources to analyze the efficiency of the allocation of regional physical education resources as an explanatory variable.

3.2.2. Analysis of Impact Factors. The study selected the value of regional sports teaching resources allocation efficiency and five hypothetical influencing factors indicators of sports teaching resources allocation efficiency for regression analysis, and calculated the degree of influence of the influencing factors of sports development on the efficiency of regional sports teaching resources allocation through Eviews5.0 software. Before and after regression analysis using two models for analysis, model I will be selected all the indicators variable data for regression analysis, through the significance test to determine the impact of indicator data on the allocation efficiency, will have a significant impact of the indicator data with model II, so as to analyze the degree of impact of the influence factor indicators on the regional sports teaching resources allocation efficiency.

1) ADF test of influencing factor indicator data. For 2014-2022, four indicators of physical education resource allocation efficiency influencing factors such as the level of regional economic development, the scale of investment in physical education resources, regional population density and policy promotion, and 1246 indicator data, the AFD unit root test was used to test the smoothness of the influencing factor indicator data, and to further validate the smoothness of the influencing factor indicator data. The AFD test results As shown in Table 5. Where E_s indicates the value of the allocation efficiency of physical education teaching resources in each region, $InGDP$ indicates the per capita development level of the regional economy, $InScale$ indicates the scale of the input of physical education teaching resources, and $InDensity$ indicates the number of population per 10,000 square meters in the region. "*" indicates a significant difference at the 0.05 level, and "**" indicates a highly significant difference at the 0.01 level.

From the results of the test of the data on the indicators of the influencing factors in Table 5, the values of the efficiency of the allocation of physical education teaching resources in each region and the data on the indicators of the regional population density are at the level of 0.05, which shows a high degree of stability. Regional economic development level, physical education teaching resources scale and other indicators data are at the 0.01 level, showing highly significant stability. It indicates that the data subjected to the test are highly stable and can effectively support the results of model validation.

Table 5. ADF test results

Influence factor variable name	ADF statistics	P value	Test result
ES	-0.887405	0.0342*	Data stationary
InGDP	-0.832659	0.0001**	Data stationary
InScale	-0.805432	0.0002**	Data stationary
In Density	-0.776538	0.0117*	Data stationary

2) Influencing factors indicator data regression analysis. The results of Tobit regression analysis of the influencing factors of physical education resource allocation efficiency are shown in Table 6. Among them, Geography denotes regional geographic location, Policy denotes regional policy orientation, and other indicators are the same as Table 5.

Through the analysis, it is found that among the five influencing factor indicators involved in Model Ⅰ, there are four factor indicators that have significant influence on the allocation efficiency of regional physical education teaching resources. Respectively, they are regional development level, the scale of investment of physical education teaching resources, regional population density, regional geographic location and other influencing factors, while the number of regional policy documents does not have a direct significant impact on regional physical education teaching resource allocation efficiency, considering that the policy-oriented documents of physical education teaching resource allocation do not have a direct significant impact on the efficiency of physical education teaching resource allocation, but mainly indirectly play a role in the The government and relevant sports departments of resource allocation can promote the improvement of sports teaching resource allocation level. Therefore, the policy influencing factors on the efficiency of physical education resource allocation are excluded from Model II.

Table 6. Tobit regression analysis of influencing factors

Influence factor variable name	Model Ⅰ		Model Ⅱ	
	Coefficient	P value	Coefficient	P value
Constant	0.5564*	0.0418	0.5062*	0.0329
InGDP	0.2267**	0.0000	0.2164**	0.0001
InScale	0.1324**	0.0001	-0.1214**	0.0001
In Density	0.4316*	0.0427	0.4237*	0.0209
Geography	0.5728*	0.0459	0.5208*	0.0435
Policy	0.9507	0.7513	-	-

(1) From the level of regional economic development, the level of regional economic development has a highly significant effect on the allocation efficiency of public physical education teaching resources. Among them, the regression coefficient of model Ⅰ is 0.2267, with a P value of 0.0000. The regression coefficient of model Ⅱ is 0.2164, with a P value of 0.0001, and there is a significant positive correlation, which indicates that the rapid development of the regional economy can effectively promote the enhancement of the efficiency of physical education teaching resources allocation. Compared with regions with slower economic development, regions with rapid economic development have greater advantages in the structure of physical education teaching resource allocation, resource input and output mode and technology. From the regression coefficient of influencing factor index data, the regional economic development on the level of regional public sports resource allocation influences between 0.2164-0.2267, which indicates that for every one-fold increase in regional gross domestic product, the efficiency of physical education resource allocation will be effectively improved by 0.2266. Therefore, colleges and universities in economically more developed regions should effectively utilize the scientific resource allocation methods and resource allocation techniques to Promote the improvement of physical education teaching resource allocation efficiency.

(2) From the perspective of physical education teaching resources input scale, the size of regional physical education teaching resources input scale has a significant impact on the improvement of physical education teaching resources allocation efficiency. Among them, the regression coefficient of model Ⅰ is 0.1324, P value is 0.0001. The regression coefficient of model Ⅱ is -0.1214, P value is 0.0001. From the regression coefficient, the scale of input of physical education teaching resources and the efficiency of physical education teaching resources allocation is not completely positive correlation, which is specifically shown that to a certain extent, with the increase of the scale of input of physical education teaching resources, the efficiency of the regional physical education teaching resources allocation will be significantly improved. Resource allocation efficiency will be significantly improved, but after reaching a certain level with the increase in the scale of input of physical education teaching resources, it will have a negative impact on the efficiency of physical education teaching resources allocation. This shows that the pursuit of the increase in the scale of input can not effectively promote the enhancement of the allocation efficiency of physical education teaching resources, but will cause a certain degree of waste of resources. Therefore, in order to effectively promote the enhancement of regional physical education resource allocation efficiency, while guaranteeing the scale of resource input, should further improve the pure technical efficiency of physical education resource allocation, through the scientific way of resource allocation and reasonable resource allocation scale collocation, in order to effectively guarantee the effectiveness of resource input and output.

(3) From the situation of regional population density, regional population density has a significant effect on the allocation efficiency of regional physical education teaching resources. The results of both models show that there is a significant positive correlation between regional population density and the allocation efficiency of regional physical education resources. Specifically, the greater the regional population density, the higher the allocation efficiency of physical education teaching resources. From the regression coefficient, the degree of influence of regional population density on regional public

sports resource allocation efficiency ranges from 0.4316 to 0.4237. This indicates that on the basis of the current regional population size, every 10,000 square meters of population in the region increases by one time, the efficiency of the region's physical education teaching resources allocation effectively increases by 0.4230-0.4310. Therefore, in the process of supplying public physical education teaching resources, it is necessary to effectively configure the public physical education resources according to the regional population density and the teaching needs of the regional colleges and universities.

(4) From the geographical location, the regional geographical location has a significant positive effect on the efficiency of public physical education teaching resources allocation. The regression coefficient of model Ⅰ is 0.5728, and the P value is 0.0459, and the regression coefficient of model Ⅱ is 0.5208, and the P value is 0.0435. Therefore, the regional geographic location has a significant positive correlation with the efficiency of physical education resources allocation.

From the situation of the factors affecting the allocation efficiency of physical education resources, regional geographic location, regional population density, regional economic development and the scale of physical education resources investment and other factors, all to varying degrees, have different degrees of impact on the allocation efficiency of physical education resources. Therefore, in the process of physical education resources allocation, we should fully consider the geographical location of different regions, population density, the level of economic development and the scale of investment in physical education resources, and effectively adjust the scale of investment in physical education resources, the proportion of investment and investment structure according to the actual situation of the region where the university is located. At the same time should also continue to improve the level of physical education resources allocation pure technical efficiency, through the scientific way of physical education resources allocation, reasonable resource allocation supervision system, effectively promote the efficiency of physical education resources allocation.

4. Conclusions and recommendations for response

4.1. Conclusion

In this paper, principal component analysis, systematic clustering method and Tobit multiple regression model are used as research tools to analyze the influencing factors of the level of efficiency of physical education teaching resources allocation and realize the optimization of physical education teaching resources.

First, using principal component analysis, four principal components were extracted from the 17 indicators selected to respond to the distribution of physical education teaching resources, and their cumulative contribution rate reached 90.22%, indicating that the four extracted principal components had an explanatory degree of up to 90.22% to the original data.

Secondly, the systematic clustering of the dimensionality-decreased data was carried out, and the evaluation and ranking of the allocation of physical education teaching resources of the sample colleges and universities were carried out based on the clustering-principal component analysis method. The results show that the principal component-cluster analysis method used in this paper is more reasonable and more in line with the actual situation than the comprehensive evaluation ranking of the traditional principal component analysis method.

Finally, based on Tobit multiple regression to analyze the influencing elements affecting the distribution of regional physical education teaching resources. Its results show that factors such as regional geographic location, regional population density, regional economic development and the scale of investment in physical education teaching resources all have a significant impact on the allocation efficiency of physical education teaching resources to different degrees.

4.2. Recommendations for countermeasures

Based on the previous analysis and the actual situation of physical education teaching resource allocation, this paper puts forward the following strategic suggestions:

1) Change the status quo of physical education teaching resources allocation from the actual situation. While developing the economy, the development of education needs to be regulated by the national policy. Under the guidance of the national policy, for schools with insufficient students, teaching is carried out in accordance with small class size, the teacher-student ratio is adjusted according to the real situation of the school, the school infrastructure is strengthened, the strong and weak schools are allowed to run schools jointly, and some of the schools with a serious shortage of students are withdrawn, the investment in education in developed and underdeveloped regions is balanced, and the investment in physical education teaching resources in underdeveloped regions is increased to achieve a comprehensive and coordinated development of the resources for physical education teaching. The distribution of physical education resources among all levels and types of education departments, among schools at all levels and types of education, and among education departments in all regions should be balanced and rationalized.

At the same time, a survey will be conducted on the situation of physical education resources in all existing schools, and comprehensive improvements will be made in schools with poor physical education resources. Harmonize local and national policies, set hardware standards for physical education teaching in schools regardless of level, regional and urban/rural differences, and economic strength/weakness, and give each school uniform hardware and software standards for physical education teaching, develop each school without discrimination, and increase investment in weak schools, and improve the environment of hardware (physical education equipments, field construction, etc.) and software environment (teachers, etc.). Under the circumstance of improving the hardware (sports equipment, field construction, etc.) and software environment (teachers, etc.), the management of special funds for sports should be strengthened, and the amount and size of equipment and fields should be decided according to the number of students. In addition, it is necessary to strengthen the equipping of physical education teachers in schools and to improve their training and standards, so as to effectively improve the basic quality of physical education teachers and the level of education management. Strengthen the role of the government in the development of local education, reduce the income gap between teachers and maintain the principle of fairness.

2) Improve the physical education examination system. Establish and improve the national unified compulsory education physical education examination system, the national unified physical education curriculum examination standards, in the case of a sound system, under the guidance of the examination standards, and then according to the actual situation of each province, to further improve the physical education examination of the local secondary school examination. Eliminate the waste of sports resources, so as to make the best use of people and things.

3) Establishing sports poverty alleviation strategies. Establish a sports feeding mechanism, so that sports developed schools and regions help weak regions and schools, and actively carry out inter-regional and inter-school exchanges and assistance to promote the development of sports education in weak schools. Adopt the principle of compensation and equality, you can adopt the principle of compensation of sports scores in the secondary school examination, the same region to different conditions of the school assessment stereotypes, according to the school situation, to give the sports comprehensive conditions of poor school students a certain amount of scores to compensate for the balance of the poor conditions of physical education teaching and sports level of the students and the gap between the students in the good school students.

4) Strengthening Education. Strengthen the education of ideology and advanced physical education concepts, management experience, ideological education for students, teachers, school leaders and society as a whole, to change the traditional ideology of emphasizing literature over physical education, and to strengthen students' awareness of exercise. Weak schools' physical education teaching and management concepts have been improved through the organization of short-term training courses for physical education teachers, the sending of teachers from weak physical education areas to

overseas training courses, the sending of outstanding physical education teachers to weak physical education areas to teach, the full implementation of the teacher rotation system, and the emphasis on balanced equipping of the strength of each specific unit of the physical education discipline in the rotation of the teaching staff. For the backward areas, it is necessary to strive for the strength of the central financial investment on the basis of the local government's investment in education, and to strive for the help of the sports sector to the weak schools in sports, sports poverty alleviation needs to be a clear focus to ensure that the responsibility for the implementation of the person to the person, and to strengthen the supervision and assessment.
5) Balanced strategy. Enrollment of students in sports for realistic consideration, but also to be able to take the student sports performance as a reference coefficient, so that each school enrollment of students sports level is basically balanced, the enthusiasm for sports is basically the same, by the students interested in sports to drive the overall sports development. In the evaluation of teachers' titles, the acquisition of honors, and the opportunity to participate in competitions, the policy should be appropriately tilted towards schools with weak student populations, in order to stimulate the motivation and sense of achievement of both students and teachers in the schools.

Funding

Hunan Province basic education teaching reform youth funding project----A Study on the Comprehensive Evaluation and Improvement Strategies for the Physical Fitness of Adolescents in Hunan Province Under the Perspective of Healthy China (Y20230572).

References

[1] Sale, R. S. (2022). The game of teaching resource allocation. International Journal of Education & Development using Information & Communication Technology.

[2] Omoeva, C., Cunha, N. M., & Moussa, W. (2021). Measuring equity of education resource allocation: An output-based approach. International Journal of Educational Development, 87, 102492.

[3] Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L., & Koole, M. (2021). Balancing technology, pedagogy and the new normal: Post-pandemic challenges for higher education. Postdigital Science and Education, 3(3), 715-742.

[4] BenDavid-Hadar, I. (2018). Funding education: Developing a method of allocation for improvement. International Journal of Educational Management, 32(1), 2-26.

[5] Luo, Y., & Zhang, J. (2022, July). Research on the Balanced Allocation Model of Online Education Resources of Economics and Management in Colleges and Universities Based on Parallel Clustering. In International Conference on E-Learning, E-Education, and Online Training (pp. 216-230). Cham: Springer Nature Switzerland.

[6] Sibarani, H. E. I. (2020, March). Survey of Physical Education Facilities and Teaching and Learning Process of Physical Education in State High Schools in the Middle City of Tebing Tinggi Academik Year 2018/2019. In 1st Unimed International Conference on Sport Science (UnICoSS 2019) (pp. 185-188). Atlantis Press.

[7] Demeo, S. (2020). Achieving Balance through Efficiency: A Practical Resource for Time-Minded Teachers. Educe NY.

[8] Lund, J. L., & Kirk, M. F. (2019). Performance-based assessment for middle and high school physical education. Human Kinetics Publishers.

[9] Liu, K. G., Yahiro, K., & Yokoo, M. (2023). Strategyproof mechanism for two-sided matching with resource allocation. Artificial Intelligence, 316, 103855.

[10] Wright, P. M., & Richards, K. A. (2021). Teaching social and emotional learning in physical education. Jones & Bartlett Learning.

[11] Richards, K. A. R., Gaudreault, K. L., Starck, J. R., & Mays Woods, A. (2018). Physical education teachers' perceptions of perceived mattering and marginalization. Physical Education and Sport Pedagogy, 23(4), 445-459.

[12] Pressley, T., Allington, R. L., & Pressley, M. (2023). Reading instruction that works: The case for balanced teaching. Guilford Publications.

[13] Kroupis, I., Kouli, O., & Kourtessis, T. (2019). Physical Education Teacher's Job Satisfaction and Burnout Levels in Relation to School's Sport Facilities. International Journal of Instruction, 12(4), 579-592.

[14] Xia, J., Wang, J., Chen, H., Zhuang, J., Cao, Z., & Chen, P. (2022). An unsupervised machine learning approach to evaluate sports facilities condition in primary school. Plos one, 17(4), e0267009.

[15] Solihin, A. O. (2022, December). Physical Education, Health, and Recreation Study Program Students' Level of Satisfaction with Sports Infrastructures and Facilities at STKIP Pasundan. In IPEC (Vol. 1, No. 1, pp. 78-81).

[16] Magaki, Z. N., Ntabo, J. A., & Nduku, S. D. E. (2021). Status of physical facilities under the public schools infrastructure investment funds in public primary schools of Nyamira County in Kenya. Journal of African Interdisciplinary Studies, 5(6), 56-80.

[17] Rekaa, H., Hanisch, H., & Ytterhus, B. (2019). Inclusion in physical education: Teacher attitudes and student experiences. A systematic review. International Journal of Disability, Development and Education, 66(1), 36-55.

[18] Lee, Y. H. (2019). Emotional labor, teacher burnout, and turnover intention in high-school physical education teaching. European Physical Education Review, 25(1), 236-253.

[19] Rojo-Ramos, J., Manzano-Redondo, F., Adsuar, J. C., Acevedo-Duque, Á., Gomez-Paniagua, S., & Barrios-Fernandez, S. (2022). Spanish physical education teachers' perceptions about their preparation for inclusive education. Children, 9(1), 108.

[20] Abdulla, A., Whipp, P. R., & Teo, T. (2022). Teaching physical education in 'paradise': Activity levels, lesson context and barriers to quality implementation. European Physical Education Review, 28(1), 225-243.

[21] Du Plessis, A. E., & Du Plessis, A. E. (2020). Rethinking Policies: The Balancing Act of Building a Stable Workforce and Improving Teachers' Capacity on an Educational Budget. Out-of-Field Teaching and Education Policy: International Micro-Education Policy, 287-304.

[22] Hollis, J. L., Sutherland, R., Williams, A. J., Campbell, E., Nathan, N., Wolfenden, L., ... & Wiggers, J. (2017). A systematic review and meta-analysis of moderate-to-vigorous physical activity levels in secondary school physical education lessons. International Journal of Behavioral Nutrition and Physical Activity, 14, 1-26.

[23] Younes Khaled. (2025). Application of Principal Component Analysis for the Elucidation of Chemical Compositions' Effect in Frass Compost. Journal of Hazardous, Toxic, and Radioactive Waste(2).

[24] Stephanie Young & Jérôme Gilles. (2024). Use of 3D chaos game representation to quantify DNA sequence similarity with applications for hierarchical clustering. .Journal of theoretical biology111972.

[25] Youssef Er Rays & Meriem M'dioud. (2024). Evaluating the Effectiveness of Maternal, Neonatal, and Child Healthcare in Moroccan Hospitals and SDG 3: Using Two-Stage Data Envelopment Analysis and Tobit Regression. Evaluation review193841X241264863.