Analysis of the effects of college students' sports participation habits on physical fitness

Boyuan Yu*

School of Big Data and Statistics, Anhui University, Hefei, China

*Corresponding author: 1911511215@mail.sit.edu.cn

Abstract. This paper investigates the influence of contemporary college students' sports participation habits on physical fitness test scores and the logical relationship between them. Taking 200 college students from all over China as a sample, the data of four dependent and four independent variables were obtained through questionnaire survey and physical fitness test scores. Based on pre-processing and standardization of the original data, the mathematical and scientific ideas of multiple regression analysis were applied to do correlation analysis and principal component analysis on the dependent and independent variables. The degree of influence of each independent variable on the dependent variable is investigated separately. In connection with physical education, suggestions are made for further reforming physical education and guiding college students to engage in physical exercise.

Keywords: college students; physical fitness; correlation analysis; principal component analysis.

1. Introduction

Physical health is a crucial and multifaceted topic that encompasses various aspects of well-being and plays a central role in our overall quality of life. Physical education is a critical component of carrying out quality education and fostering the overall development of students' morality, intellectuality, physical fitness, aesthetics, and labor, according to The Opinions on Strengthening School Sports to Promote the Overall Development of Students' Physical and Mental Health, published by the General Office of the State Council in April 2016 [1]. It is of great significance to the construction of a healthy nation.

Since the reform and opening up, especially the 18th Party Congress, physical education in higher education institutions has made essential contributions to implementing the Party's education policy, adhering to the people-oriented, and improving the physical quality of college students. However, there still exists a large gap between the current physical quality of college students and the requirements of the country and the expectations of the people. In recent years, research centering on improving university physical education teaching quality and evaluating college students' physical quality has become increasingly interesting topic. Zhang [2] studied the development trend of college students' physical quality based on time series analysis and found that the average value of male students' 1000m test scores is generally in a downward trend, which needs to be solved urgently; the average values of female students' 800m, seated forward bending and vertical jump test scores show an upward trend as a whole, and the time series analysis predicts the performance of college students' physical quality indexes with a minor error, which is in line with the law of statistics. The prediction data have specific guiding significance to reality.

However, to take the predicted value as reference data, it must be combined with the actual situation and relevant intervention measures. Qi [3] studied the indicators of college students' physical fitness test and found that the overall qualification rate is not high and the excellence rate is low. In a word, the overall quality of physical fitness is not high; Li and Liu [4] studied aerobic exercise and resistance training on obese female college students' body composition, body shape, and physical quality indicators in order to provide a basis for weight loss exercise prescription and a new theory and method for China's obese youth weight. Based on the multiplicity of factors affecting college students' physical quality and the interaction between them, this paper uses the mathematical idea of multiple regression analysis to carry out multi-factor correlation analysis and principal component analysis of college students' physical quality.

2. Model Formulation

This paper investigates the influence of contemporary college students' sports participation habits on physical fitness test scores and the logical relationship between them. It provides theoretical references for the current reform of sports club system teaching in colleges and universities and students' independent exercise.

Factors affecting college students' physical fitness are summarized as endogenous and exogenous factors, which are the philosophical basis for this paper to use multiple regression analysis. Among them, the endogenous factors are mainly composed of innate genetic factors and acquired factors such as height, weight, and exercise habits, and the exogenous factors are mainly composed of factors such as exercise environment, exercise conditions, and physical education teaching situation. Among these factors, they are cross-influenced by each other, so analyzing factors affecting college students' physical fitness is a typical multiple regression analysis problem. In the multiple regression analysis problem, correlation analysis is commonly used to find whether there is a correlation between the variables and what kind of correlation, and then the method of principal component analysis is used to reduce the number of variables and achieve the purpose of dimensionality reduction while retaining most of the original primary information.

2.1. Research Process

2.1.1 Variable Setting

Referring to the variable setting method of Huang [5], the physical fitness performance of college students is set as the dependent variable, which consists of the dimensions of athletic agility, athletic endurance and elasticity. The corresponding quantitative indexes are: the total score of the physical fitness test, 50-meter run, standing long jump and men's 1000-meter/women's 800-meter run, which are written by y_j, j=1, 2, 3, 4, respectively. The exercise habits of college students are set as independent variables. The corresponding quantitative indicators are: weekly exercise time, weekly exercise times, the number of sports clubs and the number of sports competition, where we use x_1 to measure weekly exercise time, x_2 to be weekly exercise times, x_3 to measure the number of sports clubs and x_4 to be the number of sports competitions.

2.1.2 Data pre-processing

A sample of 212 college students is randomly selected nationwide, taking into account both the distribution of majors and the proportion of men and women, covering all types of college students.

Firstly, abnormal data is screened and excluded, mainly including data related to students with physical disabilities and incomplete participation in the physical test and finally 200 sets of data are retained as the analyzed sample (among which, 101 are male students and 99 are female students). Secondly, the screened data constitute the observation matrix of the independent and dependent variables, which are denoted as $A = (a_{ij})_{200\times 4}$ and $B = (b_{ij})_{200\times 4}$, respectively.

Because each evaluation indicator differs from the others, the data in a multi-indicator evaluation system typically have multiple scales and orders of magnitude. When the level difference between the indicators is significant, if the raw data are used directly for statistical analysis, the indicators with higher numerical levels will amplify their role in the comprehensive analysis, thus weakening the role of the indicators with relatively low numerical levels [6].

Therefore, the raw indicator data must be standardized before data analysis so as to guarantee the accuracy and reliability of the results. The indicator value a_{ij} of the independent variable is sequentially transformed into the standardized indicator value $\tilde{a}_{ij}$.

$$\tilde{a}_{ij}=\frac{a_{ij}-\mu_j^{(1)}}{s_j^{(1)}},\ i=1, 2, \ldots, 200, j=1, 2, 3, 4.$$

$$\text{where } \mu_j^{(1)}=\frac{1}{200}\sum_{i=1}^{200} a_{ij}\ ;\ s_j^{(1)}=\sqrt{\frac{1}{200-1}\sum_{i=1}^{200}(a_{ij}-\mu_j^{(1)})^2}\ , j=1, 2, 3, 4.$$

Call $\tilde{x}_j=\frac{x_j-\mu_j^{(1)}}{s_j^{(1)}}$, j=1, 2, 3, 4 as the standardized indicator vector.

Similarly, the indicator value b_{ij} of the dependent variable is transformed into the standardized indicator value $\tilde{b}_{ij}$.

$$\tilde{b}_{ij}=\frac{b_{ij}-\mu_j^{(2)}}{s_j^{(2)}},\ i=1, 2, \ldots, 200, j=1, 2, 3, 4.$$

$$\text{where } \mu_j^{(2)}=\frac{1}{200}\sum_{i=1}^{200} b_{ij}\ ;\ s_j^{(2)}=\sqrt{\frac{1}{200-1}\sum_{i=1}^{200}(b_{ij}-\mu_j^{(2)})^2}\ , j=1, 2, 3, 4.$$

Call $\tilde{z}_j=\frac{z_j-\mu_j^{(2)}}{s_j^{(2)}}$, j=1, 2, 3, 4 as the standardized indicator vector.

2.2. Variable Analysis

2.2.1 Correlation Analysis

This part mainly analyzes the strength of the impact of the four independent variables on the four dependent variables respectively. The processed data is analyzed by Eviews software, taking the total score of a physical fitness test, 50-meter running score, standing long jump score, men's 1000-meter/women's 800-meter running score as the dependent variables, and weekly exercise time, weekly exercise times, number of sports clubs, and number of sports competitions as the independent variables, and then correlation coefficients are calculated, and p-values are tested for correlation analysis. The results are shown in Table 1.

According to the meaning of the correlation coefficient, when the correlation coefficient is between 0.00 and 1.00, the two variables are positively correlated. The two variables are negatively correlated when the coefficient is between -1.00 and 0.00. Analyzing the correlation coefficients further after the p-value test is only meaningful. The initial hypothesis is disproved and there is a significant correlation between the two variables when the p-value is less than 0.05, especially when it is less than 0.01, it means highly significant; when the p-value is more than 0.05, it means that they are not correlated.

Table 1. Correlation analysis results

	x_1	x_2	x_3	x_4
y_1 **Correlation**	0.781	0.745	0.605	0.587
p-value	0.000	0.000	0.000	0.000
y_2 **Correlation**	-0.147	-0.145	-0.172	-0.164
p-value	0. 307	0.078	0.036	0.045
y_3 **Correlation**	0.178	0.159	0.217	0.208
p-value	0.030	0.052	0.008	0.011
y_4 **Correlation**	-0.533	-0.498	-0.408	-0.387
p-value	0.000	0.000	0.000	0.000

From the results of the data of the correlation analysis above, the correlation between the dependent variables and the independent variables and whether there is a linear relationship can be obtained. The conclusions are shown in Table 2.

Table 2. Conclusion of correlation analysis

Correlation	x_1	x_2	x_3	x_4
y_1	positive correlation linear correlation	positive correlation linear correlation	positive correlation linear correlation	positive correlation linear correlation
y_2	uncorrelation ——	uncorrelation ——	negative correlation ——	negative correlation ——
y_3	positive correlation ——	uncorrelation ——	positive correlation ——	positive correlation ——
y_4	negative correlation linear correlation	negative correlation linear correlation	negative correlation linear correlation	negative correlation linear correlation

2.2.2 Principal Component Regression Analysis

When using statistical analysis to study a multivariate problem, the number of variables tends to increase the complexity of the problem. The hope is that by using fewer variables, more information can be gleaned. In most circumstances, there is some overlap between the two independent variables, reflecting information about the dependent variables, because there is some connection between the independent variables. In principal component analysis, all of the initially proposed independent variables are taken into consideration, the overlapping variables (those with stronger correlations) are eliminated, and as few new uncorrelated variables as possible are created, preserving as much information as possible about the dependent variables as reflected by these new variables. Principal component regression analysis, a widely used mathematical technique for dimensionality reduction, is a statistical technique that involves recombining the original variables into a new set of several composite variables that are not related to one another as much as possible. This will reflect the information of the statistical method.

Principal component analysis is achieved through the following steps based on data standardization.

In the first step, the correlation coefficient matrix is calculated. The correlation coefficient matrix R=$(r_{ij})_{m*m}$, where $r_{ij}=\frac{\sum_{k=1}^{n}\tilde{a}_{ki}\cdot\tilde{a}_{kj}}{n-1}$, i, j=1, 2, 3, 4.

In the second step, eigenvalues and eigenvectors are calculated. Using Eviews software, the eigenvalues, contributions, and cumulative contributions of the independent variables can be calculated, as shown in Table 3.

Table 3. Eigenvalues, contribution rates and cumulative contribution rates

Eigenvalues	Contribution rates (%)	Cumulative contribution rates (%)
3.4714	86.7800	86.7800
0.3237	8.0900	94.8800
0.1622	4.0600	98.9300
0.0427	1.0700	100

In the third step, the principal components are chosen. According to the calculation formula of the information contribution rate of eigenvalues:

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k},$$

The cumulative contribution rate of features is obtained as:

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{m} \lambda_k}.$$

Generally, when selecting the number of principal components, in order to reduce the workload of the analysis, not all the independent variables can be selected as the number of principal components. Calculating the characteristics λ of the correlation coefficient matrix and the eigenvectors μ, new indicator variables composed of eigenvectors can be obtained as follows.

$$\begin{cases} y_1 = \mu_{11}\tilde{x}_1 + \cdots + \mu_{m1}\tilde{x}_m \\ \vdots \\ y_m = \mu_{1m}\tilde{x}_1 + \cdots + \mu_{mm}\tilde{x}_m \end{cases}$$

From Table 3, it is evident that the rate of cumulative contribution of the first 2 eigenvalues alone reaches more than 90% and the principal component analysis is more effective. Further, the first 2 principal components are selected for comprehensive evaluation as the number of principal components.

$$\begin{cases} y_1 = 0.508061 * x_1 + 0.510913 * x_2 + 0.496197 * x_3 + 0.484386 * x_4 \\ y_2 = -0.507145 * x_1 - 0.457463 * x_2 + 0.383348 * x_3 + 0.621752 * x_4 \end{cases}$$

In the fourth step, construct the principal component analysis model. Establish the principal component evaluation model after choosing the contribution rate of the 2 principal components as weights separately:

$$Y = 0.8678 * y_1 + 0.0809 * y_2.$$

Substituting the 2 principal component values of each sample point into the above equation, the comprehensive evaluation value of each student's physical quality can be obtained, so as to comprehensively evaluate and rank the students' physical quality.

3. Conclusions and Suggestions

The results of correlation analysis show that among the dependent variables, the total physical test score and long-distance running score have a significant correlation with the four independent variables, and the former is obviously positively correlated, and the latter is negatively correlated.

By using the idea of dimensionality reduction and establishing the principal component analysis model for the evaluation of comprehensive physical quality, the physical quality of college students can be evaluated comprehensively.

Reasonably guide students to do physical exercise and provide platforms and opportunities. From the results of this paper, it is easy to see that the four independent variables have a correlation relationship with the dependent variables, and some of them are even linear, which shows the importance of cultivating good exercise habits for improving physical quality. Schools should guide students to exercise more physically and provide more platforms for sports activities and competitions. According to Oja et al., [7], the most beneficial exercises for physical fitness and reducing the risk of cardiovascular disease mortality are swinging, swimming and aerobic exercise. These different health effects are mainly associated with different forms or types of exercise. As Chekroud et al., [8] found that different forms (programs) of exercise act on different dimensions of health. In this way, while stimulating students' interests in sports, it can also help students establish a correct view of sports. Also, the amount and spontaneity of students' sports can be improved.

Put an emphasis on achieving the goals of physical education instruction and significantly enhance the educational impact of university physical education courses. The standards for students' activity intensity both inside and outside of the classroom must be carefully developed and adjusted in order to accomplish the goals of physical education teaching. Medium - high-intensity exercise is conducive to health, which is one of the essential theoretical bases for physical activity to promote health [9]. As an evaluation index of exercise intensity in a physical education class, an experimental team has proved that when students have about 10 minutes of diversified physical fitness exercise in each physical education class, the exercise density is around 75%. The exercise intensity reaches a heart rate of 140-160 beats/minute; together with structured skill exercise and competitions, it can effectively promote children and adolescents' body mass index (BMI), cardiorespiratory fitness, speed, flexibility, muscular strength, muscular endurance, agility, and mental health [10]. It is pointed out that exercise intensity is the critical factor. However, at present, in the teaching of physical education classes, due to the lack of conditions for accompanying data testing, exercise intensity is only a vague concept, resulting in far from enough exercise intensity and a more significant dilution of physical education teaching's target requirements to the students' agility and endurance quality. Both theory and practice show that only when students adapt to the new exercise intensity and can repeat the agility skills in training will their athletic ability subsequently improve [11].

References

[1] General Office of the State Council. The Opinions on Strengthening School Sports to Promote the Overall Development of Students' Physical and Mental Health. [EB/OL]. (2016-05-06) [2021-03-11]. https:/ / baike. sougou. com/ v170725446. htm .

[2] Zhang Min. Research on the development trend of college students' physical fitness based on time series analysis [D]. Taiyuan: North University of China, 2020:20.

[3] Qi Yue. Research on Physical Fitness Indicators of College Students' Physical Fitness Test [J]. Research on Sports Science, 2018, 22(6): 61-65+84.

[4] Li Baitong, Liu Yujiao. Effects of aerobic exercise and resistance training on body composition, body shape and physical fitness indexes of obese female college students [J]. Exercise, 2013(8): 39-40.

[5] Huang Zhengfeng. Multiple regression analysis of the effect of college students' sports participation on physical fitness [J]. Journal of Anhui Normal University (Natural Science), 2022, 45(4): 395-401.

[6] Li Yanwu. Application of partial least squares regression analysis in the calibration of air quality monitoring data [J]. Journal of Huaibei Normal University (Natural Science Edition), 2020, 41(4): 16-20.

[7] Oja. P, Kelly. P, Pedisic. Z, et al. Associations of specific types of sports and exercise with all-cause and cardiovascular-disease mortality: A cohort study of 80306 British adults [J]. British Journal of Sports Medicine, 2017, (51): 812-817.

[8] Chekroud. S. R, Guerguieva. R, Zheutlin. A. B, et al. Association between physic al exercise and mental health in 1.2 million individuals in the USA between 2011 and 2015: A cross-sectional study [J]. The Lancet Psychiatry, 2018, 9(5): 739-746.

[9] Xiong Wen. Questioning and Discerning:The Intensity Correlation of Healthy Pursuit in School Physical Education---A Reflection on the Specific Exercise Intensity Based on Physical Education Curriculum Teaching [J]. Journal of Shanghai Institute of Physical Education, 2021, 45(1): 86-98.

[10] Ji Liu. Focusing on Frontier Hot Issues, Paying Attention to the Development of Physical Education, and Exploring the Future of Students' Growth---Review of 2019 International Conference on Physical Education Curriculum and Teaching [J]. Journal of Chengdu Institute of Physical Education, 2019, 45(5): 78.

[11] Tudor Bompa. Physical Training for Young Athletes [M]. Shanghai: Shanghai Culture Press, 2017: 157.