

College Teachers' Physical Health Medical Evaluation Research Based on Linear Regression

Fanfei Ge^{1, a}, Shujian Kao^{2, b}

¹ School of electronics and Information engineering, Tongji University, Shanghai, 201804, China

² School of physics science and engineering, Tongji University, Shanghai, 200092, China

^a1336269@tongji.edu.cn, ^bkaoshujian@tongji.edu.cn

Keywords: Linear Regression, Data Modeling, College Teachers, Physique Health, Medical Evaluation.

Abstract. According to current situation of college teachers' physique health, with the help of college teachers' physical examination data, on the basis of physical health evaluation system, using multiple linear regression method, established college teachers' physical health data model, designs and realizes a kind of college teachers' medical evaluation system. Thesis research work of college teachers' physical health prevention has carried on the beneficial exploration.

1. Introduction

Recent researches show that the health of college teachers is not optimistic. In the college teachers' physical test results, indicators that have greater impacts on the health get low scores. College teachers are high risk population of hypertension, hyperlipidemia, high cholesterol and other chronic diseases [1].

The paper is based on the medical indicator data, depends on the physical health evaluation standards, use multiple linear regression to design and build data model. The research provides scientific evaluations and decision supports for improving health of college teachers.

2. Functional Requirements

The college teachers' medical evaluation system mainly achieves the following features, just like figure 1.

2.1 Data Collection.

Determine the teachers' medical indicators and collection methods to collect basic information and medical data of college teachers.

2.2 Data Storage.

Storing the collected data into database for establishing data model.

2.3 Establishing Data Model.

Basing on the database, depending on the physical health evaluation standards, we use multiple linear regression to design and build data model.

2.4 Health Advice.

Depending on the established data model, we can conclude teachers' health evaluation score and offer relate prevention advice.

2.5 Visualization.

Making visualization on the results of the evaluations, and intuitive for users to view and analyze information on their medical evaluations.

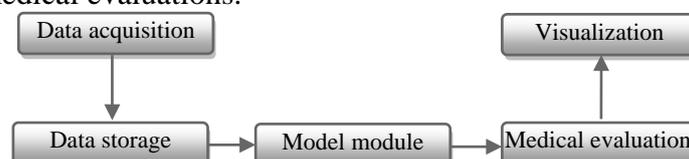


Fig. 1. System functional requirements

3. The Design of the System

The system consists of three parts, namely data acquisition and storage module, setting up system module, and medical health evaluation and visualization module.

3.1 Data Acquisition and Storage Module.

(1) Determine the indicator system

The indicators that we need to test reference to medical standards. Depending on expert knowledge, we make some improvement for taking medical tests and making comments on college teachers' physical health. At last, we determine the targeted medical indicator system, as shown in figure 2. The full name of these indicators are as follows. CR: creatinine, ALT: Alanine aminotransferase, UA: Uric acid, BUN: Urea nitrogen, TC: total cholesterol, TG: Triglyceride, HDL: High density lipoprotein cholesterol, LDL: Low density lipoprotein, GLU: Blood glucose.

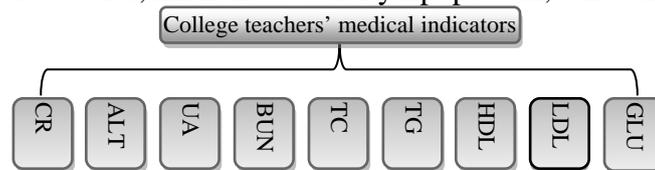


Fig. 2. Medical indicators system of college teachers' physical health

(2) Methods of data Acquisition

There are several methods to collect data: Design website to collect college teachers' basic information and medical health data; Wireless measurement equipment; Mobile; Email.

(3) Data Storage

In the system, data are divided into two categories: basic user information data and medical information data. Storing the data into database. Depending on the situation of each topic, we design its data model.

3.2 Establishing Data Model.

The system combines these three aspects to set up data model.

(1) Theoretical Preparation

Refer to domestic and international physical health research materials, understanding the current situation of domestic and foreign medical health research methods, tools and ideas.

(2) Data Acquisition

According to medical health indicators standard, we test college teachers and get test data.

(3) Regression Analysis

Regression analysis is a statistical analysis method that defines interdependent quantitative relationship between two variables or more. In this paper, we will use multiple linear regression method. The independent variables are college teachers' basic medical indicators, dependent variable is overall health evaluation. This method uses the least square method to establish model between various independent variables and dependent variable. This function is a linear combination of one or more model parameters that called regression coefficients [2].

3.3 Health Prediction and Visualization Module.

(1) Health Prediction

Tester's data are taken into relate module. Then we can get all indicators' goodness scores and overall evaluation score. Depending on their professional knowledge and experience, experts make comments on these scores and give health prevention advices.

(2) Visualization

Basing the indicators' goodness score, we use matlab to the simulation curve. Then teachers can know their health situation intuitively.

4. The Realization of the System

4.1 Data Acquisition and Storage.

We use thinkPHP structure and database technology to build the physique monitoring website to collect data. Database is used to store data.

4.2 Scoring Goodness of Indicators.

The expert interviews method depends largely on the selection of experts, because different experts may give a high score on the indicators which they are familiar with or pay more attention to. So it's improper to select so many or few experts.

The test medical data are 70% as the training set and 30% as the test set data. Depending on their professional knowledge and experience, reference to the test equipment, methods and standard values, the experts score about indicators' goodness based on the data.

The table 1 is the relative information about test equipment, methods and standard values which get from the professional hospital. Beckman is an equipment for medical tests which is made by America Beckman Kurt Ltd.

Table 1: Indicator's test equipment and its method, reference value

Indicator	Test equipment	Test method	Reference value
CR	Beckman	Enzyme colorimetric method	30-133 umol/L
ALT	Beckman	Rate method	0-60 u/L
UA	Beckman	Enzyme colorimetric method	134-319 umol/L
BUN	Beckman	Enzyme colorimetric method	2.2-6.8 mmol/L
TC	Beckman	Enzyme colorimetric method	3.1-5.7 umol/L
TG	Beckman	Enzyme colorimetric method	0.56-1.8 mmol/L
HDL	Beckman	Chemically modified enzymatic	0.78-2.0 mmol/L
LDL	Beckman	Homogeneous measurement	0-3.64 mmol/L
GLU	Beckman	hexokinase	3.8-6.1 mmol/L

Table 2: Goodness score on indicator's test data and medical evaluation (M_E) score

No	Sex	Age	CR	CR	ALT	ALT	UA	UA	BUN	BUN	TC	TC	TG	TG	HDL	HDL	LDL	LDL	GLU	GLU	M_E	
			x1	x2	x3	x4	x5	x6	x7	x8	x9	Y										
1	M	36	74	7.8	10	7.0	13	6.2	6.3	7.4	3.1	7.3	1.6	7.5	1.3	7.8	1.4	7.6	4.1	7.6	7.5	
2	F	44	50	7.4	70	6	28	7.2	2.0	6.2	2.8	6.5	0.9	7.8	0.5	6.4	0.6	7.4	3.2	6.5	6.9	
...
...	M	41	45	7.2	50	7.1	15	7.4	7.5	6	5.9	6.4	2.1	6.5	2.3	6	4.1	6.4	9.3	6.2	6.8	

After the experts score about the indicators goodness, they will make the overall medical evaluation Y, just like table 2. Then we establish the physical health medical evaluation model. CR stands for the teachers' test data of CR. CR x1 stands for CR's goodness score. M_E stands for overall medical evaluation goodness score [3].

Precautions: experts should score about one kind of indicator, then others; the score should be between 0 and 10, accurate to one decimal place. High score indicates high level health situation.

4.3 Setting up Module of Medical Health Evaluation.

(1) Setting Up Module

The model value Y is the dependent variable, all indicators' goodness score x1, x2, x3, x4, x5, x6, x7, x8, and x9 are independent variables, then establish evaluation model $Y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) = p_0 + p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 + p_5x_5 + p_6x_6 + p_7x_7 + p_8x_8 + p_9x_9 + \varepsilon$, p_0, p_1, \dots, p_9 are unknown parameters, called regression coefficients, $\varepsilon \sim N(0, \sigma^2)$.

For data $(M_i; x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}, x_{i7}, x_{i8}, x_{i9})$, $i=1, 2, \dots, n$. Then model's matrix form is

$$Y = XP + V$$

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \dots \\ M_n \end{bmatrix} X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{19} \\ 1 & x_{21} & x_{22} & \dots & x_{29} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{n9} \end{bmatrix} P = \begin{bmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{bmatrix} V = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}$$

Using the least squares method to estimate multiple linear regression coefficients, $p_0, p_1 \dots p_9$ [4] [5], we get the medical evaluation function:

$$Y = x_1 * 0.2065 + x_2 * 0.2347 + x_3 * 0.926 + x_4 * 0.02398 + x_5 * 0.01569 + x_6 * 0.0003 + x_7 * 0.7275 + x_8 * 0.1698 + x_9 * 0.135 - 9.731$$

The table 3 is model's abstract. Goodness of fit R2's span is [0,1], if its value is close to 1, it shows that actual observation point is closer to the sample line and the goodness of fit is higher. Generally, R2 is required more than 0.7. The R2 of table is 0.999 which shows that this model's goodness of fit is very high and meets the requirement.

Table 3: Goodness score on i

Model	R	R2	A djusted R2	Std. Error of the Estimate
1	1.000(a)	0.999	0.998	0.34766

a. Predict Variables: (Constant),x1,x2,x3,x4,x5,x6,x7,x8,x9.

The table 4 is ANOVA (a).

Table 4: Variance analysis

Model	R	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12.807	9	1.423	1177.279	0.000(b)
	Residual	0.008	7	0.001		
	Total	12.815	16			

Dependent variable: Y

Predict Variables: (Constant),x1,x2,x3,x4,x5,x6,x7,x8,x9.

(2) Module Validation

We take the test set data to the model and analytic results are: Minimum error: -0.221, Maximum error: 0.22, Mean error: 0.00, Absolute mean error: 0.12, Standard deviation: 0.133, and Linear correlation: 0.99.

(3) Visualization of Signal Indicator Module

Based on a single indicator of measured values and the corresponding quality scores, curve fitting which indicators measured value is horizontal axis, goodness score is the vertical axis. Figure 3 is a fitting curve CR excellent degree, wherein x is CR, y is excellent score. Teachers can clearly find their CR situation from this curve.

Get the fitting function, CR's goodness function model is as follow:

$$y = p_1 * x^4 + p_2 * x^3 + p_3 * x^2 + p_4 * x + p_5$$

Coefficients (with 95% confidence bounds):

$p_1 = -1.286e-08$ (-1.617e-08, -9.541e-09), $p_2 = 1.025e-05$ (8.257e-06, 1.225e-05), $p_3 = -0.002925$ (-0.003316, -0.002533), $p_4 = 0.2991$ (0.2716, 0.3266), $p_5 = 0.2694$ (-0.24, 0.7788)

Goodness of fit: SSE is 5.064, R2 is 0.987, Adjusted R2 is 0.9853, and RMSE is 0.4108.

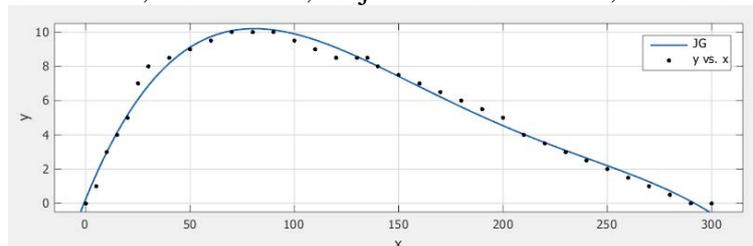


Fig. 3. Fitting curve of CR excellent degree

4.4 Medical health evaluation.

In this part of the work, tester's data are introduced into the indicators' model function. Then we can get all indicators' goodness scores and take these scores into overall evaluation model for Y's score. Depending on their professional knowledge and experience, experts make comments on these scores and give relative advices.

(1) Overall Evaluation

Experts make different comments with different medical evaluation scores. We conclude the results as table 5. Y is divided into five levels, their names and score range are worse (0-3.9), bad (4-4.9), general (5-6.9), good (7-8.4), and excellent (8.5-10). Their evaluations are as follows:

Table 5: Physical health medical evaluation rank

Rank of physical health medical evaluation	Evaluation
Worse (0-3.9)	Immediately go to hospital for treating and check in all the indicators' health situation
Bad (4-4.9)	Try to make some changes in daily life. Check in some indicators
General (5-6.9)	Change life style to make it better.
Good (7-8.4)	Keep going on and try to be excellent
Excellent (8.5-10)	Very healthy and please keep going on

(2) Indicators' Goodness Evaluation

From the model Y, we can find that x3 and x7 which stand for UA and HDL have a great influence on medical health. So college teachers should pay more attention to these indicators.

Experts give evaluations and advices on some indicators which have low score of goodness. Taking GLU for example, if its score is very low, the tester will be advised to ensure nutrition and keep GLU to be a normal value. Table 6 is GLU's goodness score and relative evaluations.

Table 6: Goodness score on GLU

Goodness score on GLU	Evaluation
Worse (0-3.9)	Your GLU value is very dangerous , please go to hospital for treating and change life style immediately
Bad (4-4.9)	Your GLU value is bad, please go to hospital
General (5-6.9)	Your GLU is close to dangerous level, please change your life style
Good (7-8.4)	Your GLU is good, please keep going on
Excellent (8.5-10)	Your GLU is excellent, please keep going on

5. Summary

In this paper, with the help of college teachers' physical examination data, on the basis of physical health evaluation system, using multiple linear regression method, designs and realizes a kind of college teachers' medical evaluation system. On the basis of further research in the future, the function of the system will be improved constantly.

References

- [1] LIU Jian, ZHANG Yu. "Liaoning university teachers' sub-health condition detection and cause analysis", Journal of Shenyang Sport University, vol. 31, No. 2, pp 63-66, 2012.
- [2] WANG Huiwen, MENG Jie. "Predictive modeling on multivariate linear regression", Journal of Beijing University of Aeronautics and Astronautics, Vol. 33, No. 4 pp 500-504, 2007.
- [3] Hamid Moghaddasi, Reza Rabiei. "A Model for Measuring e-Health Status Across the World", Telemedicine and e-Health, Vol. 19, No. 4 pp 322-7, 2013.
- [4] Font-i-Furnols, M. "Use of linear regression and partial least square regression to predict intramuscular fat of pig loin computed tomography images", Chemometrics and intelligent laboratory systems, Vol. 122, No. 4 pp 58-64, 2014.
- [5] Agha, SR. "Neural network and multiple linear regression to predict school children dimensions for ergonomic school furniture design", Applied ergonomics, Vol. 43, No. 6 pp 979-984, 2012.