MEDS Public Health and Preventive Medicine (2021) 1: 59-62
Clausius Scientific Press, Canada
DOI: 10.23977/phpm.2021.010110
ISSN 2616-1915

Research based on Analytical model on indicators affecting oxygen saturation

Shuge Yuan

International Business College, Dongbei University of Finance and Economics, Dalian, Liaoning, 116025

Keywords: Principal component analysis, Regression analysis, Correlation analysis

Abstract: The oxygen saturation concentration is an important physiological parameter of the respiratory cycle [1]. In order to explore the relationship between the influencing factors of oxygen saturation, there is an analysis of indicators related to the average oxygen saturation. In this part, using the method of correlation analysis. The five independent variables are Gender, BMI, Age, Smoking Status1, Smoking Status2 and the dependent variables are average oxygen saturation. Then, the partial correlation analysis model of five independent variables with the average oxygen saturation is established and analyzed, and it is found that the age factor has the largest relationship with the average oxygen saturation. By establishing multiple linear regression models, we use the least square estimation method to analyze the regression coefficient and study the influence degree of the age factor on the average oxygen saturation. Finally by extracting the principal components and calculating the principal component scores, the conclusion can be drawn: the older the age, the greater the influence of the age factor on oxygen saturation.

1. Background

Oxygen saturation is the percentage of the volume of oxygen-bound oxyhemoglobin in the blood to the total volume of hemoglobin that can be combined. Pulse oximetry is a conventional instrument to detect the oxygen saturation of the patient [2]. This type of instrument uses light-emitting diodes to emit red light and infrared light, passing through appropriate peripheral tissues on the human body, and reaching the optical sensor on the opposite side. The respiratory department often uses oxygen saturation monitoring as an index to judge the patient's respiratory function [3]. So, the research on oxygen saturation has great biological and medical significance.

2. Model preparation

As shown in Figure 1, the correlation coefficient between Gender and average oxygen saturation is -0.185 showing a negative correlation; the correlation coefficient between BMI and average oxygen saturation is -0.188 showing a significant negative correlation; The correlation coefficient between average oxygen saturation and Age is -0.413 which is the smallest correlation coefficient.And the correlation coefficient between Smoking Status1 and average oxygen saturation -0.327, showing a significant negative correlation.The correlation coefficient between Smoking Status2 and average

oxygen saturation is -0.272, showing a significant negative correlation.

There is a negative correlation between the average oxygen saturation and the content of the other 4 indicators, which can be fitted to the commonly used primary linear model, quadratic model, or cubic model.

		Gender	BMI	Age	Smoking Status1	Smoking Status2	average oxygen saturation
Correlation	Gender	1.000	.413	.084	.117	.030	-.185
	BMI	.413	1.000	.304	.214	.064	-.188
	Age	.084	.304	1.000	-.080	.319	-.413
	Smoking Status1	.117	.214	-.080	1.000	.564	-.327
	Smoking Status2	.030	.064	.319	.564	1.000	-.272
	average oxygen saturation	-.185	-.188	-.413	-.327	-.272	1.000

Figure 1: Correlation matrix

The linear fitting model is:

$$y = \beta_0 + \beta_1 x \tag{1}$$

The quadratic model is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{2}$$

The cubic model is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \tag{3}$$

y is the average oxygen saturation,and x is the content of one of the other five sub-indexes for pairwise analysis with the average oxygen saturation.

Establish a general model of multiple linear regression:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \tag{4}$$

Among them, the independent variable x_1 represents gender, x_2 represents BMI, x_3 represents age, x_4 represents Smoking Status1,and x_5 represents Smoking Status2.The dependent variable y represents the average oxygen saturation.ε is a random error,subject to $N(0 \cdot \sigma^2)$.

There are 36 sets of data:

$$\begin{bmatrix} 1 & y_1 \\ 2 & y_2 \\ 3 & y_3 \\ \vdots & \vdots \\ 36 & y_{36} \end{bmatrix} \tag{5}$$

Bring into the above formula: $y = \beta_1 x_i + \beta_2 x_i + \beta_3 x_i + \beta_4 x_i + \beta_5 x_i + \varepsilon$

$$\mathrm{i} = 1,2,3 \cdots 36$$

For convenience, the matrix notation is introduced:

$$\mathrm{Y}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{36} \end{bmatrix} \qquad \mathrm{X}=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{36} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} \tag{6}$$

Choose $\hat{\beta}\ as$ an estimated value of β to minimize the sum of squares of random error ε:

$$min\varepsilon^t\varepsilon = min(Y - X\beta)^T(Y - X\beta) \quad (7)$$

$$Q(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5) = minQ(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \quad (8)$$

$$\frac{\partial Q}{\partial \beta_j} = 0(j = 1,2,3,4,5) \quad (9)$$

$$\sum_{i=1}^{n} x_i \sqrt{Q(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5)} = 0 \quad (i = 1,2\ldots36) \quad (10)$$

Bring $\hat{\beta}$ into the model, get $\hat{Y} = X^T\hat{\beta}$

Using the SPSS software,the influence of the five independent variables on the average oxygen saturation of the dependent variable can be obtained when the remaining four independent variables are fixed.It can be find that the partial correlation between average oxygen saturation and age is -0.464,the partial correlation between average oxygen saturation and BMI is 0.120, and the partial correlation between average oxygen saturation and gender is- 0.155,the partial correlation between average oxygen saturation and smoking status1 is -0.383,and the partial correlation between average oxygen saturation and smoking status2 is 0.134. The analysis shows that among the above five independent variables, the average oxygen saturation has a relatively obvious negative correlation with age, and the correlation coefficient is relatively large.

3. Modeling and solving

3.1 Model Establishment

As the age factor has a large influence on the average oxygen saturation, according to the cubic model, a fitting model is established for the independent variable age and the average oxygen saturation of the dependent variable with better fitting effects, as shown in the following figure:

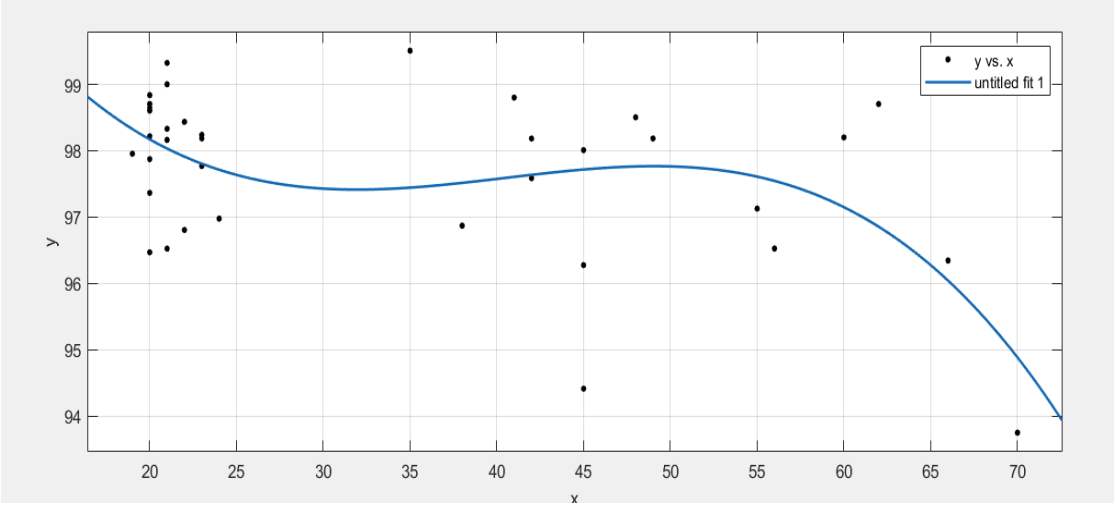

Figure 2: Fitting effect diagram of average oxygen saturation and age

Then the fitted cubic model of average oxygen saturation and age is:

$$y=-0.000141x^3+0.01714x^2-0.6632x+105.7 \quad (11)$$

Suppose the three principal components are Y_1, Y_2 and Y_3.And Gender, BMI, Age, Smoking Status 1, and Smoking Status 2 are set to X_1, X_2, X_3, X_4 and X_5. According to the calculated principal component coefficients, we can get the Linear combination of Y_1, Y_2, Y_3:

$$Y_1 = 0.360X_1 + 0.477X_2 + 0.363X_3 + 0.490X_4 + 0.520X_5 \quad (12)$$

$$Y_2 = 0.499X_1 + 0.481X_2 + 0.223X_3 - 0.469X_4 - 0.500X_5 \quad (13)$$

$$Y_3 = -0.416X_1 - 0.103X_2 + 0.785X_3 - 0.396X_4 + 0.209X_5 \quad (14)$$

3.2 Model solving

It can be seen from the above formula that the absolute value of the coefficients of Smoking Status 1 and Smoking Status 2 in the principal component Y_1 is greater than that of other variables, so the principal component Y_1 is a comprehensive reflection of personal smoking status. In the principal component Y_2, The absolute value of the coefficient of the Gender and BMI is greater than that of other variables, so the principal component Y_2 is a comprehensive reflection of the combination of Gender and BMI. In the principal component Y_3, the absolute value of the coefficient of Age is greater than the absolute value of the coefficients of other variables, so the principal component Y_3 is the individual's age reflection.

From the perspective of the principal component score Y_3, that is, the principal component reflecting the individual's age, the score basically shows a trend of increasing with age. This can indicate that the older the age, the higher the influence of age on the oxygen saturation.

4. Evaluation of Model

4.1 Advantages

When discussing the influence of age on oxygen saturation, we use bivariate correlation and partial correlation analysis, so as to more accurately and scientifically analyze the real impact of a variable on the dependent variable. After the conclusion is drawn, it is checked by principal component analysis to make the conclusion more convincing.

4.2 Disadvantages

Only 36 sets of data provided were used. The established model and the conclusion may have little errors because of the small amount of data.

References

[1] SHEN Jun,ZHENG Zheng.Application of pulse blood oxygen saturation monitoring in elderly patients with chronic heart disease.Department of Nursing,Zhongshan Hospital,Fudan University,Shanghai 200032,China.

[2] LIU Guang-wei, MAO Lu-hong, XIE Sheng, MENG Qing-bin, ZHANG Qing-rong. A monitoring method of blood-oxygen saturation using reflection type photoelectric sensor.1.School of Electronic and Information Engineering, Tianjin 300192,China; 2.Binhai College of Nankai University, Tianjin 300270; 3.Tianjin Hexi Hospital,Tianjin 300202, China.

[3] LYU Jian, YANG Hong-xia. Effects of different factors of operation and anesthesia methods before and after bronchial foreign body surgery on blood oxygen saturation. Department of Otorhinolaryngology, Taian Maternal and Child Health Hospital, Tian 271000, China.