ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

Exploration of the Factors of Blood Oxygen Saturation Measurement Mode based on Regression Analysis

Yawen Huang

School of Statistics and Applied Mathematics, Anhui University of Finance & Economics, Bengbu, 233030, China.

Abstract

Blood oxygen saturation is one of the most important parameters of human physiology. Non-invasive, real-time and continuous monitoring of blood oxygen saturation is of clinical importance. Currently, the more conventional monitoring method is pulse oximetry, but the ability to directly calculate oxygen saturation using a specific model has always been an issue. Based on certain assumptions, this paper first establishes a multiple linear regression model, uses least squares estimation to solve the unknown coefficients of the multiple linear regression equation, and obtains a significant linear relationship between smoking status and age and blood oxygen saturation, while gender, body mass index and blood There is no significant correlation between oxygen saturation, and qualitative and quantitative analysis are combined again, and stepwise regression is used to analyze the degree of influence of various factors on blood oxygen saturation. Finally, it is found that blood oxygen saturation is significantly negatively correlated with age. The method of goodness of fit validates the model.

Keywords

Blood Oxygen Saturation; Multiple Linear Regression; Least Square Estimation; Stepwise Regression.

1. Introduction

Oxygen saturation refers to the percentage of hemoglobin saturated with oxygen, that is, the ratio of hemoglobin's oxygen content to oxygen binding capacity multiplied by 100. Blood oxygen saturation indirectly reflects the size of blood oxygen partial pressure, and is an index to understand the degree of hemoglobin oxygen content and the buffering capacity of the hemoglobin system [1-3]. Blood oxygen saturation plays a very important role in research in the field of clinical medicine. Doctors will understand the patient's physical condition based on blood oxygen saturation. For example, pneumonia caused by a new type of coronavirus is also a respiratory infection. If the oxygen saturation is lower than 90%, timely diagnosis and treatment is required. It requires a series of tedious examinations such as doctors taking pictures. At the same time, nucleic acid testing is also required to rule out new coronavirus infections. Therefore, it is of great significance to study blood oxygen saturation measurement models in biological and medical field [2]. This article attempts to analyze the main influencing factors of blood oxygen saturation through modeling, and explore effective methods for measuring blood oxygen saturation.

2. Data Sources and Assumptions

In order to describe the pattern of oxygen saturation, we use experimental data to develop a model based on the test conducted on 36 subjects and the data recorded on the above subjects. We use these data to find a model of oxygen saturation change, using several parameters to characterize a person, including age, BMI, gender, smoking history and current smoking status,

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

and any medically important conditions that may affect reading. In order to ensure the rigor and reasonableness of the research, we make the following assumptions: (i) Assume that the information recorded by pulse oximetry is absolutely complete and accurate, and is not interfered by accidental factors; (ii) The collected data is representative Sex and usability; (iii) It is assumed that the four factors of age, BMI, gender, smoking history and current smoking status can better explain the changes in blood oxygen saturation; (iv) The errors caused by the recording equipment are negligible.

3. Multivariate model of blood oxygen saturation

3.1. Data Processing

According to the data provided by the subject, each person's data is monitored by the pulse oximeter, there is no correlation between the data of each subject, and each data point is independent of each other, there is no connection between them [3], because of the differences between the pulse oximeter sensor and chip, the amount of data obtained is not exactly the same, but all of them are within the gold standard (arterial blood sampling, using blood gas analysis equipment). The electrochemical analysis, which measures the partial pressure of oxygen and performs calculations to obtain an accurate oxygen saturation of blood, is within 1% of the error margin of the difference value and meets national standards.

In addition, the number of data provided is large enough that the mean value reflects the central position of the data as a whole; each subject is independent of each other at each data point, with no logical connectivity, and according to the descriptive statistical analysis, there are no outliers in the current data. It is suggested that data analysis can be performed directly using the mean values.

Therefore, we ventured to infer that the oxygen saturation data from the first six groups of subjects could be analyzed using the mean values of the data from each group for all 36 subjects [4].

Table 1. The statistical analysis of the first six groups.

Name	Sample size	Minimum	Maximum	Mean Standard	Deviation	Median
1	60	96.900	99.100	98.220	0.370	98.300
2	60	95.900	98.600	97.598	0.798	97.550
3	60	95.200	97.900	96.470	0.631	96.400
4	60	96.200	99.000	97.870	0.646	97.900
5	60	96.600	98.600	98.027	0.515	98.000
6	60	95.600	98.800	98.177	0.597	98.400

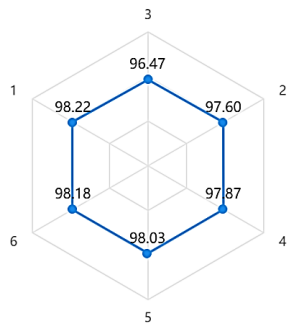

Figure 1. Radar plots of the first six mean values

3.2. Analysis of the Model

The problem was to obtain oxygen saturation data for 36 subjects using pulse oximetry at 1 HZ for approximately 1 h of continuous testing, giving each subject's age, BMI, gender, smoking

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

history and current smoking status, and any other important medical conditions that affect the test data. It is required to find a classical model of oxygen saturation changes using the given values so that the parameters can be used to represent the characteristics of an individual.

This question investigates the effect of several factors such as age, BMI, gender, smoking history and current smoking on oxygen saturation. A multiple regression model needs to be developed and the unknown parameters of the multiple linear regression equation solved using the least squares method in order to develop a complete model of oxygen saturation, which can also be used to characterize a person [5-6].

3.3. Model Establishing and Solving

Construct a matrix of independent variables $[x_1, x_2, x_3, x_4]$. Where x1 indicates gender, x2 indicates smoking history or current smoking status, x3 indicates BMI index, x4 indicates age, and y indicates oxygen saturation index.

$$\tilde{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 \quad (1)$$

It is a constant and a bias coefficient that reflects the degree of influence of the independent variable on the dependent variable oximetry.

For n sets of observations, the multiple linear regression Model is expressed as.

$$\begin{cases} y_1 = b_0 + b_1x_{11} + \cdots + b_4x_{14} \\ y_2 = b_0 + b_1x_{21} + \cdots + b_4x_{24} \\ \cdots \\ y_n = b_0 + b_1x_{n1} + \cdots + b_4x_{n4} \end{cases} \quad (2)$$

Among them,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, x = \begin{bmatrix} 1, x_{11}, x_{12} \ldots x_{14} \\ 1, x_{21}, x_{22} \ldots x_{24} \\ \cdots \\ 1, x_{n1}, x_{n2} \ldots x_{n4} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad (3)$$

Solve the partial regression matrix coefficient b using the least squares method. Substituting data from 36 individuals on Gender, Smoking Status, BMI, and Age as independent variables and OS as dependent variables into SPSS, where the categorical variables Gender and Smoking Status are assigned such that F=1, M=2, Ex-Smoker=1, Non-Smoker=2 and Smoker = 3. Multiple linear regression analysis was performed and the following results were obtained.

Table 2. Results of linear regression analysis. (a)

Variables	B	Standard Error	Beta
Constant	10010.50%	148.80%	-
Gender	-43.50%	34.70%	-18.10%
Smoking Status	-91.80%	24.80%	-50.80%
BMI	6.60%	7.10%	14.20%
Age	-3.80%	1.10%	-50.40%

Table 3. Results of linear regression analysis. (b)

Variables	t	p	VIF
Constant	-	-	-
Gender	6726.50%	0.00%	-
Smoking Status	-125.20%	22.00%	117.30%
BMI	-370.40%	0.10%	105.80%
Age	93.40%	35.80%	129.70%

Table 4. Results of linear regression analysis. (c)

R^2	Adjustment R^2	F
0.45	0.379	F(4,31)=6.335, p=0.001

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

From the above table, we can see that the median value of the model is 0.450, which means that Gender, Smoking Status,BMI,Age can explain the 45.0% variation in OS. The F-test of the model (F= 6.335, p= 0.001<0.05)), which means that at least one of Gender,Smoking Status,BMI,Age will have an impact on the OS, gives the following preliminary relationship [7].

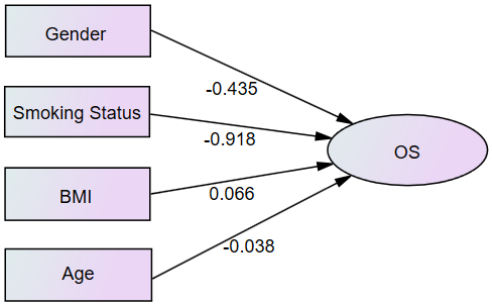

Figure 2. Model I model results graph

The regression coefficient of the independent variable Gender is -0.435 (t= -1.252, p= 0.220>0.05), implying that Gender does not have an impact on OS; similarly, the regression coefficient of the independent variable Smoking Status is -0.918 (t= -3.704, p= 0.001<0.01), implying that Smoking Status produces a significant negative impact relationship on OS; the regression coefficient value of the independent variable BMI is 0.066 (t= 0.934, p= 0.358>0.05), implying that BMI does not have an impact relationship on OS; the regression coefficient value of the independent variable Age is -0.038 (t= -3.580, p= 0.001< 0.01), implying that Age would have a significant negative relationship on OS.

Using SPSS software, a scatter plot was drawn to further argue that BMI has no significant relationship with OS.

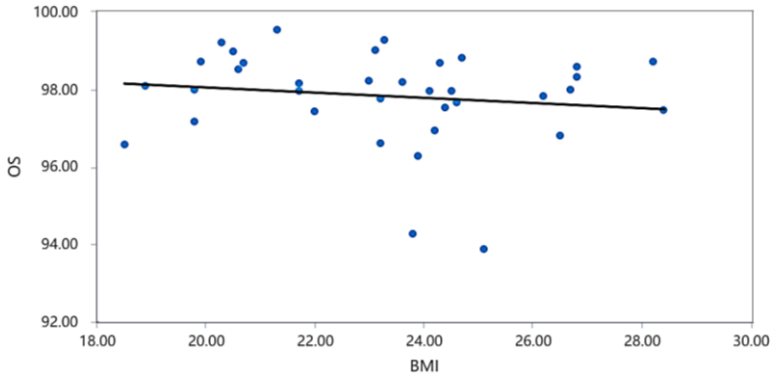

Figure 3. OS and BMI chart

Summary analysis showed that Smoking Status, Age had a significant negative effect on OS. However, Gender, BMI does not have an effect on the relationship.

The final linear regression model formula is obtained as follows.

$$y_{OS} = -0.918x_S - 0.038x_A + 100.105 \quad (4)$$

3.4. Analysis of Results

From the data in the table, we can see that the D-W value is near the number 2, thus indicating that there is no autocorrelation in the model, there is no correlation between the sample data, and the VIF values in the model are all less than 5, which mean that there is no co-collinearity problem. Therefore, it is indicated that the results of the linear regression are accurate and reliable.

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

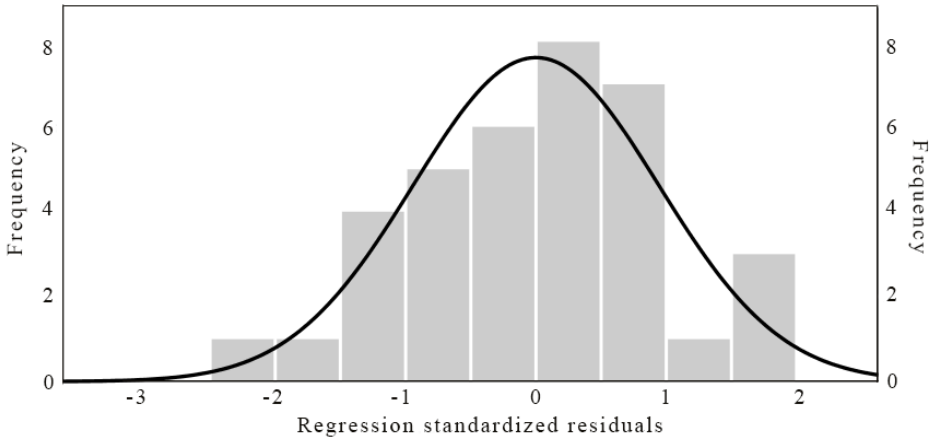

Figure 4. Normal distribution curve for OS

In the figure above, the black curve is normal, meaning that the residuals in this model follow a normal distribution. Therefore, it can be concluded that smoking status and age have a significant linear relationship with blood oxygen saturation, while gender, body mass index and blood oxygen saturation have no significant correlation [10].

4. Model of Blood Oxygen Saturation Based on Stepwise Regression

4.1. Analysis of the Model

The question requires an analysis of the relationship between age and the pattern of oxygen saturation as demonstrated in Question 1, i.e., what changes occur in the elderly compared to the young, and which are of biological or medical significance.

The method used in this topic is stepwise regression analysis, which combines qualitative and quantitative analysis to analyze the stepwise regression process of OS on the four factors, to evaluate and analyze the data, in order to draw conclusions. Finally, reliable information is collected to make reasonable prediction estimates of oxygen saturation in different age groups to make the prediction closer to the actual value.

4.2. Model Establishment

The four variables are first introduced one by one, x_G,x_S,x_B,x_A, and then eliminated when the original introduced independent variable becomes insignificant due to the introduction of subsequent independent variables, and F-tests are performed at each step to ensure that only significant variables are included in the regression equation before each new variable is introduced. It is repeated until only significant variables remain, and finally the optimal regression subset is obtained. Among them,the introduction of a significance level of a variable to less than the significance level rejected the argument, that is $\alpha_{entry} < \alpha_{removal}$.

Table 5. Results of stepwise regression analysis. (a)

Variables	B	Standard Error	Beta
Constant	10091.10%	66.00%	-
Smoking Status	-90.20%	24.20%	-49.90%
Age	-3.60%	1.00%	-47.90%

Table 6. Results of stepwise regression analysis. (b)

Variables	t	p	VIF
Constant	15293.00%	0.00%	-
Smoking Status	-372.30%	0.10%	101.70%
Age	-357.70%	0.10%	101.70%

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

Table 7. Results of stepwise regression analysis. (c)

R^2	Adjustment R^2	F
0.42	38.20%	F (2,33)=11.815, p=0.000

Using SPSS software, we performed stepwise regression analysis with Gender, Smoking Status, BMI, and Age as independent variables and OS as dependent variables. And Age is the two independent variables that explain 41.7% of the variation in the dependent variable OS. The abbreviated relationship pattern is shown in the following diagram.

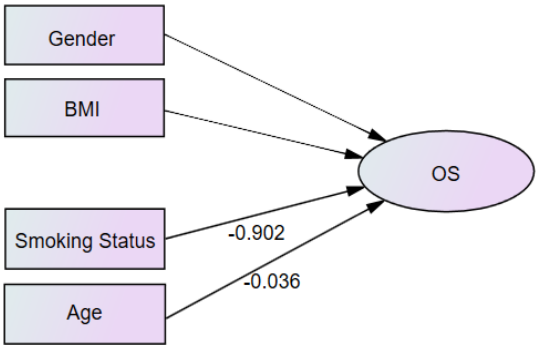

Figure 5. Model II model results graph

From the table, the value of the regression coefficient of the independent variable Smoking Status is -0.902 (t= -3.723, p= 0.001<0.01), implying that Smoking Status will have a significant negative impact on OS; similarly, the value of the regression coefficient of the independent variable Age is -0.036 (t= -3.577, p= 0.001 < 0.01), implying that Age will also have a significant negative effect on OS.

Summarizing the analysis, we can see that Smoking Status, Age will have a significant negative effect on OS. Model formula:

$$y_{OS} = -0.902x_S - 0.036x_A + 100.911 \quad (5)$$

4.3. Analysis of Results

The model passed the F test (F= 11.815, p= 0.000<0.05), indicating that the model is valid; in addition, the multicollinearity of the model was tested and found that all the VIF values in the model are less than 5, meaning that there is no co-collinearity problem; and the D-W value is near the number 2, thus indicating that the model does not have autocorrelation, there is no correlation between the sample data, the model is of great practical significance. In addition, the regression coefficient is -0.039, when all other influences are constant, the oxygen saturation x_A decreases by 0.039% for each year of age and increases by 0.039% for each year of age. This means that, under the same conditions, oxygen saturation is lower in the elderly compared to the young.

5. Conclusion

In order to study the mechanism of age, BMI, gender, smoking history and current smoking status affecting blood oxygen saturation, this article first establishes a multiple linear regression model and uses least squares estimation to solve the unknown coefficients of the multiple linear regression equation to obtain various factors On the influence of blood oxygen saturation, it is concluded that age and current smoking status have a greater influence on blood oxygen saturation. Then, stepwise regression analysis is used to combine qualitative and quantitative analysis to analyze the gradual effect of OS on these four factors. In the regression process, the data is evaluated and analyzed, so as to draw the conclusion that age and blood

ISSN: 2692-7608 DOI: 10.6981/FEM.202108_2(8).0005

oxygen saturation are significantly negatively correlated on the basis of other conditions. Based on the data and experimental results provided in the title, as well as our own review of relevant information, we have overcome the subject data, too few influencing factors and other problems, and established a typical blood oxygen saturation model. This model has certain typicality and accuracy, and has certain practical value for the medical practice of blood sample monitoring.

References

[1] Hasan Md. Impact of high dose of baricitinib in severe COVID-19 pneumonia: a prospective cohort study in Bangladesh[J]. BMC Infectious Diseases, 2021,21(1).

[2] S A Niu, S M Duan, Y K Sun. Clinical application status and progress of pulse oximetry monitoring technology [J]. Chinese Medical Journal,2020,55(06):585-586.

[3] J Xu, C Wang.Wireless wearable blood oxygen saturation measurement system[J].Communication World, 2019, 26(06):260-261.

[4] Qian Guangsong.The current status and existing problems of domestic oximetry detection [J]. Metrics and Testing Technology, 2010,37(09):19-20.

[5] Y Xu, J Li, X Chen, J X Wei. Brief analysis of pulse oximeter industry standards[J]. China Medical Equipment, 2017, 32(05): 68-72.

[6] B Liu, X L Liu, L Y Wang. An Unconventional Proof of the Second Integral Mean Value Theorem[J]. Progress in Applied Mathematics, 2020, 09(11).

[7] D D Sun. Selection of the Linear Regression Model According to the Parameter Estimation[J]. Wuhan University Journal of Natural Sciences, 2000, 5(4):400-405.

[8] C R Qiu, Y D Jiang. Design of a production configuration parameter modeling system based on the least square method [J]. Modern Electronic Technology, 2021, 44(04): 83-87.

[9] Hosmane B S. Improved likelihood ratio tests and pearson chi-square tests for independence in two dimensional contingency tables[J]. Communications in Statistics, 1986, 15(6):1875-1888.

[10] Y P Du, J Shu, X B Hou. Prediction model of tobacco moisture in the whole process before drying based on normal distribution statistical analysis [J]. Industrial Technology Innovation, 2021, 08(02): 119-124.