ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

Design of Variability of Oxygen Saturation Based on BP-ANOVA Combined Algorithm

Zelong Xiang[1,a], Siyao Li[2,b], Yangchenmiao Cui[3,c], Xiyuan Su[4,d], Mingze Chen[5,e]

[1]School of computer science and engineering, Wuhan Institute of Technology, Wuhan 431700, China;

[2]School of economics and management, Northwest A&F University, Yangling, 712100, China;

[3]School of Statistics, Xi'an University of Finance and Economics, Xi'an 710100, China;

[4]Shoolof arts and sciences, Shaanxi University of Science & Technology, Xi'an 710021, China;

[5]Department of computer and communication, Lanzhou University of Technology, Lanzhou 730050, China.

[a]1792300986@qq.com; [b]1664827654@qq.com; [c]m13152118218@163.com; [d]tfasxy@163.com; [e]2398639492@qq.com

Abstract

This paper mainly studies the variation of oxygen saturation. The mathematical models such as data envelopment analysis (DEA), BP neural network (BP neural network) and analysis of variance (ANOVA) were established to solve the problems of establishing the change mode of oxygen saturation and the relationship between age and oxygen saturation. Calculate the percentage of oxygen saturation between 95% - 97%; calculate the deviation between the average value of oxygen saturation and the normal value (95% - 97%), that is, if the average value is within the normal range, then assign a value of "1"; if the average value is not within the normal range, then use Mahalanobis distance discrimination mathematical model to calculate, and then take the number; first, calculate the standard deviation, so as to obtain the reciprocal of coefficient of variation. Data envelopment analysis (DEA) mathematical model was established to evaluate the oxygen saturation of 36 people, and the comprehensive evaluation value of oxygen saturation was obtained. Taking the comprehensive evaluation value of oxygen saturation, age, BMI, gender and other data of these 36 people as the original input data, the BP neural network mathematical model is established and the function is constructed. The fitting degree of these factors and oxygen saturation can be known, and the change mode of oxygen saturation can be known. Through MATLAB programming calculation, we can use age, BMI, smoking history or current smoking status to characterize a person.

Keywords

Data Envelopment Analysis; BP Neural Network; ANOVA; Correspondence Analysis.

1. Introduction

Blood oxygen saturation (SaO2) is the percentage of the capacity of oxygenated hemoglobin (HBO2) in the blood to the capacity of total hemoglobin (HB), that is, the concentration of blood oxygen in the blood. It is an important physiological parameter of respiration and circulation. Functional oxygen saturation is the ratio of HBO2 concentration to HBO2 + Hb concentration, which is different from

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

the percentage of oxygenated hemoglobin. Therefore, monitoring arterial oxygen saturation (SaO2) can be used to estimate lung oxygenation and hemoglobin oxygen carrying capacity. The oxygen saturation of arterial blood is 98% and that of venous blood is 75%.

The traditional method of measuring blood oxygen saturation is to take blood from human body first, then use the blood gas analyzer for electrochemical analysis, measure the partial pressure of oxygen PO2, and calculate the oxygen saturation of bleeding. This method is troublesome and can not be used for continuous monitoring.

Now we use the fingertip photoelectric sensor. When measuring, we just need to put the sensor on the human finger, use the finger as a transparent container for hemoglobin, and use the red light with wavelength of 660 nm and 940 nm Nm near-infrared light is used as the light source to measure the light transmission intensity through the tissue bed to calculate the hemoglobin concentration and blood oxygen saturation. The instrument can display the human blood oxygen saturation, which provides a continuous and non-invasive blood oxygen measurement instrument for clinical.

2. Analysis of question

Through Baidu, Google and other search engines, query shows that oxygen saturation between 95% and 97% is the most appropriate, that is, from this indicator alone, this person is the most healthy. According to the oxygen saturation data of 36 individuals, calculate the percentage of oxygen saturation between 95% and 97% for each person; calculate the absolute value of the deviation value between the average value of oxygen saturation and the normal value of 95% - 97% (if the average value is within the normal range, assign "1" if the average value is within the normal range; if not, use the Mahalanobis distance discriminant mathematical model to calculate the comprehensive deviation value, and then take the number)The reciprocal of coefficient of variation. The average value can represent the oxygen saturation of a person over a long period of time; the coefficient of variation can indicate the stability of a person's oxygen saturation over a long period of time, that is, the greater the reciprocal of the coefficient of variation is, the more stable the oxygen saturation of the person is. Finally, data envelopment analysis (DEA) mathematical model was established to evaluate the oxygen saturation of 36 individuals. This value can represent the comprehensive situation of a person's oxygen saturation. The higher the comprehensive evaluation value is, the better the condition of the person's oxygen saturation is. Then, the evaluation value of oxygen saturation and age, BMI, gender, smoking history and current smoking status of these 36 people were taken as the original input data, and the BP neural network mathematical model was established, and the function was constructed to know the influence of these factors on oxygen saturation, and the most influential factor on oxygen saturation was found out.

3. Establishment of model

Excel was used to calculate the percentage of oxygen saturation between 95% and 97%.

3.1 Calculation of average oxygen saturation:

$$\overline{x}_i = \frac{1}{n}(x_{ij}, x_{ij}, \cdots, x_{ij}), i = 1,2,\cdots,36; j = 1,2,\cdots,n$$

Firstly, judge whether it is in the normal range. If it is in the normal range, it is assigned as "1"; if not, it is determined by Mahalanobis distance discrimination.

3.2 The mathematical model of Mahalanobis distance discrimination is constructed:

This paper mainly analyzes the situation of two populations. The problem of multi population discrimination can be transformed into the discriminant analysis of two populations. Consider n p-dimensional samples from population G_1, G_2.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

$$X^{(i)}=\begin{bmatrix} x_{11}^{(i)} & x_{12}^{(i)} & \cdots & x_{1p}^{(i)} \\ x_{21}^{(i)} & x_{22}^{(i)} & \cdots & x_{2p}^{(i)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n_i1}^{(i)} & x_{n_i2}^{(i)} & \cdots & x_{n_ip}^{(i)} \end{bmatrix}^T = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_{n_i}^{(i)} \end{bmatrix} (i=1,2)$$

Where $n_i(i=1,2)$ is the sample number of the i population, $n=n_1+n_2$. Before introducing the projection method, we define several essential parameters.

Sample mean vector of population G_i: $\bar{x}^{(i)}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_j^{(i)}$;

Total sample mean vector: $\bar{x}=\frac{n_1}{n}\bar{x}^{(1)}+\frac{n_2}{n}\bar{x}^{(2)}$;

Inter group dispersion matrix: $S_b=\sum_{i=1}^{2}\frac{n_i}{n}(\bar{x}^{(i)}-\bar{x})(\bar{x}^{(i)}-\bar{x})^T$;

Sample total intra group dispersion matrix: $S_w=\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_j^{(i)}-\bar{x}^{(i)})(x_j^{(i)}-\bar{x}^{(i)})^T$;

Inter group dispersion matrix of population G_2 "relative" to population G_1:

$$B=\sum_{j=1}^{n_2}(x_j^{(2)}-\bar{x}^{(1)})(x_j^{(2)}-\bar{x}^{(1)})^T;$$

Intra group dispersion matrix of population G_1:$E_1=\sum_{j=1}^{n_1}(x_j^{(1)}-\bar{x}^{(1)})(x_j^{(1)}-\bar{x}^{(1)})^T$;

Intra group dispersion matrix of population G_2:

$$E_2=\sum_{j=1}^{n_2}(x_j^{(2)}-\bar{x}^{(2)})(x_j^{(2)}-\bar{x}^{(2)})^T$$

3.3 Fisher Projection:

In the same way, the p -dimension is projected to one dimension, but we only want the difference between groups to be as large as possible after projection, without considering the intra group difference (this is mainly to avoid the singularity of within group deviation matrix S_w of sample matrix in p -dimensional space), that is, the projection vector w_2 is:

$$w_2=arg\max_{w^Tw=1} w^TS_bw,$$

Obviously, w_2 is the eigenvector corresponding to the largest eigenvalue of S_b. The criterion of projection is consistent with that of Fisher projection, that is, for any given sample x, if

$$\left|w_1^Tx-w_1^T\bar{x}^{(t)}\right|=\min_{i=1,2}\left|w_1^Tx-w_1^T\bar{x}^{(i)}\right|$$

Then judge the sample $x\in G_t$.

3.4 Calculation of coefficient of variation:

The coefficient of variation can represent the stability of a person's oxygen saturation over a long period of time, that is, the greater the reciprocal of the coefficient of variation is, the more stable the oxygen saturation of a person is.

$$s_i=\sum_{i=1}^{n}\sqrt{\frac{1}{n}(\overline{x}_i-x_{ij})^2}, i=1,2,\cdots,36; j=1,2,\cdots,n$$

$$c_i=\frac{s_i}{\overline{x}_i}, i=1,2,\cdots,36$$

3.5 Establishment of mathematical model of data envelopment analysis:

There are n DMUs, each DMU has m inputs and s outputs. Let $x_{ij}(i=1,\cdots,m,j=1,\cdots,n)$ denote the i -th input of the j -dmu, $y_{rj}(r=1,\cdots,s,j=1,\cdots,n)$ represents the r -th output of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

j -dmu, $v_i(i=1,\cdots,m)$ represents the weight of the i -th input, and $u_r(r=1,\cdots,s)$ represents the weight of the m-output of the r -dmu.

The vectors $X_j, Y_j(j=1,\cdots,n)$ represent the input and output vectors of DMU j respectively, and v *and* u represent the input and output weight vectors respectively

$$X_j=(x_{1j},x_{2j},\cdots,x_{mj})^T, Y_j=(y_{1j},y_{2j},\cdots,y_{sj})^T, u=(u_1,u_2,\cdots,u_m)^T, v=(v_1,v_2,\cdots,v_s)^T$$

The efficiency evaluation index of DMU j is defined as

$$h_j=(u^TY_j)/(v^TX_j), j=1,2,\cdots,n$$

The mathematical model for evaluating j_0 efficiency of DMU is

$$max\frac{u^TY_{j0}}{v^TX_{j0}}$$

$$s.t\begin{cases}\dfrac{u^TY_j}{v^TX_j}\le 1, j=1,2,\cdots,n,\\ u\ge 0, v\ge 0, u\ne 0, v\ne 0。\end{cases}$$

Through Charnes Cooper transformation:

$$w=tv, U=ut, t=\frac{1}{v^TX_{j0}},$$

The model can be transformed into an equivalent linear programming problem

$$maxV_{j0}=U^TY_{j0},$$

$$s.t\begin{cases}w^TX_j-U^TY_j\ge 0, j=1,2,\cdots,n,\\ w^TX_{j0}=1,\\ w\ge 0, U\ge 0。\end{cases}$$

It can be proved that the two models are equivalent. The dual programming model of linear programming has clear economic significance. The dual form is as follows:

$$min\,\theta$$

$$s.t\begin{cases}\sum_{j=1}^{n}\lambda_jX_j\le\theta X_{j0},\\ \sum_{j=1}^{n}\lambda_jY_j\ge Y_{j0,}\\ \lambda_j\ge 0, j=1,2,\cdots,n。\end{cases}$$

If: $V_{j0}=1$ is defined, DMU j_0 is said to be weakly DEA efficient.

If there is an optimal solution $w^*>0, u^*>0,$ and it has the most objective value $V_{j0}=1$, then the DMU j_0 is said to be DEA efficient.

4. Model solution

Search data show that the oxygen saturation between 95% and 97% is the most suitable for human body. The oxygen saturation data of 36 people are processed, and the proportion of each person's oxygen saturation data in the normal range to each person's total oxygen saturation data is calculated by using Excel's function function function, and the average value of each person's oxygen saturation data is calculated. The average value can represent a person in a long period of time If the average value is between 95% and 97% of the normal range, the coefficient of variation is 1; otherwise, the comprehensive deviation value is calculated by means of Mahalanobis distance differential

ISSN: 2414-1895
DOI: 10.6919/ICJE.202105_7(5).0006

mathematical model, and the reciprocal and the reciprocal of coefficient of variation are taken. Finally, take the "oxygen saturation" as the input and output indicators of the "lingo" as the input and output indicators of the "lingo"; finally, use the data envelopment value of "lingo" as the input and output indicators of the software Grid data. This data can be used to represent a person's oxygen saturation.

Finally, the BP neural network mathematical model is established, and the 36 individuals' comprehensive index value, age, BMI, gender, smoking history or current smoking status are taken as the original input data. According to the output results, the influence of each index on oxygen saturation is assigned, and the influence of each index on oxygen saturation can be directly observed. And calculate the percentage of the impact of each index in the total index, and the results are as follows:

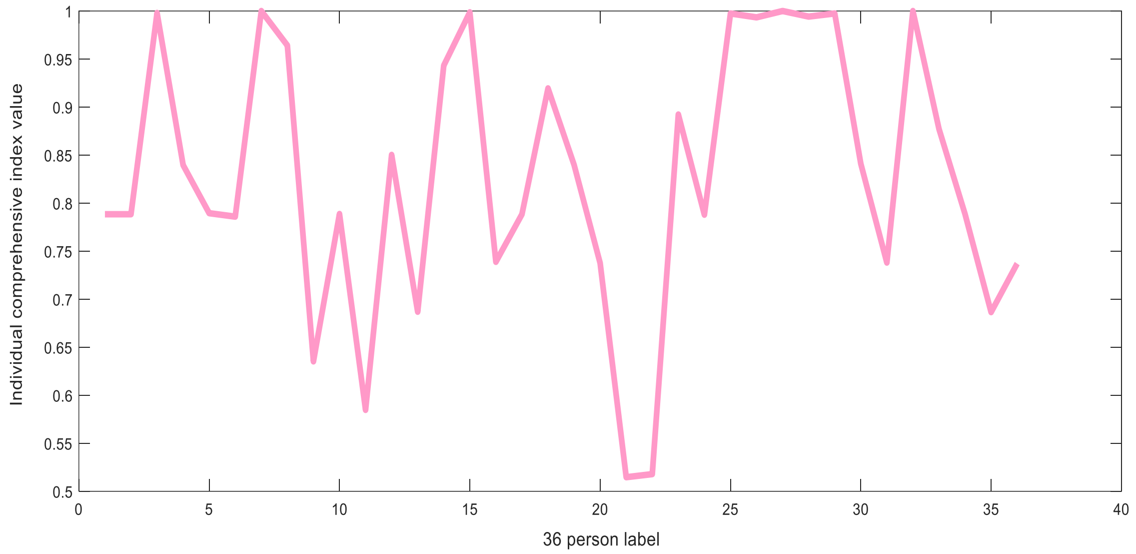

Figure 1. Comprehensive index value corresponding to 36 individuals

By observing the comprehensive index value, one can directly see the oxygen saturation of a person. The closer the comprehensive index value is to 1, the more normal the oxygen saturation is. Other people's oxygen saturation is based on the corresponding comprehensive index value.

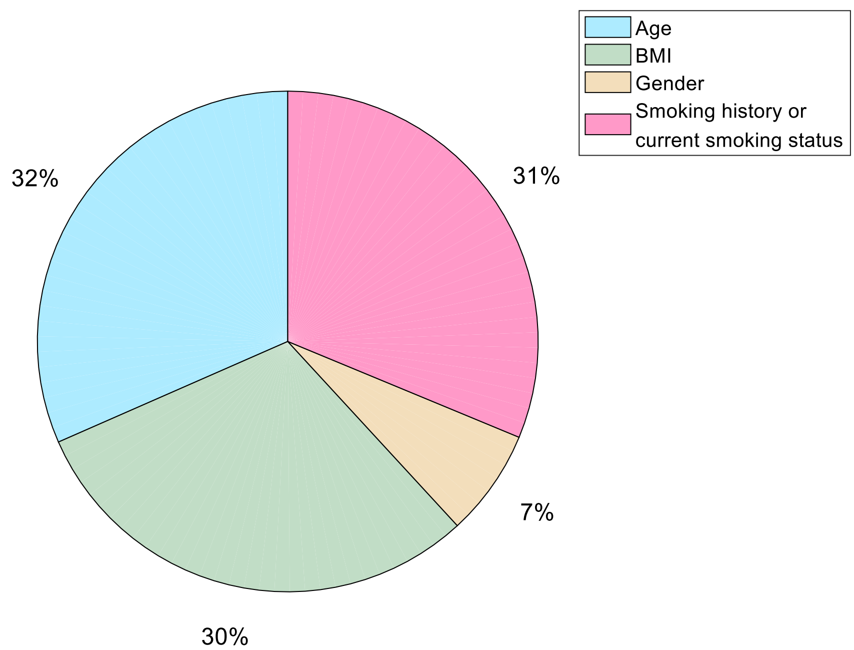

Figure 2. Percentage of influence on oxygen saturation in four index values

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

It can be seen that age, BMI, smoking history or current smoking status have great influence on oxygen saturation, among which age has the greatest influence, while gender has the least influence. This is more in line with the actual situation. According to the data, personal ventilation function, blood transport efficiency and various conditions of bronchi, lungs and respiratory tract or diseases have great influence on oxygen saturation. With the growth of age, these functions of the elderly are generally worse than those of the young; obesity or emaciation affect BMI, and physical conditions also have a great impact on the above functions; and smoking or damage to the bronchi and lungs. Therefore, in fact, these three indicators really affect oxygen saturation. Therefore, the accuracy of the results is high.

5. Conclusion

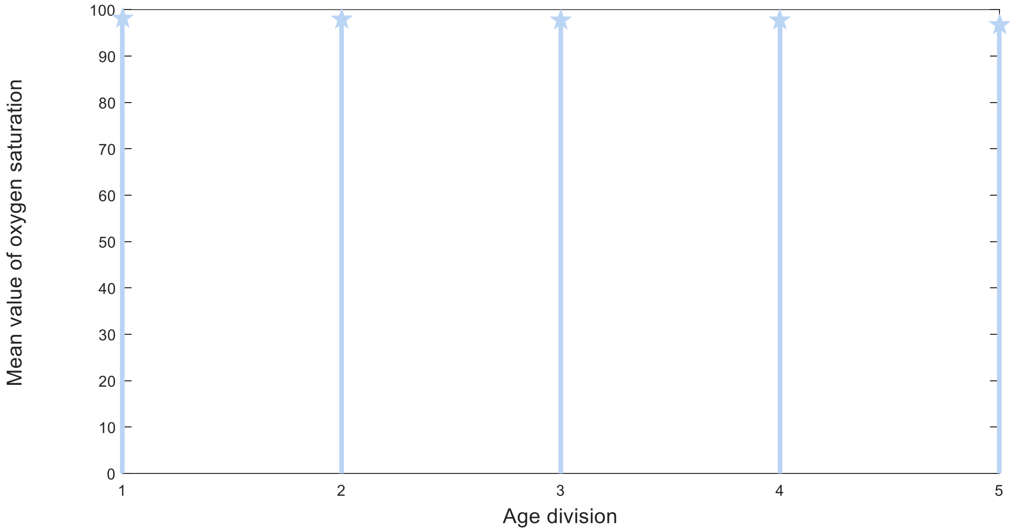

Figure 3. Mean value of oxygen saturation

It can be seen from the Figure that with the increase of age, the average value of oxygen saturation is decreasing, which indicates that age has an impact on oxygen saturation. The older the age, the lower the oxygen saturation is. The younger the oxygen saturation is, the higher the oxygen saturation is within the normal range. It shows that the younger you are, the healthier you are.

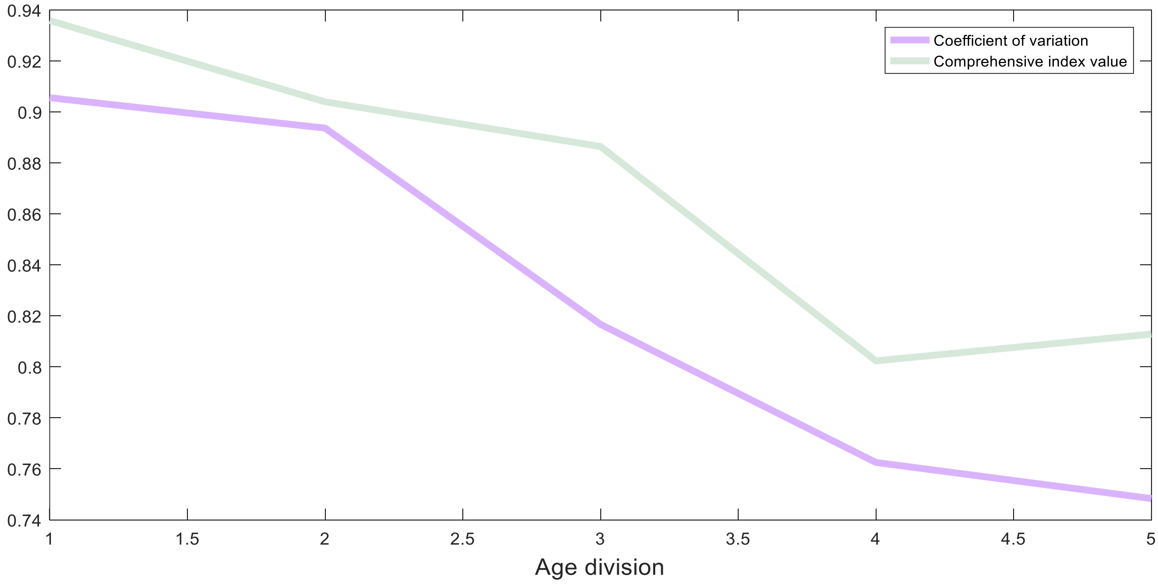

Figure 4. Coefficient of variation and comprehensive index values of five age groups

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0006

The coefficient of variation also showed a downward trend, indicating that the older the age, the worse the stability of oxygen saturation; the maximum comprehensive index value is around 20 years old, which indicates that the comprehensive state of oxygen saturation value is the most healthy during this period of age, and generally speaking, the health status of oxygen saturation index is inversely related to age. Therefore, the pattern of oxygen saturation sequence is related to age and has inverse correlation. The older the age, the less healthy.

Comparing the data of the elderly with other people at different stages, it is found that the average value of oxygen saturation is lower than that of other stages, and the coefficient of variation is also smaller, which indicates that the more unstable the oxygen saturation is, the smaller the comprehensive index value is, and the health degree is less than that of other age stages. The larger the age difference is, the greater the difference is, which indicates that the function related to oxygen saturation of the elderly increases with age Degradation occurs.

References

[1] Peixoto Maria Gabriela Mendon ça, Musetti Marcel Andreotti, Mendonça Maria Cristina Angélico de. Performance management in hospital organizations from the perspective of Principal Component Analysis and Data Envelopment Analysis: the case of Federal University Hospitals in Brazil[J].Computers & Industrial Engineering,2020,150(prepublish):.

[2] Pan Wen-Tsao, Zhuang Mei-Er, Zhou Ying-Ying, Yang Jia-Jia. Research on sustainable development and efficiency of China's E-Agriculture based on a data envelopment analysis-Malmquist model[J]. Technological Forecasting & Social Change,2021, 162:.

[3] Feng Yong-qiang, Liu Yu-Zhuang, Wang Xin, He Zhi -Xia, Hung Tzu-Chen, Wang Qian, Xi Huan. Performance prediction and optimization of an organic Rankine cycle (ORC) for waste heat recovery using back propagation neural network[J]. Energy Conversion and Management,2020, 226:.

[4] Utsav Pandey, Sanjeet Singh. Data envelopment analysis in hierarchical category structure with fuzzy boundaries[J]. Annals of Operations Research,2020, (prepublish):