International Journal of Frontiers in Sociology
ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

Study on variation of oxygen saturation based on cluster and multiple regression analysis

Yan Liang[1,*], Jintao Su[1], Xutong Ran[2]

[1] *Institute of Finance, Anhui University of Finance and Economics, Bengbu 233030, China*
[2] *School of international trade and economics, Anhui University of Finance and Economics, Bengbu 233030,China*
[*] *Corresponding Author*

ABSTRACT. *This paper mainly uses SPSS software to construct clustering and multivariate regression analysis model to study the variation of blood oxygen saturation from five indexes of sex, age, smoking, body mass index and blood oxygen saturation. The variation law of blood oxygen saturation and the use of indicators to describe a person's model method are obtained. Then the biological and medical significance of the characteristics of blood oxygen saturation change is analyzed.*

KEYWORDS: *oxygen saturation, cluster analysis, multiple regression*

1. Introduction

Pulse oxygen saturation is a routine measure to monitor the level of blood oxygen saturation in patients.A cluster analysis and multiple regression analysis model are constructed by SPSS software. The variation of blood oxygen saturation and how to use the index to describe the two problems of one person are analyzed from five indexes: sex, age, smoking, body mass index and blood oxygen saturation.

2. Overview of the status of research

For the study of oxygen saturation, many experts and scholars at home and abroad have carried out a number of studies, here is a brief summary of the existing research conclusions. Han Shuai [1] believes that oxygen saturation can reflect the health of human respiratory system and cardiovascular system and plays an important role in the prevention and diagnosis of human diseases. At present, the measurement of oxygen saturation and other components of blood by photoelectric pulse wave detection has become one of the hot spots in the field of medicine, which is of great significance. Wang Xiaofei[2] believes that the noninvasive detection of blood oxygen saturation in clinic is mainly based on the principle of double

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

wavelength pulse oxygen saturation measurement, but its detection accuracy still needs to be further explored and improved. Many researchers at home and abroad use three-wavelength or even eight-wavelength methods to measure oxygen saturation to some extent reduce the error. You Chenyu[3] believes that the oxygen saturation of human tissues refers to the comprehensive value of oxygen saturation of blood in various microvessels in local tissues to reflect the oxygenation of local tissues. Non-destructive, continuous and rapid monitoring of oxygen saturation in human tissues is of great clinical significance. Downing [4] believes that enough oxygen is the material basis of all life activities, and oxygen saturation is an important parameter to reflect the oxygen content in the blood. Timely detection of oxygen content in the blood is essential to maintain life and determine whether there are disorders in the respiratory system and circulatory system.

3. Establishment and Solution of Model

3.1 Establishment and solution of cluster analysis model

3.1.1 Establishment of cluster analysis model

Firstly, the blood oxygen saturation level of 36 individuals was classified by cluster analysis. According to the distance result of each cluster member in the cluster analysis ,36 samples were divided into 3 categories, and the number of individuals in each cluster result was obtained.The cluster membership table and the number of cases in each cluster are shown in Table 1, Table 2.

Table 1 Cluster membership table

Case number	Cluster clustering	Distance	Case number	Cluster clustering	Distance	Case number	Cluster clustering	Distance
1	1	33.639	13	1	33.844	25	3	40.281
2	3	75.098	14	3	48.087	26	3	41.623
3	3	53.181	15	1	35.472	27	3	67.661
4	1	58.774	16	1	43.682	28	3	42.082
5	1	48.312	17	1	47.246	29	3	47.175
6	1	48.649	18	2	53.169	30	1	52.132
7	3	72.016	19	1	44.206	31	1	37.257
8	1	31.720	20	1	32.738	32	3	73.606
9	1	45.334	21	1	26.353	33	2	53.169
10	1	44.759	22	1	33.598	34	1	36.040
11	1	64.285	23	3	68.749	35	1	30.074
12	1	67.063	24	1	49.314	36	1	23.798

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

Table 2 Number of cases per cluster

Cluster clustering	1	23.000
	2	2.000
	3	11.000

The results show that the number of individuals in cluster 1 is 23, the number of individuals in cluster 2 is 2, and the number of individuals in cluster 3 is 11. All 36 individuals are valid individuals, indicating that there is a missing value in the value of no individual.

3.1.2 Analysis description of different clusters

According to the results of the three clustering analysis, the changes of oxygen saturation of individuals corresponding to different clustering results were classified and averaged. The time series changes of oxygen saturation of individuals in different clustering results were obtained by SPSS analysis.

Through the analysis of cluster 1, the fitting degree results of cluster 1, the statistical tables of cluster model and the time series analysis results of one hour oxygen saturation change of 23 individuals are obtained. The results are as follows.

Table 3 Clustering 1 model fit degree

Fitting statistics	Average value	Standard error	Minimum value	Maximum value	percentile						
					5	10	25	50	75	90	95
R smooth	0.044	.	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
R party	0.867	.	0.867	0.867	0.867	0.867	0.867	0.867	0.867	0.867	0.867
RMSE	0.044	.	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
MAPE	0.032	.	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
MaxAPE	0.389	.	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.389	0.389
MAE	0.032	.	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
MaxAE	0.382	.	0.382	0.382	0.382	0.382	0.382	0.382	0.382	0.382	0.382
BIC of normality	-6.262	.	-6.262	-6.262	-6.262	-6.262	-6.262	-6.262	-6.262	-6.262	-6.262

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

Table 4 Cluster 1 Model Statistics

Model	Number of predicted variables	Model fit statistics	Q(18) Yang-Box			Number of outliers
		R smooth	Statistics	DF	Significant significance	
@ Cluster 1 Mean - Model _1	0	0.044	26.287	15	0.035	0

From tables 3 and 4, it can be seen that the R of cluster 1 is higher, reaching 0.867, indicating that the independent variable can explain 86.7% of the population. The standard error of the model is zero, which indicates that the error deviation of the cluster analysis results is small.A stable R of model fit statistics is 0.44, which indicates that the overall fit of the model is better. The significance of the model is 0.035, which indicates that the model has passed the significance test.

Blood oxygen saturation of individuals in cluster 1 was then measured, and the blood oxygen saturation value of about 1 hour was continuously tested at 1 hz frequency. In order to simplify the analysis of time series model, we take the average value of the change value of blood oxygen saturation and integrate the analysis results, as shown in figure 1.

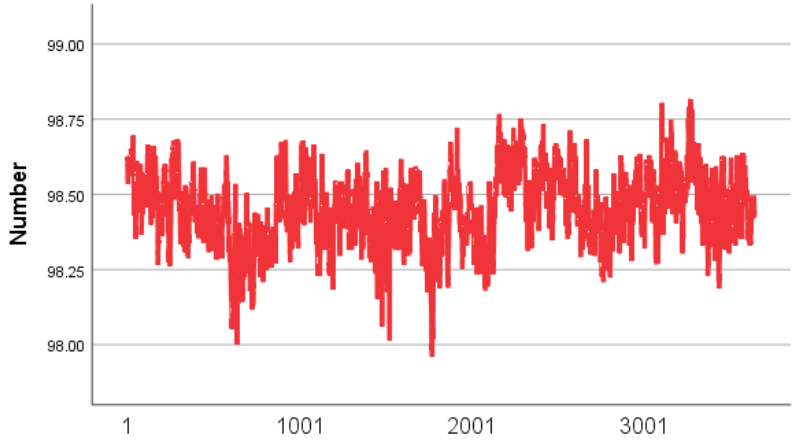

Figure. 1 Time series analysis of poly 1 oxygen saturation

Figure 1 shows that the level of oxygen saturation in cluster 1 is 98-98.75, the overall level of oxygen saturation is higher, and the level of oxygen saturation is at the level of normal people.The individuals in cluster 1 may be affected by age, sex, smoking and body mass index.

In the same way, we can get the analysis process of cluster 2 and cluster 3.

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

By analyzing cluster 2, the R of cluster 2 is higher, reaching 0.950, the standard error of the model is zero, and the stable R of the model fitting degree statistics is 0.072, which indicates that the overall fitting degree is better. The significance of the model was 0.036, which passed the significance test. The results of time series analysis showed that the level of oxygen saturation in cluster 2 was 91-96, and the level of oxygen saturation in some time testers was lower than 94%, even close to 91%. The results showed that the level of oxygen saturation in cluster 2 was lower than that in normal people.

By analyzing cluster 3, the R of cluster 3 is higher, At 0.939, The standard error of the model is zero, R of model fit statistics is 0.060, It shows that the overall fitting degree is good. The significance of the model is 0.0375, Passed the significance test. The results of time series analysis showed that the level of oxygen saturation in cluster 3 was 95.5 to 97.5, For most of the time, The blood oxygen saturation was between 96 and 97 percent, Only a few times did the testers have a 95 to 96 percent oxygen saturation, indicating that the level of oxygen saturation in the individuals in cluster 3 is slightly higher than the normal level, It can be determined that the tester's human oxygen supply is normal.

3.2 Establishment and solution of multiple regression models

The multivariate regression analysis model was established by taking the total index as dependent variable, age, sex, body mass index, smoking condition and blood oxygen saturation as independent variables, and then the model was described by using the index. The process of model solving and analysis is as follows.

Before establishing the multivariate regression analysis model, the correlation analysis of the six indexes is carried out to determine the correlation coefficient between the indexes, so as to facilitate the establishment of the multivariate regression model. The analysis results of the correlation analysis are as follows.

Table 5 Correlation analysis

		Total indicators	Gender	Smoking Status	BMI	Age	Oxygen saturation
Pearson Correlation	ZBB	1.000	0.105	0.240	-0.336	-0.995	0.493
	Gender	0.105	1.000	0.073	-0.413	-0.084	0.185
	Smoking Status	0.240	0.073	1.000	-0.140	-0.180	0.330
	BMI	-0.336	-0.413	-0.140	1.000	0.304	-0.188
	Age	-0.995	-0.084	-0.180	0.304	1.000	-0.414
	SpO2	0.493	0.185	0.330	-0.188	-0.414	1.000

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

The results of correlation analysis showed that there was a positive correlation between total index and sex smoking and oxygen saturation, while there was a negative correlation between body mass index and age and total index.At the same time, the absolute value of the correlation coefficient between age and oxygen saturation and the total index is larger, indicating that there is a strong correlation between the two indexes and the total index, while the absolute value of the correlation coefficient between sex, smoking and body mass index and the total index is smaller, indicating that the correlation between the index and the total index is small.

Table 6 Correlation coefficient tables

Model				Standardized coefficients	t	Significant significance	B 95.0% confidence intervals		Correlation		
		B of unstandardized coefficients	Standard error	Beta			Lower limit	Upper limit	Zero	Partial	Part I
1	Constants	2.047E-7	0.000		0.002	0.998	0.000	0.000			
	Gender	-2.730	0.000	-0.004	-1160710.317	0.000	-2.730	-2.730	0.105	-1.000	-0.004
	Smoking Status	19.506	0.000	0.038	10691756.738	0.000	19.506	19.506	0.240	1.000	0.036
	BMI	-3.334	0.000	-0.029	-7443500.095	0.000	-3.334	-3.334	-0.336	-1.000	-0.025
	Age	-18.743	0.000	-0.945	-246861259.296	0.000	-18.743	-18.743	-0.995	-1.000	-0.827
	SpO2	21.760	0.000	0.085	21836463.253	0.000	21.760	21.760	0.493	1.000	0.073

Then the coefficient parameter table of multivariate regression model is obtained by multivariate regression analysis of six indexes.We assume that the five independent variables of the multivariate regression model: sex, smoking, body mass index, age and oxygen saturation are replaced by X 1, X2, X3, X4, X5, and the dependent variables are replaced by Y, and the expression of the multivariate regression model is as follows:

$$Y = 2.047E\text{-}7 - 2.730X1 + 19.506X2 - 3.334X3 - 18.743X4 + 21.760X5$$

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

Table 7 Collinear diagnosis

Model	Dimensions	eigenvalues	Conditional indicators	Proportion of variance					
				(Constants)	Gender	Smoking Status	BMI	Age	Oxygen saturation
1	1	5.014	1.000	0.00	0.01	0.01	0.00	0.00	0.00
	2	0.444	3.362	0.00	0.65	0.03	0.00	0.04	0.00
	3	0.419	3.459	0.00	0.10	0.74	0.00	0.02	0.00
	4	0.116	6.567	0.00	0.03	0.13	0.01	0.76	0.00
	5	0.007	27.161	0.00	0.20	0.02	0.99	0.05	0.00
	6	5.844E-5	292.914	1.00	0.02	0.07	0.00	0.13	1.00

Then the multiple collinearity analysis of each independent variable is carried out. The linear correlation between independent variables from Table 7 is low and the collinearity between independent variables is weak. Therefore, the model does not have multiple collinearity.

Finally, the multivariate regression model has passed the significance test and the multiplex collinearity test, which shows that the model has certain analytical ability and prediction ability. Through the multivariate regression model, we can describe a person's health level better by inserting several independent variables in the model.

4. Biological and Medical Significance Analysis of Oxygen Saturation

From the results of model analysis, it can be seen that by measuring the value of oxygen saturation, people's health level can be analyzed to a certain extent, and then it can bring some help and significance to biological research and medical research. Therefore, we will analyze the relationship between oxygen saturation and these five indexes in biology and medicine.

First of all, people's smoking situation can be divided into three categories: non-smoking, smokers and smoking history. The blood oxygen saturation level of non-smoking population is the highest, which indicates that non-smoking is helpful to improve blood oxygen saturation. For smokers and those with a history of smoking, the oxygen saturation of smokers was significantly lower than that of those with a history of smoking, indicating that continuous smoking would further reduce the oxygen saturation of people to some extent, and if there was a history of smoking but then quit smoking, the oxygen saturation would rise and return to a level close to normal.

Second, the relationship between oxygen saturation and human sex is not particularly obvious. For men and women in the same physical condition, the oxygen saturation value is close, there is no obvious difference. The difference of

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

blood oxygen saturation between men and women with great difference in physical fitness is obvious. Therefore, it can be concluded that the effect of sex on the level of blood oxygen saturation is weak.

Moreover, for the relationship between oxygen saturation and body mass index, when a person's body mass index is at a normal level, his oxygen saturation level will be at a higher level. If the body mass index is too high at overweight or malnutrition is too low, the oxygen saturation level will be lower than normal.

Finally, oxygen saturation and age are negatively correlated, indicating that when a person increases with age, his or her oxygen saturation level decreases gradually. The main reason for this phenomenon is that when people grow older, the metabolic rate in the human body will decrease, and the physiological functional levels of people will gradually weaken. It makes it difficult for the body to maintain the normal level of blood oxygen content physiological behavior. Therefore, this can also explain to some extent the physiological situation of hypoxia in some elderly people from time to time.

5. Conclusion

By establishing clustering and multiple regression models and analyzing the corresponding results, we can get that the change of blood oxygen saturation has a great relationship with smoking and body weight. The change of blood oxygen saturation in people with smoking, smoking history and substandard weight will be abnormal, and the blood oxygen saturation level will decrease gradually with the increase of age. Therefore, in daily life, we should develop good living habits and maintain a healthy living state.

References

[1] Han Shuai. Design and Application of Oxygen Saturation Monitor[D].Tianjin University of Technology ,2016.

[2] Wang Xiaofei, Zhao Wenjun. Multi-wavelength pulse oximetry based on dynamic spectroscopy[J].Spectroscopy and Spectral Analysis ,2014,34(05):1323-1326.

[3] You Chenyu. Non-invasive detection of tissue oxygen saturation using visible near infrared spectroscopy[D].Zhejiang University ,2015.

[4] Downing. Application of Image Processing in the design of Android mobile phone blood oxygen saturation monitoring software[D].Donghua University ,2014.

[5] Zhang Hao. Research on Oxygen Saturation Algorithm Based on Photoelectric Volume Pulse Wave [D]. Nanjing University of posts and Telecommunications, 2018.

[6] Pilcher Janine Marie,Kearns Cilein,Beasley Richard. Searching for the optimal oxygen saturation range in acutely unwell patients[J]. Emergency Medicine Journal,2020.

ISSN 2706-6827 Vol. 2, Issue 9: 172-180, DOI: 10.25236/IJFS.2020.020924

[7] Zhou Tangxing. A Study on FPGA Pulse and Oxygen Saturation Monitoring System[D]. Donghua University ,2015.
[8] Chen Yuhong, Zhao Chai, Zhao Qian, Hu Zhenjie. Central venous oxygen saturation combined with venous carbon dioxide partial pressure difference to guide capacity management in sepsis patients[J].Chinese General Medicine 2016,19(11):1276-1281.
[9]Hu Bi, Wang Maoyun, Wang Yiwei, Zhang Run, Lei Fei, Liang Zongan. Significance of nocturnal dynamic oxygen saturation monitoring in patients with metabolic syndrome complicated with obstructive sleep apnea hypopnea syndrome[J]. Chinese General Medicine 2017,20(11):1287-1293.
[10]Jia-Ming Zhu,Lu Wang,Jia-Bao Liu,Chung-Min Liao. Eradication of Ebola Based on Dynamic Programming[J]. Computational and Mathematical Methods in Medicine,2016.
[11]Le Yi. Diabetic retinopathy and saturation of retinal vessels in diabetic retinopathy and Vogt-of willow field syndrome[J].Tianjin Medical University ,2016..
[12]Li Gang, Bao Lei, Zhou Mei, Lin Ling, Liu Rui, Zhao Chunjie. A new method for measuring arterial oxygen saturation[J].Spectroscopy and Spectral Analysis 2016,36(01):196-200.
[13]Fei Ba, Chen Weiping, Xu Li, Yin Jun. Research on Measurement System of Oxygen Saturation Based on Photoelectric Volume Pulse Wave[J].Industrial Instrumentation and Automation ,2015(05):14-16.
[14]Xu Panpan, Xu Bingqiao, Xu Wenlong. AFE4400- based pulse oximetry system[J].Laser and Infrared ,2015,45(03):320-324.