Cardiac and Cardiovascular Research (2021) Volume 2 Issue 1, 6 *pages*.
doi: 10.54517/ccr.v2i1.1896

ORIGINAL RESEARCH ARTICLE

Prediction of cardiovascular mortality based on Grey Markov model

Zikuan Wang, Haiwen Yu**

Nanchang University, Nanchang 330031, Jiangxi, China. E-mail: yuhaiwen@ncu.edu.cn

ABSTRACT

With the development of society and economy, the number of cardiovascular diseases continues to increase. Accurate prediction of cardiovascular mortality can guide the prevention of cardiovascular events and the sustainable development of public health. This paper mainly studies the grey Markov prediction model of cardiovascular disease mortality, proposes an improved grey Markov prediction model, and then uses the statistical data of cardiovascular disease mortality in rural and urban areas from 1991 to 2018 for numerical simulation. The simulation results show that the improved grey Markov prediction model is effective.

Keywords*:* Grey Markov; GM (1,1) model; cardiovascular disease; forecast; simulation

1. Introduction

With the development of society and economy, people's living standards and lifestyles have undergone profound changes. With the aging of the population and the acceleration of urbanization, the epidemic trend of cardiovascular risk factors is becoming more and more obvious, resulting in the continuous increase in the number of cardiovascular disease patients, and the number of cardiovascular patients will continue to increase in the next 10 years [1]. According to the data in the 2019 China Health Statistical Yearbook [2], the death of cardiovascular diseases (heart disease and cerebrovascular disease) ranks first among the total causes of death of urban and rural residents. In 2018, the rural mortality of cardiovascular diseases was 322.31/(100000), including 162.12/ (100000) for heart disease and 160.19/ (100000) for cerebrovascular disease; the death rate of cardiovascular disease in the city was 275.22/ (100000), of which the death rate of heart disease was 146.34/ (100000), and the death rate of cerebrovascular disease was 128.88/ (100000); in 2018, the ratio of cardiovascular disease deaths in rural and urban areas to all causes of death was 46.66% and 43.8% respectively. The occurrence of cardiovascular disease events is mainly attributed to the high rise of blood pressure and total cholesterol. Four nationwide hypertension sampling surveys were conducted in 1958-1959, 1979-1980, 1991 and 2002 [3]. The prevalence of hypertension among people over 15 years old was 5.1%, 7.7%, 13.6% and 17.6% respectively, showing an overall upward trend. The burden of cardiovascular disease is increasing,

ARTICLE INFO
Received: January 1, 2021 | Accepted: February 13, 2021 | Available online: March 2, 2021

CITATION
Wang Z, Yu H. Prediction of cardiovascular mortality based on Grey Markov model. Cardiac and Cardiovascular Research 2021; 2(1): 6 pages.

which has become a major public health problem [1]. Whether the mortality rate of cardiovascular disease can be accurately predicted is conducive to the prevention of cardiovascular events and the sustainable development of public health. Therefore, it is of great significance to study the prediction methods of cardiovascular mortality to guide the healthy development of public health in China.

The core and foundation of grey prediction theory is GM (1,1) model [4–6], which is applicable to prediction objects with short time, few data, small fluctuation and long-term trend. For objects with large random fluctuations, the predicted value will be high or low, and the degree of fitting is poor. Markov prediction model is suitable for a dynamic system with random changes. It predicts the future development trend of the system according to the transition probability between states. At present, many researchers have combined the grey GM (1,1) model with the Markov prediction model [7 – 10], that is, the grey Markov prediction model. This paper mainly studies the grey Markov prediction model of cardiovascular disease mortality, proposes an improved grey Markov prediction model, and uses the statistical data from 1991 to 2018 [2, 11] for numerical simulation. The numerical simulation results show that the improved grey Markov prediction model proposed in this paper is effective.

2. Grey Markov prediction model

2.1. Grey GM (1,1) model

GM (1,1) model is the simplest grey dynamic prediction model [4 – 6]. Its modeling principle and process are described below.

Set $x^{(0)}=\left\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\}$ as the original data sequence, where $x^{(0)}(i)>0, i=1,2,\cdots,n$. Make $x^{(0)}$ first-order accumulation $x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i), k=1,2,\cdots,n$ and record $x^{(1)}=\left\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right\}$. The definite solution problem of $x^{(1)}$ whitening differential equation of GM (1,1) model is constructed from the first-order accumulated data

$$\begin{cases}\dfrac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=u,\\ x^{(1)}(1)=x^{(0)}(1),\end{cases} \tag{1}$$

Where a and u are $x^{(1)}$ parameters to be determined by. By integrating both sides of the equation in equation (1) on the $[k,k+1]$ upper surface, it is obtained that

$$x^{(1)}(k+1)-x^{(1)}(k)+a\int_{k}^{k+1}x^{(1)}(t)dt=u \tag{2}$$

Obviously, $x^{(1)}(k+1)-x^{(1)}(k)=x^{(0)}(k+1)$ equation (2) is equivalent to

$$x^{(0)}(k+1)+a\int_{k}^{k+1}x^{(1)}(t)dt=u \tag{3}$$

When k from 1 to $n-1$, the overdetermined equations are obtained

$$\begin{bmatrix}-Z^{(1)}(2) & 1\\ -Z^{(1)}(3) & 1\\ \vdots & \vdots\\ -Z^{(1)}(n) & 1\end{bmatrix}\begin{bmatrix}a\\ u\end{bmatrix}=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix} \tag{4}$$

Where $Z^{(1)}(k+1)=\int_{k}^{k+1}x^{(1)}(t)dt$, $k=1,2,\cdots,$ $n-1$, it is usually called the background value of GM (1,1) model. The trapezoidal numerical integration formula [12], $Z^{(1)}(k+1)\approx\frac{1}{2}\left[x^{(1)}(k+1)+x^{(1)}(k)\right]$ can be used. Note that the least square solution of equation group (4) is $[\hat{a},\hat{u}]^{T}$. By substituting the sum of $\hat{a}$ the $\hat{u}$ least square solution into equation (1), the discrete approximate solution of equation (1) is obtained.

$$\hat{x}^{(1)}(k)=\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}(k-1)}+\frac{\hat{u}}{\hat{a}} \tag{5}$$

$k=1,2,\cdots,n$

Therefore, the GM (1,1) prediction model is

$$\hat{x}^{(0)}(k+1)=\left(1-e^{\hat{a}}\right)\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k} \tag{6}$$

$k=1,2,\cdots,n$

2.2. Markov prediction model

Let $\left\{Y_n, n=0,1,2,\cdots\right\}$ it be a sequence of discrete random variables, and the possible values of

Y_n belong to E, but E is a set containing finite elements, which is a set of finite states. If the conditional probability of Y_{n+1} depends only on Y_n the value of, that is $P\left(Y_{n+1}=i_{n+1} \mid Y_n=i_n, Y_{n-1}=i_{n-1}, \cdots, Y_0=i_0\right)=P\left(Y_{n+1}=i_{n+1} \mid Y_n=i_n\right)$, the stochastic process is called a finite state Markov chain [7-8]. When describing the probability distribution of Markov chain, the most important is conditional probability $p_{ij}(k)=P\left(Y_{k+1}=j \mid Y_k=i\right)$, which represents the probability of taking k value Y_k at i the next time under the condition of Y_{k+1} taking j value at the time. It $p_{ij}(k)$ is k generally called one-step transition probability of time. If the conditional i probability of random j variable transferring from state to state k is independent of time, that is $p_{ij}(k)=P\left(Y_{k+1}=j \mid Y_k=i\right) p_{ij}$, it is called homogeneous Markov chain.

Set P as the matrix of the one-step transition probability of the p_{ij} text Markov chain, i.e

$$P=\begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0m} \\ p_{10} & p_{11} & \cdots & p_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m0} & p_{m1} & \cdots & p_{mm} \end{bmatrix} \quad (7)$$

It P is called one-step transition probability matrix, $\sum_{j=0}^{m} p_{ij}=1$ and. Generally, only the transition probability matrix needs to be P considered. If the inspected object is in the state k and the row of P the k matrix is $\max_j p_{ij}=p_{kN},(j=1,2,\cdots,m)$ inspected, it is considered that the next moment is most likely to k change from the N state to the state. If P there k are 2 or more identical elements in the row of the matrix, the multi-step transition probability matrix must be further investigated.

2.3. Grey Markov prediction model

The grey Markov prediction model uses the fitting value of GM (1,1) model which reflects the change of original data to construct the state space of Markov chain, and then realizes the model prediction. First, a $m+1$ curve parallel $\hat{x}(k)$ to the trend E is m used $F_1, F_2, \cdots, F_m$ to divide into states:

$$F_i=\left[F_{i1}, F_{i2}\right](i=1,2,\cdots,m)$$
$$F_{i1}=\hat{x}(k)+\alpha_i \bar{X}, F_{i2}=\hat{x}(k)+\beta_i \bar{X}$$

Where α_i, β_i is the translation constant, and $\alpha_{i+1}=\beta_i$ is $\bar{X}$ the mean value $x^{(0)}$ of the F_{i1}, F_{i2} original data, which changes with time. Assume that the $x^{(0)}(k)$ original i data $F_i=\left[F_{i1}, F_{i2}\right]$ is in $\hat{z}_M(k)=\frac{F_{i1}+F_{i2}}{2}$ the $x^{(0)}(k)$ state, then take the fitting value of; $x^{(0)}(n)$ If k the $F_k=\left[F_{k1}, F_{k2}\right]$ last data $x^{(0)}(n+1)$ is in the state, the predicted value is

$$\hat{z}_M(n+1)=\sum_{j=1}^{m} \frac{F_{j1}+F_{j^2}}{2} p_{kj} \quad (8)$$

Where $F_{j1}=\hat{x}(n+1)+\alpha_j \bar{X}, F_{j2}=\hat{x}(n+1)+\beta_j \bar{X}$, $\hat{x}(n+1)$ is the prediction value obtained from the grey GM (1,1) model, and equation (8) is called the grey Markov prediction model.

2.4. Improved grey Markov prediction model

In order to further improve the prediction accuracy, this paper proposes the following improved grey Markov prediction model, namely

$$\begin{aligned}\hat{z}(k)&=\beta_0+\beta_1 \hat{z}_M(k)+\beta_2 \hat{x}(k)+\\ &\beta_3 \hat{z}_M(k) \hat{x}(k)+\beta_4\left(\hat{z}_M(k)\right)^2+\beta_5(\hat{x}(k))^2\end{aligned} \quad (9)$$

$x^{(0)}(k)$ is the new fitted value and predicted value of $\beta_0, \beta_1, \beta_2$, where $\beta_3, \beta_4, \beta_5$ is the pending parameter. In the following prediction of cardiovascular mortality, it is directly calculated by using the function regress in MATLAB $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ software, and the format is:

$$b=\text{regress}(y, A, \text{alpha})$$

Where y is the original $x^{(0)}(k)$ data, a is the matrix corresponding $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ to the regression coefficient, alpha is α the confidence $\alpha=0.05)$ level (default; b the output is a six dimensional column vector, and its $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ components are the calculated values of (**Figure 1**).

2.5. Calculation method of state division

In order to give the state division in the grey Markov prediction model, the definition of the central trend curve is given first.

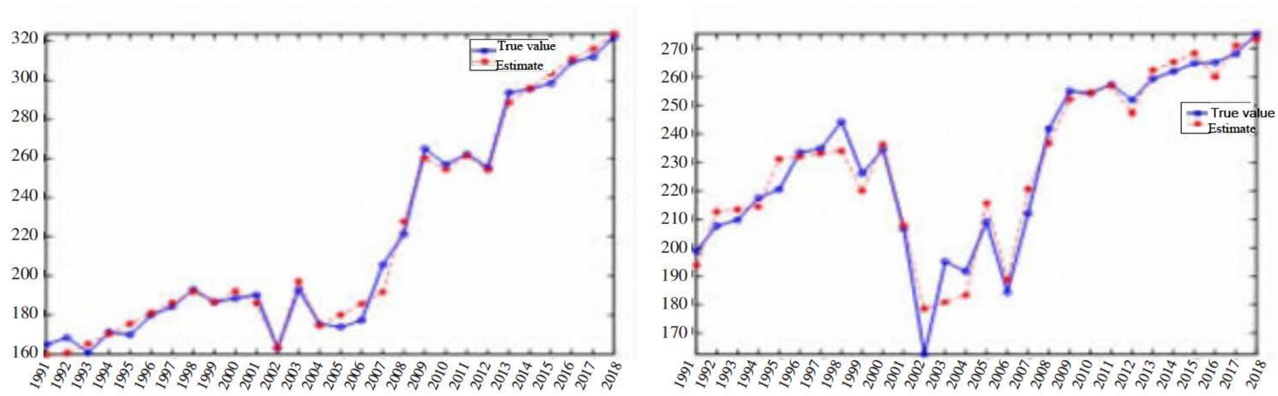
Definition 1[9-10]: $F(k)=\hat{x}(k)+\gamma\bar{X}$ set $F(k)$ if

$$\min_{\gamma}\sum_{k=1}^{n}\left[x(k)+\gamma\bar{X}-x^{(0)}(k)\right]^{2}$$

If yes, it $F(k)$ is called the central trend curve of the original data.

Definition 2[9-10]: set $T=\max_{k}\left\{x^{(0)}(k)-F(k)\right\}$, it T is called the positive deviation of the original data from the central trend curve. If $D=\min_{k}\left\{x^{(0)}(k)-F(k)\right\}$, then D is called the negative deviation of the original data from the central trend curve.

According to the above definition, it can be obtained from $r=\frac{\sum_{k=1}^{n}\left[x^{(0)}(k)-\hat{x}(k)\right]}{n\bar{X}}$ the knowledge of calculus. When dividing the status, first calculate the central trend curve, positive deviation and negative deviation, and take the central trend curve as the center Then the states are divided according to the characteristics of the data. In this article, $F_i=\left[F_{i1},F_{i2}\right](i=1,2,3,4)$ is taken, as $F_{i1}=F(k)+\alpha_i\bar{X}$, $F_{i2}=F(k)+\lambda_i\bar{X}$ the four states of the original data, where $\alpha_1=\left(D/\bar{X}-\varepsilon_1\right)$, $\lambda_4=T/\bar{X}+\varepsilon_2$, $\lambda_1=c_1\alpha_1$, $\alpha_2=\lambda_1$, $\lambda_2=0$, $\lambda_3=c_2\lambda_4$, $\alpha_3=\lambda_2$, $\alpha_4=\lambda_3$, ε_1, ε_2 are decimals greater than 0, . $0<c_1,c_2<1$. Among them, in Section 2, cardiovascular disease prediction, $\varepsilon_1=-0.1D/\bar{X}$, $\varepsilon_2=0.1T/\bar{X}$, $c_1=c_2=1/3$ take (**Table 1**).



(a) Rural cardiovascular mortality (b) urban cardiovascular mortality

Figure 1. Prediction effect of Improved Grey Markov model.

Table 1. Fitting results and accuracy test of Improved Grey Markov model and GM (1,1) model

Time	Raw data	GM (1,1) fitting value	Improved grey Markov fit value	Raw data	GM (1,1) fitting value	Improved grey Markov fit value
	Mortality rate of cardiovascular disease in rural areas (1/ (100000))			Urban cardiovascular disease mortality (1/ (100000))		
1991	164.960 0	164.960 0	159.804 5	198.840 0	198.840 0	193.828 3
1992	168.440 0	146.101 2	160.477 0	207.760 0	202.901 5	212.718 9
1993	160.890 0	150.471 0	165.407 4	209.820 0	204.846 6	213.508 2
1994	171.080 0	154.971 4	170.450 5	217.520 0	206.810 4	214.381 6
1995	170.030 0	159.606 5	175.607 5	220.580 0	208.793 0	231.243 2
1996	179.990 0	164.380 1	180.879 7	233.500 0	210.794	232.095 0
1997	184.240 0	169.296 6	186.268 0	234.870 0	212.815 5	233.036 2
1998	193.120 0	174.360 1	191.773 5	244.300 0	214.855 7	234.069 2
1999	186.560 0	179.575 1	186.676 1	226.100 0	216.915 4	220.094 0
2000	188.630 0	184.946 0	192.267 6	234.610 0	218.994 9	236.420 3
2001	190.320 0	190.477 6	185.877 3	206.780 0	221.094 3	207.805 3
2002	163.050 0	196.174 6	163.390 6	162.490 0	223.213 9	178.562 4
2003	193.060 0	202.042 0	197.319 9	195.130 0	225.353 8	180.953 6
2004	175.470 0	208.084 9	174.415 3	191.600 0	227.514 2	183.460 5
2005	173.870 0	214.308 5	180.077 9	209.240 0	229.695 3	215.857 2
2006	177.320 0	220.718 3	185.839 3	184.410 0	231.897 3	188.833 0
2007	205.700 0	227.319 8	191.698 2	212.080 0	234.120 4	220.575 5

Table 1. (*continued*)

Time	Raw data	GM (1,1) fitting value	Improved grey Markov fit value	Raw data	GM (1,1) fitting value	Improved grey Markov fit value
2008	221.260 0	234.118 7	227.839 5	241.790 0	236.364 8	236.828 7
2009	264.980 0	241.121 0	260.185 6	255.090 0	238.630 8	252.235 5
2010	257.050 0	248.332 7	254.461 4	254.340 0	240.918 4	254.578 9
2011	262.370 0	255.760 1	261.291 0	257.41	243.228 0	257.050 7
2012	255.450 0	263.409 6	254.142 9	251.97	245.559 8	247.370 3
2013	293.690 0	271.288 0	288.552 0	259.4	247.913 9	262.393 1
2014	295.630 0	279.402 0	295.911 1	261.99	250.290 6	265.270 3
2015	298.420 0	287.758 6	303.370 5	264.84	252.690 0	268.289 6
2016	309.330 0	296.365 2	310.925 8	265.11	255.112 5	260.110 3
Accuracy test	Average relative error	0.082 3	0.019 2	Average relative error	0.084 8	0.024 7
	Mean square deviation ratio	0.4065	0.102 5	Mean square deviation ratio	0.812 0	0.237 2
	Small error probability	0.846 2	1.000 0	Small error probability	0.653 8	1.000 0

3. Grey Markov model of cardiovascular mortality

In this paper, the rural and urban cardiovascular mortality from 1991 to 2018 are selected as the prediction objects. The data from 1991 to 2016 are used as the original data to establish the grey Markov chain prediction model, and the data from 2017 and 2018 are used as the prediction sample data to observe the prediction effect.

Firstly, the data from 1991 to 2016 are substituted into equation group (4) to obtain the least square solution of equation group (4). Therefore, the grey GM (1,1) prediction models of cardiovascular disease mortality in rural and urban areas are obtained as follows:

$$\hat{x}^{(0)}(k+1)=\left(1-e^{-0.0295}\right)\left(x^{(0)}(1)+\frac{139.075}{0.0295}\right)e^{0.0295k}$$

$$,k=1,2,\cdots,n,\hat{x}^{(0)}$$

$$(k+1)=\left(1-e^{-0.0095}\right)\left(x^{(0)}(1)+\frac{200.0379}{0.0095}\right)e^{0.0095k}$$

$$k=1,2,\cdots,n,$$

Then, calculate the central trend curve, positive deviation and negative deviation, and divide the original data into states according to the calculation method given in section 1.5. Then, the one-step transition probability matrix is calculated according to the Divided States. Finally, from the $\hat{x}(k)$ fitting $\hat{z}_M(k)$ values and $\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5$ calculated parameters, the improved grey Markov prediction models are

$$\hat{z}(k)=-31.8304+0.5449\hat{z}_M(k)+0.7976\hat{x}(k)$$
$$+0.0113\hat{z}_M(k)\hat{x}(k)-0.0045\left(\hat{z}_M(k)\right)^2-0.0076(\hat{x}(k))^2$$
$$\hat{z}(k)=533.1165+2.2658\hat{z}_M(k)-5.8985\hat{x}(k)$$
$$-0.0053\hat{z}_M(k)\hat{x}(k)-0.0004\left(\hat{z}_M(k)\right)^2-0.0157(\hat{x}(k))^2$$

The prediction effect of the improved grey Markov model is shown in **Figure 1**, and the comparison with GM (1,1) model and accuracy test are shown in Table 1.

According to the classification criteria for accuracy test in table 2[5] and the calculation results in **Table 2** and **Table 3**, it can be seen that the improved grey Markov model proposed in this paper can well predict the mortality of cardiovascular diseases, and has better results for the prediction of the mortality of cardiovascular diseases in rural areas. The reason may be that the mortality of cardiovascular diseases in urban areas in 2002 was very low compared with other years, resulting in the poor accuracy of the model.

Table 2. Classification standard for precision inspection

Classification	Relative error α	Mean square deviation ratio C0	Small error probability P0
Level 1 (good)	0.01	0.35	0.95
Level II (qualified)	0.05	0.50	0.80
Level III (barely)	0.10	0.65	0.70
Level IV (unqualified)	0.20	0.80	0.60

Table 3. Prediction results of Improved Grey Markov model and GM (1,1) model

Time	Raw data	GM (1,1) fitting value	Improved grey Markov fit value	Raw data	GM (1,1) fitting value	Improved grey Markov fit value
	Mortality rate of cardiovascular disease in rural areas (1/ (100000))			Urban cardiovascular disease mortality (1/ (100000))		
2017	311.880 0	305.229 2	316.090 9	268.190 0	257.558 1	271.091 4
2018	322.310 0	314.358 4	323.776 4	275.220 0	260.027 3	273.303 4
	Average relative error	0.022 998	0.009 026	Average relative error	0.047 423	0.008 891

4. Conclusions

Based on the statistical data of cardiovascular disease mortality in rural and urban areas from 1991 to 2018, this paper selects the grey Markov model to predict it, and puts forward an improved grey Markov prediction model of cardiovascular disease mortality. The numerical simulation results show that the improved grey Markov prediction model is effective, especially for the current prediction of cardiovascular disease mortality in rural areas, this will help prevent and guide the development of public health undertakings for cardiovascular disease prevention and treatment. In fact, cardiovascular disease is related to many factors, such as hypertension, dyslipidemia, air pollution, etc. How to establish the causal relationship between cardiovascular events and important factors, so as to establish a better prediction model, needs further research.

Conflict of interest

The authors declare no conflict of interest.

References

1. National Cardiovascular Center China cardiovascular disease report. Beijing: China Encyclopedia press; 2019.
2. Editorial staff of China Health Statistical Yearbook China Health Statistical Yearbook. Beijing: China Union Medical College Press; 2019.
3. Liliming, raokeqin, konglingzhi, et al Survey on nutrition and health status of Chinese residents in 2002. Chinese Journal of epidemiology 2005; 26(7): 478–484.
4. Liusifeng Grey system theory and its application. Beijing: Science Press; 2008.
5. Xiao Zhengming, Li Huan Application of Grey Theory in urban land price prediction -- a case study of Xiamen City. Jiangxi science 2011; 29(5): 663–666.
6. Sun Aimin Prediction of power demand in Xi'an Based on metabolic grey prediction model. Practice and understanding of mathematics 2019; 49(23): 298–305.
7. Huang Yinhua, Peng Jian, lichangchun, et al. Application of Markov theory in medium and long term load forecasting. Journal of power system and automation 2011; 23(5): 131–136.
8. Zhouzhijian, fuzetian, wangruimei, et al Application of Grey Markov model in cotton yield prediction. Statistics and decision making 2005; 183 (2): 48– 49.
9. Liu Miao Grey prediction model and its application in power demand. Xi'an: Xi'an University of architecture and technology; 2012.
10. Yellow steel Research on emergency blood support characteristics and demand prediction model for unconventional emergencies. Chengdu: Southwest Jiaotong University; 2012.
11. Editorial board of China Health Yearbook China Health Yearbook. Beijing: People's Health Publishing House; 1992–2002.
12. Xuhuilin, Gao Xingjun. Numerical differentiation algorithm based on integral. Jiangxi science 2014; 32(1): 1–4.