ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

SIR-Based Flu Epidemic Model

Xiaojun Zhou[1], Yuxi Mao[1], Jiayu Liu[1], Xiaoqin Li[1], Xuehan Wu[1, *]

[1]College of Science, Chongqing University of Technology, Chongqing, 400054, China

Abstract

Influenza is an important public health issue of global concern. This paper conducts short-term predictions by studying the changes in the population of China and the characteristics of influenza transmission, and makes recommendations for the prevention and control of influenza. Investigate the law of population change in China, select the data from 1978-2018 for 41 years to establish the model of population change law, compare the polynomial curve fitting and logistic algorithm, and select the logistic population model to predict the population of 139.912 million in 2019. The final rule is: the population base is large, and the total population is on the rise. After reaching a peak at a certain moment, it will show a negative growth.

Keywords

Influenza transmission; logistic population model; SIR model.

1. INTRODUCTION

The population problem is related to the spread of influenza. The spread of influenza includes contact and air transmission [1]. China has a large population base and a small per capita land area, which is likely to cause local outbreaks of influenza. To explore the spread of the flu epidemic, we must first study the changing laws of the Chinese population [2]. We decided to use the polynomial curve fitting and logistic algorithm to select the national population data for 41 years from 1978 to 2018, calculate the population growth rate, and use the fitted Formula, select the past few years as a test set, and compare the errors[3]. The model with small error and high fitness is used to predict the population of China in 2019.

2. SYMBOL DESCRIPTION

X (t): Population at time t x_0: Population at the initial moment r: Population growth rate

x_m: Maximum population capacity δ: Relative error x: Predictive value μ: actual value

S: Susceptible person I: Sick population I: Sick population R: Exit the crowd

λ: Incidence rate v: Cure rate N: Total population

3. MODEL ESTABLISHMENT

3.1. Polynomial Fitting Model

Polynomial fitting has its place in many fields because the main function is to find smooth curves to better represent noise-bearing measurement data. The larger the order, the higher the fitting accuracy. The expansion factor is determined using a least squares fit.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

$$p(x)=p_1x^n+p_2x^{n-1}+\ldots+p_nx+p_{n+1} \tag{1}$$

3.2. Logistic Model

The Logistic model was proposed by Verhulst-Pearl in the revision of the non-density equation in 1938. He believes that there is always an upper limit to the growth of the population in a certain environment. When the population gradually increases toward the upper limit, the actual growth rate will gradually decrease.

The population at time t is $x(t)$ and $x(t)$ Think of a continuous, differentiable function. The population at the initial moment (t=0)is x_0. Suppose the population growth rate is constant r, that is, unit time $x(t)$ the increment is equal to r multiplied by $x(t)$. We consider t to $t+\Delta t$ the increase in population during the time, there is

$$x(t+\Delta t)-x(t)=rx(t)\Delta t \tag{2}$$

Make $\Delta t\to 0$, then get $x(t)$ satisfy the following differential equation

$$\frac{dx}{dt}=rx, x(0)=x_0 \tag{3}$$

The block effect is reflected in the impact on the population growth rate r, so that r decreases as the population x increases. If the population growth rate r is expressed as a function of the population quantity x, r(x), then r(x) is a decreasing function, so (3) can be written as:

$$\frac{dx}{dt}=r(x)x, x(0)=x_0 \tag{4}$$

Let r(x) be a linear function of x, ie:

$$r(x)=r-sx(r>0,s>0) \tag{5}$$

Here r denotes the growth rate when the population is small (theoretically x = 0), that is, the natural growth rate of the population without environmental and resource constraints. In order to clarify the meaning of the parameter s, the maximum population capacity is introduced x_m, the largest population that can be accommodated by natural resources and environmental conditions. Then $x=x_m$ When the population growth rate is zero, that is, the growth rate $r(x_m)=r-sx_m=0$ to get $s=\frac{r}{x_m}$, then (5) can be rewritten as:

$$r(x)=r(1-\frac{x}{x_m}), x(0)=x_0 \tag{6}$$

Substituting (6) into (4) the following Logistic model:

$$\frac{dx}{dt}=rx(1-\frac{x}{x_m}), x(0)=x_0 \tag{7}$$

The general solution of equation (7) is obtained by the variable separation method: $\frac{x}{x_m-x}=ce^{rt}$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

Using the initial conditions: $c=\frac{x_0}{x_m-x_0}$, put c into the general solution and simplify it:

$$x(t)=\frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}} \tag{8}$$

(8) The formula can be abbreviated as:

$$x=\frac{x_m}{1+ae^{-bt}} \tag{9}$$

Among them $a=\frac{x_m}{x_0}-1, b=r$.

3.3. Sir Model

The population is divided into three categories: susceptible person S, disease group I and withdrawal group R (including death and healers). The proportion of these three types of people in the total number of people at time t is $S(t)$, $I(t)$, $R(t)$.

The susceptible person and the affected person become effective after contact with the disease. Let the number of susceptible people with effective daily contact per affected person be $\lambda S(t)$, $NI(t)$, the average incidence of the disease $\lambda S(t)$, $NI(t)$, a susceptible person becomes a virus lurker. So have:

$$\frac{dS(t)}{dt}=-\lambda S(t)I(t) \tag{10}$$

The change in the withdrawal time per unit time is equal to the reduction in the affected population, ie:

$$\frac{dR(t)}{dt}=vI(t) \tag{11}$$

The change in the affected population is equal to the number of susceptible people transferred, ie

$$\frac{dI(t)}{dt}=\lambda S(t)I(t)-vI(t) \tag{12}$$

The standard model equations 11, 12, and 13 are further assumed to assume that the newborn is a susceptible person. The incidence of disease is two-line λSI, v Is the recovery rate coefficient, the mover is no longer infected. The sir box of the flu is shown in Figure1:

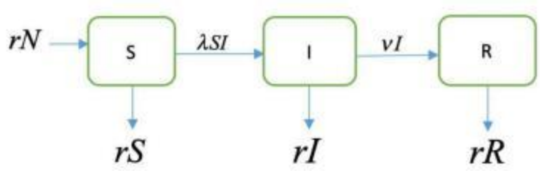

Figure 1. Flu sir model box diagram

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

Then the sir model is corrected to:

$$\begin{cases} \frac{dS}{dt} = rN - \lambda SI - rS \\ \frac{dI}{dt} = \lambda SI - vI - rI \\ \frac{dR}{dt} = vI - rR \end{cases} \tag{13}$$

Add the above 3 formulas to get:

$$\frac{d(S+I+R)}{dt} = \frac{dN}{dt} = 0 \tag{14}$$

Which is:

$$N(t) = S(t) + I(t) + R(t) = K \tag{15}$$

3.4. Sir Optimization Model

The incidence and cure rate of the SIR model above are not realistic. Therefore, based on the actual situation of China, based on the logistic model of the first question, consider the maximum capacity of the environment. x_m and the growth law of population s, the flow chart of sir optimization model is shown in Figure2:

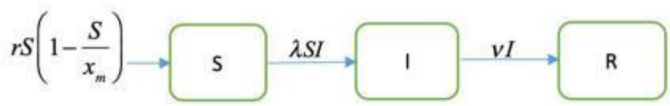

Figure 2. Sir optimization model flow chart

Then the model equation is:

$$\begin{cases} \frac{dS}{dt} = rS(1 - \frac{S}{x_m}) - \lambda SI \\ \frac{dI}{dt} = \lambda SI - vI \\ \frac{dR}{dt} = vI \end{cases} \tag{16}$$

3.5. Solution of The Model

3.5.1 Polynomial fitting model

Step one:

A total of 41 years of data for the Chinese population from 1978 to 2018 was selected. Set 1978 as the 0th year.

Step two:

The polyfit function is used to perform a 2nd order polynomial fitting, a 4th order polynomial fitting, a 5th order polynomial fitting, and a fitting residual sum.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

Table 1. Fit residuals of different orders

Order	Fitting residuals and
2	1.1085×107
4	9.7414×105
5	4.1629×105

According to Table1, the fifth-order polynomial fitting residual is the smallest, which is 4.1629×105. It can also be seen from Figure3 that the 5th-order polynomial has the best fitting effect, then the 5th-order polynomial model is selected for prediction.

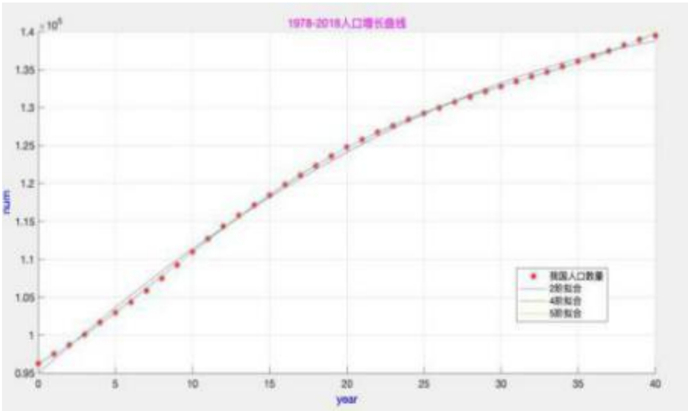

Figure 3. China's population quantity polynomial fitting map

Step three:

The polyval function is used to predict that China's population will be 1404.8 million in 2019.

Step four:

Using the established 5th-order fitting polynomial, 2005, 2010, and 2015 are predicted as test sets, and the relative error between the predicted value and the true value is compared to judge the pros and cons of the model. The total population predicted is: 137.47 million, 1347.6 million, and 1381.7 million.

Through the relative error formula:

$$\delta = \frac{x-\mu}{\mu} \times 100\% \tag{17}$$

The relative errors calculated in 2005, 2010 and 2015 were: 0.547%, 0.499%, and 0.515%, respectively. The average error is about 0.5%, which indicates that the model is realistic and the predicted results are more accurate.

5.5.2 Logistic model

Step one:

A total of 41 years of data for the Chinese population from 1978 to 2018 was selected. Set 1978 as the 0th year.

Step two:

Under the linear assumption of Logistic, there are the following Logistic models:

$$x = \frac{x_m}{1+ae^{-bt}} \tag{18}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

(1) Seeking x_m

The deformation of (7) is obtained:

$$\frac{\frac{dx}{dt}}{x}=r-\frac{r}{x_m}x \tag{19}$$

Make $\frac{\frac{dx}{dt}}{x}=r_k$ for the annual growth rate, use the numerical differentiation method to calculate the growth rate based on the population data. r_k and then r_k Linear fitting can be obtained $r_k=cx+d$, from which we can get: $x_m=|\frac{d}{c}|=15.1127$.

(2) Seeking a and b

Transform (9) into:

$$\frac{x_m}{x}-1=ae^{bt}(a=\frac{x_m}{x_0}-1,b=r) \tag{20}$$

Take the logarithm on both sides:

$$\ln(\frac{x_m}{x}-1)=\ln a-bt \tag{21}$$

Make $y=\ln(\frac{x_m}{x}-1)$, $B=\ln a$, $A=-b$. The more cumbersome exponential form of the solution can be transformed into a linear function, and the strong curve is transformed into a linear model: $Y=At+B$.

Using Matlab software, you can fit the values of A and B to find out $a=e^B$, $b=-A$ the value of the population model determines the specific form of the population model. Solve: $a=0.5710$, $b=0.0476$.

Step three:

The logistic regression model for the Chinese population is specifically:

$$x=\frac{x_m}{1+ae^{-bt}}=\frac{15.1227}{1+0.5710e^{-0.0476t}} \tag{22}$$

It is predicted that the population of China in 2019 will be 139.912 million.

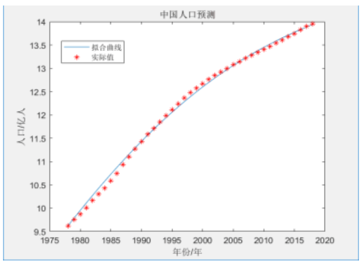

Figure 4. Logistic China population forecast map

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

The fitting of Matlab software can visually see the change of the value (see Figure4), and the predicted data and the actual data curve have a high degree of fit.

Step four:

Using the established logistic population model, 2005, 2010, and 2015 were used as test sets to predict the relative error between the predicted and actual values, and to judge the pros and cons of the model. The estimated total population is 130.08 million, 135.189 million, and 137.722 million.

Through the relative error formula, the relative errors calculated in 2005, 2010 and 2015 are: 0.1131%, 0.3369%, 0.189%. The average error is about 0.2%, which indicates that the model is established in accordance with the actual situation and the predicted result is accurate.

Summary of problems

Using 2005, 2010, and 2015 as test sets, the relative error of the polynomial fitting model is 0.5%, and the relative error of the logistic population model is 0.2%. By comparison, the logistic population model was finally selected, and the population of 2019 was predicted to be 139.912 million.

The logistic regression model of the Chinese population is specifically:

$$x = \frac{x_m}{1+ae^{-bt}} = \frac{15.1227}{1+0.5710e^{-0.0476t}}$$

According to the model, we can draw the law of population change in China:

China has a large population base;

The total population of China has been on an upward trend;

The population growth rate has gradually slowed down and the rate has gradually approached zero.

5.5.3 Sir model

Step one:

The data is processed, and the data from January 2018 to January 2019 is selected to fit the parameters, and the data is reduced by 10000000 times, which is convenient for the program to run.

Step two:

Make a susceptible population $S=1$ population growth rate $r=0.006$, total population $N=13.9913$, setting two parameters for disease incidence λ and cure rate v, the estimated value of the given parameter is 0.01, 1. Bringing (4), (5), using Matlab unconstrained nonlinear optimization function fminserarch operation $\lambda=2.2013$, $v=3.6370$, error is 1.0637×10^{-4},the fitting effect is shown in Figure5. It is observed that the fitting effect is quite good. However, the disease incidence rate and the cure rate calculated by the optimization algorithm are both over 1, which is not in line with the actual situation. Therefore, further optimization is performed on the basic sir model.

5.5.4 Sir optimization model

Step one:

The data is processed, and the data from January 2018 to January 2019 is selected to fit the parameters, and the data is reduced by 10000000 times, which is convenient for the program to run.

Step two:

Set four variables: disease incidence λ Cure rate v, population growth rate r and maximum environmental capacity x_m. The given estimated parameter values are: 0.1, 0.5, 0.01, 18.Make a susceptible population $S=1$. Bringing (9), (10), using Matlab unconstrained nonlinear

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

optimization function fminserarch operation $\lambda = 0.0918$, $v = 0.8629$, $r = 0.1758$, $x_m = 46.9714$. Error is 1.0817×10^{-9} the fitting effect is shown in Figure6.

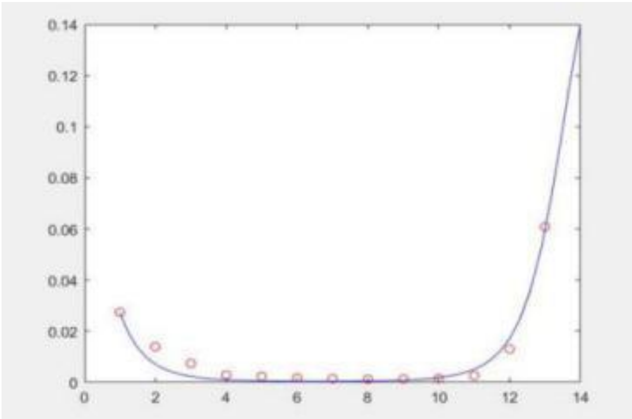

Figure 5. SIR flu fit diagram

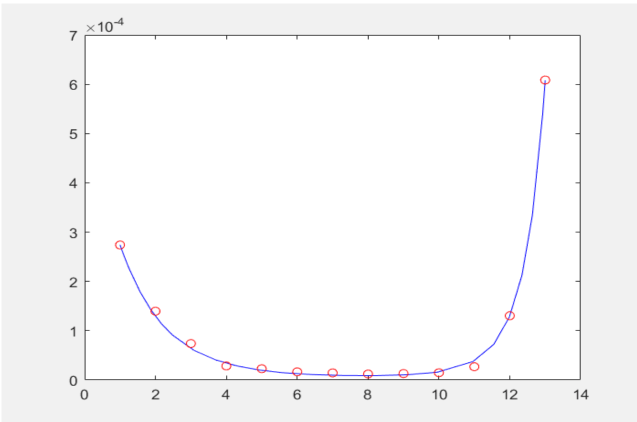

Figure 6. SIR optimization model fit diagram

According to Figure6, it is found that the optimized sir model fits well with the original data. It shows that the optimized model has certain reference significance in selecting parameters.

Step three:

Calculation threshold R_0, R_0 Is a node that distinguishes whether a disease is prevalent, if $R_0 < 1$, indicating that the disease is dying, $R_0 > 1$, indicating that the disease is endemic or forms an endemic disease. According to the data of this question, it is calculated: $R = \frac{\lambda x_m}{v} = 4.6842$. Threshold R_0 Greater than 1, indicating that the current (2019.1) period of influenza is prevalent.

Step four:

Bring the data from 2019.2-2019.6 into the original model to get a fitted map and predict the number of cases in July and August.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201912_5(12).0028

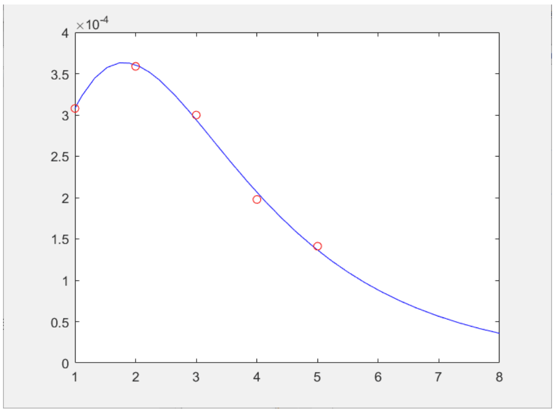

Figure 7. SIR optimization model prediction map

From Figure7, it can be seen that the fitting effect is better, and the number of cases of 2019.7 and 2019.8 is 85880 and 57090, respectively.

REFERENCES

[1] Shen Xiaofeng, Applicability Test of Logisitic Population Model in Ethnic Areas——Taking Guizhou Autonomous Prefecture as an Example [J]. Beijing: China Collective Economy, 2011.

[2] Yang Gongli, model and analysis of seasonal influenza in Hong Kong [J]. Qingdao: 2015.

[3] Ma Zhien, Zhou Yichang, Mathematical Modeling and Research of Infectious Disease Dynamics [M]. Beijing: Science Press, 2004, 116-142.