

# Forecast of China's Future Population under the Universal Two-child Policy

# Yan Peiyu

Basic Teaching Department, Shandong Huayu University of Technology, Dezhou, Shandong, China, 253034

Keywords: Population structure; Economic growth; Regression analysis; Least squares

**Abstract:** Firstly, according to the total population of China from 1996 to 2015 collected and sorted, the curve function was fitted using MATLAB tools and the polynomial difference method. According to the function predicted by the fitting function, the population of China in 2016 was 1.769.8 billion under the premise of no open two-child policy. Secondly, the functional relationship between mortality rate and time was obtained by using MATLAB tools and least squares algorithm, logistic model and curve fitting. Therefore the mortality rate in 2016, as well as the future mortality rate of China's population was predicted. Finally, a differential equation model and recursive mode were set up. Under the assumption that the birth rate of approximately equal to 1.2916% per year, generation into the prediction model, it is concluded that under the universal two-child policy in our country in 2016 and the total population of about 1.38242 billion people, the result that tallies with the China's population will reach 1.385 billion in 2016, and its forecast effect is better. At the same time, the annual population of China is predicted for the next ten years.

# **1. Introduction**

China is a big country with a population of 1.3 billion. The population problem will inevitably directly affect the development of social economy. From the initial implementation of family planning to the gradual opening of the birth policy, to the "universal two-child" policy, China's population and natural growth rate are constantly changing, and the demographic structure also presents different characteristics. Nowadays, the implementation of the "universal two-child" policy in the serious situation of an aging population structure, a prosperous old age and a serious minority child will inevitably lead to a rapid increase in the number of newborns in China. But there is no first-hand data in the national census. It is necessary to construct a prediction model for the future population, and implement the algorithm by combining the future population mortality rate obtained from the prediction of China's future population under the policy of "universal two-child".

# 2. China's population forecast before the policy of "universal two-child".

According to the relevant data of the total population of China in the past 20 years provided by the National Bureau of Statistics, and the future population of China is predicted when the policy is not open by using the MATLAB software tool and curving fitting in the polynomial difference method. China's population in 1996-2015 is shown in Table 1.

year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Total number of people (10,000)	122389	123626	124761	125786	126743	127627	128453	129227	129988	130756
year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Total number of people (10,000)	131448	132129	132802	133450	134091	134735	135404	136072	136782	137462

Table 1: Population of China from 1996 to 2015

The above data was processed by MATLAB software and EXCEL to make a scatter plot of the population of China, shown in Figure 1, in preparation for the later comparative analysis.





Figure 1 Fragment of population in China from 1996 to 2015

From the chart, it can be seen that in order to improve the prediction accuracy and reduce the error, the prediction of China's population under the policy of not fully opening the second child can be determined by polynomial regression analysis, and combined with the paired t test to enhance the prediction accuracy.

### 2.1 Polynomial regression analysis

A polynomial regression analysis of a dependent variable and one or more independent variables is called polynomial regression<sup>[1]</sup>.

In the one-way regression analysis, if the relationship between variables exhibits a nonlinear and complex feature, a one-dimensional polynomial regression can be used. A one-dimensional polynomial regression equation is established.

$$S(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

#### 2.2 Least squares algorithm

Finding the curve function with the smallest sum of squared errors in the data and observation data is called the least squares fit. In geometric language, it is called the least squares method of curve fitting<sup>[2]</sup>.

(1) Based on the observed data, assume the form of the functional relationship. S(x) is a linear form. If  $\varphi_k(x)$  it is a polynomial, S(x) is a polynomial, as shown below.

$$S(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \dots + a_n \varphi_n(x) \qquad (n < m)$$

(2) Establish a weighted sum of squared errors:  $\|\delta\|_2^2 = \sum_{i=0}^m \omega(x_i) [S(x_i) - f(x_i)]^2$ 

(3) Determine the least squares error:

$$\|\delta\|_{2}^{2} = \sum_{i=0}^{m} \delta_{i}^{2} = \sum_{i=0}^{m} \left[S^{*}(x_{i}) - y_{i}\right]^{2} = \min_{S(x)\in\varphi} \sum_{i=0}^{m} \left[S(x_{i}) - y_{i}\right]^{2}$$

Which is to ask for a function  $y = S^*(x)$ . Make the minimum, it translates into a multivariate function,

$$I(a_0, a_1, \cdots, a_n) = \sum_{i=0}^m \omega(x_i) \left[ \sum_{j=0}^n a_j \varphi_j(x_i) - f(x_i) \right]^2$$

The problem of the minimum value of the function, the necessary conditions for finding the extremum of the multivariate function are,



$$\frac{\partial I}{\partial a_k} = 2\sum_{i=0}^m \omega(x_i) \left[ \sum_{j=0}^n a_j \varphi_j(x_i) - f(x_i) \right] \varphi_k(x_i) = 0 \ (k = 0, 1, \dots, n)$$

(4) Establish a regular equation system:  $\sum_{j=0}^{m} (\varphi_k, \varphi_j) a_j = d_k \quad (k = 0, 1, \dots, n)$ 

Among them, to make the equation have a unique solution, there are:

$$G = \begin{bmatrix} \left( \varphi_0, \varphi_0 \right) & \left( \varphi_0, \varphi_1 \right) & \cdots & \left( \varphi_0, \varphi_n \right) \\ \left( \varphi_1, \varphi_0 \right) & \left( \varphi_1, \varphi_1 \right) & \cdots & \left( \varphi_1, \varphi_n \right) \\ \vdots & \vdots & \vdots & \vdots \\ \left( \varphi_n, \varphi_0 \right) & \left( \varphi_n, \varphi_1 \right) & \cdots & \left( \varphi_n, \varphi_n \right) \end{bmatrix} \end{bmatrix} = 0$$

Solving through the equations, finally the most suitable curve function relationship is found.

### 2.3 Algorithm application

(1) Model establishment Assume that the population function function in China is:

$$y = a_3 t^3 + a_2 t^2 + a_1 t + a_0$$

(2) Raw data preprocessing <sup>[3]</sup>

Considering that the units of the data are inconsistent, the data in the table is preprocessed first. The preprocessing formula is shown in Equation.

$$\frac{\lambda - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}$$

(3) Curve fitting

Using MATLAB software to program the fit<sup>[4],</sup> the relationship between the total population and time is shown in Equation.

$$y = 0.0001t^3 - 0.0046t^2 + 0.1312t + 12.1171$$

Its fitting curve is shown in Figure 2.



Figure 2 Population population fit curve

(4) Paired t-test<sup>[5]</sup>

Assumption:  $H_0: \mu_0 = 0$ . The difference between the function value and the experimental data is zero.

Assumption:  $H_0: \mu_0 \neq 0$ . The difference between the function value and the experimental data is not zero.



(5) Population forecast

If x = 21, it is predicted that the population of 2016 will be 137.698 million people without the policy of opening two-child policy.

#### 3. China's future population mortality forecast

A Logistic relationship model between death population and time was established. Through the curve fitting, the separable variable method was used to predict the annual mortality rate of China in the future.

Through the compilation of the National Bureau of Statistics' 1996-2015 China Population Mortality Release Data, MATLAB software was used to make a scatter plot of population mortality rate about time in China <sup>[6]</sup>, shown in Figure 3.



Figure 3 scatter diagram of China's population mortality from 1996 to 2015

It can be seen from the intuitive picture that the mortality rate of the Chinese population increased from 1996 to 2015 with the increase of the year, but after increasing to a certain extent, the growth rate gradually slowed down. It can be predicted that the fitting curve of this data is s-type curve, so the retarded growth Logistic model can be used for the forecast analysis.

#### 3.1 Introduction to Logistic model

The Logistic model is an important mathematical model proposed by Dutch mathematical biologist Ve Rhulst in 1938 to study the growth law of biological population in limited space<sup>[7]</sup>. Ecologists have used it to simulate and describe the growth process of various biological populations, and it has been widely used in similar fields.

Its basic form is shown in Equation.

$$\begin{cases} \frac{dx}{dt} = rx(t) \left( 1 - \frac{x(t)}{N} \right) \\ x(0) = 0 \end{cases}$$

Where, x(t) represents the population quantity at any time, r is the intrinsic growth rate (i.e. the growth rate minus the mortality rate) of China's population, and N is the maximum population that the environment can accommodate.

#### 3.2 Use the separable variable method to solve

The separable variable method in ordinary differential equation is applied to solve the equation.

$$x(t) = \frac{N}{1 + \left(\frac{N}{x(0)} - 1\right)e^{-r}}$$



# 3.3 Demographic mortality prediction analysis

First, establish the Logistic model:  $r_0 = \frac{1}{1 + ae^{kt}}$ Secondly, simplify by using the logarithm method shown in Equation.

$$\ln(\frac{1}{r_0} - 1) = \ln a + kx$$
,  $R = \ln(\frac{1}{r_0} - 1)$ ,  $A = \ln a$ ,

The unary linear regression model can be obtained: Y = kx + A.

Finally, using MATLAB software programming, the solution is: k = -0.0068, A = 5.0621. The Logistic model of mortality is shown in Equation.

$$r_0 = \frac{1}{1 + e^{5.0621}e^{-0.0068t}} = \frac{1}{1 + 157.9218e^{-0.0068t}}$$

The specific image is obtained and shown in Figure 4.



Figure 4 China's population mortality curve from 1996 to 2015

Thus, data of China's future mortality rate in 2017-2030 can be predicted, shown in Table 2.

year	2017	2018	2019	2020	2021	2022	2023
mortality(%)	0.69711	0.68832	0.67785	0.6657	0.65184	0.63636	0.61917
year	2024	2025	2026	2027	2028	2029	2030
mortality(%)	0.6003	0.57975	0.55752	0.53361	0.50822	0.48075	0.4518

Table 2 China's future mortality rate in 2017-2030

# 4. The opening-up policy

The differential equation and the recursive model are established, and a comparative analysis of China's population in the future is made under the policy framework of "universal two-child policy".

# 4.1 Prediction of the differential equation model

It is known that population growth rate r is the difference between population birth rate  $r_1$  and population death rate  $r_0$ . As the population growth is shown in Equation.

$$x(t+\Delta t) - x(t) = \Delta t x(t) (r_1 - r_0)$$

Therefore, the differential equation model is established and shown in Equation.



$$\begin{cases} \frac{dx(t)}{dt} = x(t)(r_1 - r_0) \\ x(21) = x_0 \end{cases}$$

It is known that 17.86 million people were born in 2016 and the total population is 137.698 million. The birth rate in 2016 can be obtained.

$$r_{1} = \frac{1786}{137698} \times 100\% = 1.2916\%$$
  
So 
$$\begin{cases} \frac{dx(t)}{dt} = x(t) \left[ 0.0129 - \frac{1}{1 + 157.9218e^{-0.0068t}} \right] \\ x(21) = 137698 \end{cases}$$

The general solution of the equation is obtained by the method of separating variables, shown in Equation.

$$\ln x(t) = C - 0.9871t - \frac{\ln(1 + 157.9218e^{-0.0068t})}{0.0068}$$

A special solution:

Therefore, the prediction model of population size is shown in Equation.

x(21) = 137698 C = 757.0514.

$$\ln x(t) = 757.0514 - 0.9871t - \frac{\ln(1 + 157.9218e^{-0.0068t})}{0.0068}$$

#### 4.2 Prediction of recursive model

As it can be predicted from the above problems, the mortality rate of the population in 2016 is 0.7308%.

Total population in 2016 = total population in 2015 + number of births - number of  $deaths = 137462 + 1786 - 137698 \times 0.7308\% = 1382420000$ .

This result is consistent with Tan Yong hong's [9] prediction that the total population of China will reach 1.385 billion in 2016.

After comparison and analysis, the prediction results of the recursive model are more ideal. Therefore, it is assumed that after the full implementation of the two-child policy, the annual fertility rate is,

$$b(n) = 1.2916\%$$

According to the population recurrence relations shown in Equation.

$$y(n+1) = y(n) + y(n+1)[b(n+1) - d(n+1)].$$

The recursion model of total population established is shown in Equation.

$$y(n+1) = \frac{y(n)}{[1-b(n+1)+d(n+1)]}$$

By using this model and MATLAB software programming, the total annual population of China in 2017-2030 can be predicted, shown in Table 3.

year	2017	2018	2019	2020	2021	2022	2023
Total number of people (10,000)	139016	139787	140558	141327	142095	142860	143624
year	2024	2025	2026	2027	2028	2029	2030
Total number of people (10,000)	144386	145146	145903	146658	147410	148160	148907

Table 3 China's annual population in 2017-2030

# 5. Summary

Based on the total population of China collected from 1996 to 2015 in this paper, the recursive model and the regression analysis model are established. By using the least squares algorithm, the Logistic model and curve fitting methods, the functional relationship between mortality and time is obtained. Thus, the annual mortality after 2016 is predicted, and finally the total population of China under the universal two child policy is predicted.

# Acknowledgements

Fund Project: This work is the phased result of the research of "The Practical Study of the Second Classroom of Applied Undergraduate Mathematics and the Cultivation of College Students' Innovative Ability" (("Twelfth Five-Year Plan" of Education Science Project in Shandong Province, Code: CBS15007) and "Adolescent Cybercrime Prediction Research Based on Quantitative Analysis" (Humanities and Social Sciences Research Plan Project in Shandong Province in 2016, Code: J16WB01)

# References

[1] Xian Wen, Haifeng Lei, Qiuhe Huang. Prediction of Real Estate Prices in Liuzhou Based on Polynomial Regression Model [J]. Marketing Research, 2013, 8: 53-54.

[2] Qingyang Li, Nengchao Wang, Dayi Yi. Numerical Analysis [M]. Beijing: Tsinghua University Press, 2001.

[3] Zhentao Li. Application of Improved Population Migration Algorithm in Function Optimization [D]. Hebei: North China Electric Power University, 2007.

[4] Qiyuan Jiang. Mathematical Modeling [M]. Beijing: Higher Education Press, 2011.

[5] Ganchang Wu. Probability Theory and Mathematical Statistics [M]. Beijing: China Renmin University Press, 2006.

[6] Pin Zhou, Zhengfeng He. MATLAB Numerical Analysis [M]. Beijing: China Machine Press, 2009.

[7] Hui-zhen Yan. Application of Logistic Model in Population Prediction. Journal of Dalian Polytechnic University, 2008, 27(4):333-335.