The Study for the Prediction Model of China Population Growth

Jingguo Qu
College of Light industry, Hebei Polytechnic University, 063000, China
Email: qujingguo@163.com

Yuhuan Cui\textsuperscript{1}, Aimin Yang\textsuperscript{2}, Yamian Peng\textsuperscript{2} and Huancheng Zhang\textsuperscript{1}
\textsuperscript{1}College of Light industry, Hebei Polytechnic University, 063000, China
\textsuperscript{2}College of Science, Hebei Polytechnic University, 063000, China
Email: yuhuan.cui@yahoo.com.cn

Abstract—China has a huge population. Now, it is facing a more complex trend of population change than the one in the 20th century. Population issue is the key factor restricting China’s development. As further slowing-down population growth, improving population quality, developing human resources, and properly solving the issue of population with overall consideration have become the key factor promoting China’s economic and social development, scientific and feasible prediction of the population growth trend is of great significance. First based on figures presented in the historical data, this paper first uses Malthus’ model of population growth as reference and establishes the mathematical model 1 about China’s population size and time by adopting the methods of regression and fitting; second through the adaptation of linear multiple regression method to make regression and fitting on the total number of population and the above four factors, a ternary linear regression model 2 is made; finally we find two factors: mortality rate, birth rate have great significance for the number of population, by using multiple regression, establishes the model 3.

Index Terms—Significance test, Fertility rate, Logistic regression, 0-1 dummy variable

I. INTRODUCTION

China is a developing country with the biggest population in the world. Population issue has been one of the key factors restricting China’s development. In recent years, China’s population change owns some new features, such as, speeding up of aging, constant rising of sex ratio at birth, urbanization of rural population and so on. All these factors influence China’s population growth. Various researches have been made and a large number of data have been collected about China’s population issue.

In accordance with basic national conditions and these above features of population growth, this paper creates a mathematical model for China’s population growth by using related data in the Appendix as reference (related literature and new data are also available) to predict the trend of China’s population growth and point out the advantages and disadvantages in this model.

Based on figures presented in the historical data, this paper first uses Malthus’ model of population growth as reference and establishes the mathematical model 1 about China’s population size and time by adopting the methods of regression and fitting:

Model (1) shows the growth rate of China’s population is changing during a certain period. To define certain factors which influence population growth, we adopt fitting in processing related data collected from year 1985-2005 and get the figures and fitting relations in terms of time and the following factors: fertility rate $x_1$, mortality rate $x_2$, urbanization $x_3$, birth rate $x_4$, proportion of aged population $x_5$, and sex ratio of population $x_6$. According to the conclusion of model (1) and being concerned about present state of China’s population, Significance Test about the above six factors is made through stepwise regression analysis. And results are as follows: levels of urbanization $x_3(t)$, proportion of aged population $x_5(t)$, sex ratio of population $x_6(t)$ have significant influences on population growth. Considering the relation between proportion of aged population and fertility rate and using Logistic regression model to connect them, we get the relation of the two factors:

$$x_5(t) = \frac{1}{1 + e^{-(0.3131t + 0.0360)}} .$$

Through the adaptation of linear multiple regression method to make regression and fitting on the total number of population and the above three factors, a ternary linear regression model (2) is made:

To predict the trend of China’s population growth, models are first verified. Take $t$ as 21 (2005-1985) and the predict number of population is 1.37154 billion. Compared with actual population number 1.30756 billion of 2005, the predict number is in the allowable error band. Take $t$ as 26 (year 2010) and $t$ as 36 (year 2020) as short-term and medium-term predicts, the population of 2010 would be 1.42226 billion and the population of 2020 would be 1.47243 billion. After a period of social evolution, China’s urbanization will reach the level of
middle-income developed countries’. So influence of urbanization tends to zero on population growth and model (3) is established which excluding urbanization and making regression analysis on other related factors. We find two factors: mortality rate \(x_2(t)\), birth rate \(x_4(t)\) have great significance for the number of population, by using multiple regression, the relations of total population, mortality rate and birth rate are:

Model (3) can be employed to predict the trend of population growth in a long-term. Take \(t\) as 66 (year 2050), the predict number of population would be 1.53044 billion.

II. EXPLANATION OF SYMBOLS

\(t\): Year

\(x_1(t)\): Average fertility rate of each woman at childbearing age in the year \(t\)

\(x_2(t)\): Mortality rate, the frequency of death among China’s population during the year \(t\)

\(x_3(t)\): Levels of urbanization during the year \(t\), Level of urbanization = total amount of urban population/total population

\(x_4(t)\): Birth rate in the year \(t\)

\(x_5(t)\): Proportion of aged population to the total population in the year \(t\), which is the proportion of senior citizens over 60 years’ old in the country’s total population

\(x_6(t)\): Sex ratio at birth in the year \(t\)

\(y\): China’s total population in the year \(t\)

III. MODEL HYPOTHESIS

1. All the factors influence population changes are defined under the average sense of entire social population;
2. No wars, epidemic diseases or natural disasters would influence total population change in quite a long certain period;
3. Take the population as a whole or a system for research;
4. Year 1985 is the starting year for research;
5. Aged population refers to population of age 60 and above;
6. The statistical data are collected among various groups of entire population.

IV. MODEL ANALYSIS AND SOLVING

Thomas Malthus released his work Population Theory in late 18th century and put forward the mathematical model about exponential growth of population. According to his model, population growth shows a pattern of geometric progression, that is, population growth rate function graph is a piece of straight line via the origin with constant value of slope.

Model I

To study the relation between China’s population and time hence make further investigation of China’s population change and birth rate of population, polynomial fitting method is adopted to make fitting analysis for given statistics. Changes of China’s total population during 1985-2005 are shown in Table I.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total population</th>
<th>Year</th>
<th>Total population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>105851</td>
<td>1997</td>
<td>123626</td>
</tr>
<tr>
<td>1989</td>
<td>112704</td>
<td>1998</td>
<td>124761</td>
</tr>
<tr>
<td>1990</td>
<td>114333</td>
<td>1999</td>
<td>125786</td>
</tr>
<tr>
<td>1991</td>
<td>115823</td>
<td>2000</td>
<td>126743</td>
</tr>
<tr>
<td>1992</td>
<td>117171</td>
<td>2001</td>
<td>127627</td>
</tr>
<tr>
<td>1993</td>
<td>118517</td>
<td>2002</td>
<td>128453</td>
</tr>
<tr>
<td>1994</td>
<td>119850</td>
<td>2003</td>
<td>129227</td>
</tr>
<tr>
<td>1995</td>
<td>121121</td>
<td>2004</td>
<td>129988</td>
</tr>
<tr>
<td>1996</td>
<td>122389</td>
<td>2005</td>
<td>130756</td>
</tr>
</tbody>
</table>

And we make a scatter diagram of total population and time. (Fig. 1)

It can be found in the above figure that the general trend of China’s total population is raising year by year. To accurately describe population growth rate, we use least square method to make polynomial fitting of these scatters, that is

\[ y = 103668 + 2314.54t - 150.664t^2 + 13.4359t^3 - 0.641633t^4 + 0.0110848t^5 \]  

(1)

See the fitting effect in Fig. 2

Make derivation of expression (1) and, Population growth rate is:

\[ y' = \frac{dy}{dt} = 2314.54 - 301.328t + 40.3077t^2 - 2.566532t^3 + 0.055424t^4 \]  

(2)
Make inference from expression (2), it can be seen that population growth rate function is not a piece of straight line via origin with constant value of slope. So Malthus’s model of population growth cannot completely match the prediction of China’s present population growth[3].

Through the analysis of statistics in Report on China’s Population Development Strategy Research, population growth rate is influenced by various factors[4]. In less than 30 years, China has achieved a historic change from a reproductive pattern of a high birth rate, low mortality rate and high natural increase rate to one of a low birth rate, low mortality rate and low natural increase rate. Total fertility rate dropped from 5.8 in 1970’s to 1.8 at present, which is below replacement level. China has being listed among countries with a low birth rate about more than half century’s earlier than those developing countries which also have population issues. For the present national conditions, factors influencing China’s population growth mainly includes birth rate, mortality rate, population aging, urbanization, etc. Analysis the influences of these factors on population growth can provide theoretical reference for population control.

Model 2

Followings are some new terms in model 2 and their explanations:

(1) Fertility rate: the ratio of total number of birth in one year to women at childbearing ages

Fertility rate = total number of birth in one year/total number of women at childbearing ages (women at childbearing ages refer to the group of women between the age 15 to the age 49)

(2) Mortality rate: the ratio of total amount of death in one year to average population in the same period

Mortality rate = total amount of death in one year/average population in the same period

(3) Birth rate: the ratio of total number of birth in one year to average population in the same period

Birth rate = total number of birth in one year/average population in the same period

(4) Urbanization: the process of farming population transferring to town

Levels of urbanization = total amount of urban population/total population

(5) Sex ratio at birth: the ratio of male to female

Sex ratio at birth = male / female ×%

(6) Proportion of aging population: the percentage of senior citizens over the age of 60 in the total population

To define certain factors which influence population growth, we first make separate analysis about the next six factors[5]: fertility rate, mortality rate, urbanization, birth rate, proportion of aging population and sex ratio one by one. Then we adopt polynomial fitting in processing related data from year 1985-2005 to get the relation of time and those factors. Next, by using stepwise regression method we screen some significant factors and finally establish the functional relation model about total population and related factors through multiple linear regressions.

1) Mortality rate: at present, health level of Chinese people has been listed in the premier position among other developing countries and above the one in middle-income countries. With the improvement of health care and medicine, mortality rate in China keeps dropping. Use least square method to make 5 times polynomial fitting for the data, the expression of mortality rate with time changes is as follow:

\[ x_2(t) = 6.89382 \times 10^{-3} - 3.21042 \times 10^{-5} t - 1.91029 \times 10^{-5} t^2 + 4.29741 \times 10^{-6} t^3 - 3.02777 \times 10^{-7} t^4 + 6.74117 \times 10^{-9} t^5 \]  

The fitting effect is shown in Fig. 3:

2) Birth rate: birth rate in China keeps dropping down as time changes (Fig. 3), which is mainly because of government control. Since the adoption of family planning policy in 1974, total trend of birth rate in China is falling down year by year, shown in Figure. 4. The relation of birth rate with time changes is:
\[ x_4(t) = 0.0230268 - 0.00064032t + 0.0000976841t^2 - 0.0000105614t^3 + 0.000000271805t^4 \] (4)

3) Aging of population: China has a large population. Influences of population structural imbalance on social stability are becoming serious. Meanwhile, China also has the largest number of senior citizens in the world. We establish a function \( x_5(t) \) of aging population and time \( t \) to study the influence of aging population on population growth. Through data analysis, China would become an aging society in 2000. Suppose aging as 1 and non-aging as 0, and take the features of 0-1 dummy variables as reference, so,

\[
x_5(t) = \begin{cases} 
0 & t < 16 \\
1 & t \geq 16 
\end{cases}
\] (5)

4) Sex ratio of population: the data from year 85-05 shows sex ratio varies as time changes. Suppose sex ratio of population in the year \( t \) is \( x_6(t) \) and make polynomial fitting, then we can get:

\[
x_6(t) = -0.0005t^4 + 0.0204t - 0.2526t^2 + 0.7632t + 106.32
\] (6)

Sex ratio varies with time is shown in Fig. 5:

The above figure shows the sex ratio curve changes as time changing from 1985-2005 and is in the shape of saddle. Take year 1994 (the 10th year) as a boundary point, sex ratio drops as total population grows during the first half of these years while sex ratio raises as total population grows during the next half range[6].

5) Level of urbanization: regional and industrial imbalance of population, widen of income gap between urban and rural areas, districts and fields would lead to various social conflicts, which seriously constrains the harmonious construction of China’s socialism. The analysis of population data from 1985-2005 shows the number of urban citizens is raising year by year. So the term “level of urbanization” is introduced (Level of urbanization = total amount of urban population/total population). It varies with time and directly influences China’s population change. Take level of urbanization as \( x_7(t) \) and make polynomial fitting, the expression would be:

\[
x_7(t) = 0.228459 + 0.0112352t - 0.0015104t^2 + 0.000123571t^3 - 0.0000025818t^4
\] (7)

Following is its variation trend

Consider the influences of above factors on the growth of population[7], by the adaptation of multiple linear regression, we get expression 11:

\[
y(t) = a_1x_1(t) + a_2x_2(t) + a_3x_3(t) + a_4x_4(t) + a_5x_5(t) + a_6x_6(t) + a_7x_7(t) + k
\] (8)

As too many factors are taken in when considering the causes of population growth and these factors may have multicollinearity[8], especially high interdependence among different explaining variable, the estimation result of coefficient would be unreasonable and the regression model is unavailable. To get an available regression model, we adopt stepwise regression analysis method. We first select birth rate, mortality rate, and fertility rate, proportion of aging population, sex ratio and urbanization as major factors of population growth to have incidence analysis and use matlab to solve it. The result is shown in Fig. 7.

By using stepwise regression, three variables \( x_3(t), x_5(t), x_6(t) \) are introduced into the model and variables \( x_1(t), x_2(t), x_4(t) \) are deleted, the expression of the model would be:
For parametric estimation values of model variables, see Table 2:

TABLE II
RESULTS OF STEPWISE REGRESSION

<table>
<thead>
<tr>
<th>Year</th>
<th>Total population</th>
<th>Year</th>
<th>Total population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>105851</td>
<td>1997</td>
<td>123626</td>
</tr>
<tr>
<td>1989</td>
<td>112704</td>
<td>1998</td>
<td>124761</td>
</tr>
<tr>
<td>1990</td>
<td>114333</td>
<td>1999</td>
<td>125786</td>
</tr>
<tr>
<td>1991</td>
<td>115823</td>
<td>2000</td>
<td>126743</td>
</tr>
<tr>
<td>1992</td>
<td>117171</td>
<td>2001</td>
<td>127627</td>
</tr>
<tr>
<td>1993</td>
<td>118517</td>
<td>2002</td>
<td>128453</td>
</tr>
<tr>
<td>1994</td>
<td>119850</td>
<td>2003</td>
<td>129227</td>
</tr>
<tr>
<td>1995</td>
<td>121121</td>
<td>2004</td>
<td>129988</td>
</tr>
<tr>
<td>1996</td>
<td>122389</td>
<td>2005</td>
<td>130756</td>
</tr>
</tbody>
</table>

Goodness of fit of the whole model is $R^2 = 0.97236$, $F = 199.349$ and corresponding value of $P$ is 1.91958E-13, which is far below significance level $\alpha = 0.05$. The goodness of fit of the whole model is fine.

Then based on the given figures, we get:

$$y(t) = a_3 x_3(t) + a_4 x_4(t) + a_5 x_5(t)$$

+ $a_6 x_6(t) + k$  \hspace{1cm} (9)

For parametric estimation values of model variables, see Table 2:

$$y(t) = 143370 x_3(t) - 8660 x_5(t)$$

- $19400 x_6(t) + 280820$  \hspace{1cm} (10)

As the definition of aging population in Appendix 1, the main factor influencing aging is the slow-down of fertility rate. So employing statistic software Spss to establish the function of aging level and fertility rate, the expression would be:

$$X_5(t) = \frac{1}{1 + e^{-(0.0311\ln(x_3(t))) + 0.0363}}$$ \hspace{1cm} (11)

The relation of total population and level of urbanization, sex ratio and fertility rate is:

$$y(t) = 143370 x_3(t) - 8660 x_5(t) - 19400 x_6(t)$$

+ $280820$  \hspace{1cm} And

$$X_5(t) = \frac{1}{1 + e^{0.0311\ln(x_3(t)) + 0.0363}}$$ \hspace{1cm} (12)

Connect the relations of urbanization level, sex ratio, fertility and time, following is the curve of total population and time (Figure 8):

Separately take $t$ as 26 (year 2010) and $t$ as 36 (year 2020) as short-term and medium-term predict, the population of 2010 would be 1.42226 billion and the population of 2020 would be 1.47243 billion.

Model 3

When considering long-term predict, as growth of the economy, improvement of population quality, rationalization of population distribution and employment structure, China’s urbanization will reach the level of middle-income developed countries’. So influence of urbanization tends to zero on population growth therefore can be excluded. Use regression analysis again to solve model 3 and we get Fig. 9:

Through stepwise regression in Figure 9, two variables $x_3(t), x_4(t)$ are introduced into the model, and variables $x_1(t), x_2(t), x_5(t)$ are deleted, the expression of the model is:

$$y = a_2 x_2(t) + a_4 x_4(t) + k$$

For parametric estimation values of model variables, see Table 2:

Fit Goodness of the whole model is $R^2 = 0.823889$, $F = 26.5099$, and corresponding value of $P$ is 1.22333E-006, which is far below significance level $\alpha = 0.05$. Fit Goodness of the whole model is fine.

Then based on the given figures, we get:

$a_2 = -2863500$, $a_4 = 1059000$, $k = 105400$

The relation of total population, mortality rate and birth rate is:

$$y(t) = -2863500 x_3(t) + 1059000 x_4(t) + 105400$$
Take \( t \) as 66 (year 2050) for long-term predict of population trend, the predict figure of population in year 2050 would be 1.53044 billion.

Table III. Results of stepwise regression

<table>
<thead>
<tr>
<th>variable ( x(t) )</th>
<th>coefficient</th>
<th>( t )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1(t) )</td>
<td>-1.85185E+007</td>
<td>-2.4786</td>
<td>0.0240</td>
</tr>
<tr>
<td>( x_2(t) )</td>
<td>-1.07426E+006</td>
<td>-3.7635</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

VI. MODEL EVALUATION AND PROMOTION

Advantages
1. Predicting trend of population growth in the future through this model is feasible;
2. It is suitable to use Logistic regression model to connect proportion of aging population and fertility rate;
3. Model screens factors with significant effects through stepwise regression analysis method;
4. This model is also available for the prediction of other problems;
5. The relation between levels of aging population and total population is solved by using 0-1 dummy variable;
6. The method of stepwise regression can not only select and exclude factors with insignificant effects, but also evaluate the degree of regression.

Disadvantages
1. Curve fitting is an approximate process, which can not accurately describe the trend and changes of number of population and time;
2. For the limitation of data, the predict results can not be completely accurate;
3. For an actual population group, there are various factors influencing its population change. For only 6 factors are considered in this model, the predicted value and the actual value have some deviation.

VII. PROMOTION

Through the adoption of stepwise regression method, this paper establishes the predict model of China’s population growth. Mainly in accord with its hypothetic conditions, this model matches the features of China’s population growth and predicts population changes in the future. This regression model used in this paper is also available in the fields of medicine, forest management, group population prediction, etc.

VIII. CONCLUSIONS

Through the adoption of stepwise regression method, this paper establishes the predict model of China’s population growth. Mainly in accord with its hypothetic conditions, this model matches the features of China’s population growth and predicts population changes in the future. This regression model used in this paper is also available in the fields of medicine, forest management, group population prediction, etc.

APPENDIX PROGRAM

Program 1:

```plaintext
x={1,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21};
MatrixForm[x];
y={105851,112704,114333,115823,117171,118517,119850,121121,122389,123626,124761,125786,126743,127627,128453,129227,129988,130756};
MatrixForm[y];
LL0=Table[{x[[k]],y[[k]]},{k,1,18}];
```
Program 2:

Program 3:

Apply the fundamental model to the data.

ACKNOWLEDGMENT

This work is supported by the Scientific Research Fund Project (qy201004 and qz201006) of College of Light Industry, Hebei Polytechnic University.
Jingguo Qu: Male, born in 1981, Master degree candidate. Now he acts as the Math teacher in College of Light Industry, Hebei Polytechnic University. He graduated from Harbin University of Science and Technology, majoring fundamental mathematics. His research directions include mathematical modeling and computer simulation, the design and analysis of parallel computation, and etc.

Yuhuan Cui: Female, born in 1981, Master degree candidate. Now he acts as the Math teacher in College of Light Industry, Hebei Polytechnic University. She graduated from Yanshan University, majoring computational mathematics. Her research directions include mathematical modeling and computer simulation, the design and analysis of parallel computation, elastic problems numerical simulation, and etc.