A study of China's population forecast based on a combination model

Long Wang

School of Software, Jiangxi Normal University, Nanchang, 330022, China

Abstract: The population issue has always been one of the key factors affecting the development of China. Building a strong socialist country faces a serious situation with a large population. With the rapid development of China's economy and the implementation of the national population policy, the question of how many people there will be in China in the future becomes extremely important, and the establishment of a population model, which can forecast the number of people more accurately, is a prerequisite for controlling the population growth. Therefore, this paper roots out a model to predict the future population change in China by analyzing the data related to China's population.To address problem, this paper first establishes a logistic model, and a gray prediction model. Under the assumptions, the prediction accuracy of the models is analyzed according to the historical data of China, and the results show that the accuracy of the models is high. Then, the models are used to predict the population in 2030, and the corresponding results are obtained. In order to further improve the prediction accuracy, this paper adopts the combination prediction method with unequal weights to establish the combination prediction model, calculates the weights of the prediction models separately and obtains the prediction values of the combination prediction model, and the results show that the total population of China is expected to be 150,893,000 in 2030.

Keywords: Logistic model, gray prediction model, prediction

1. Introduction

1.1 Background

The national population census is organized by the state to conduct a comprehensive survey and registration of the existing population in the country generally,[1] household by household and person by person according to the law, and the focus of the census is to grasp the data of the change of the existing population, gender ratio and urban and rural population in each place so that the state can make the next development policy. Since the total population of China is very large, each national census work costs a lot of human, material and financial resources, it would be very meaningful if mathematical modeling could be used to give effective predictions of various data of the census.

Population forecasting, as an important direction of population research, has shown an increasingly important role in the process of social and economic development. So far, seven national censuses with modern significance have been conducted in China, in 1953, 1964, 1982, 1990, 2000, 2010 and 2020 respectively. [2] The main results of the seventh national census show that the population of China in 2020 was 141.178 million, compared with 133.97 million in 2010, an increase of 72.06 million, or 5.38%; the average annual growth rate was 0.53%. 2020 national population, the male population was 723,339,956, accounting for 51.24%; the female population was 688438768 people, accounting for 48.76%.

1.2 The problem to be solved in this paper

Using statistical data published on the website of the National Bureau of Statistics and a combination of reasonable and credible data that you find, answer through mathematical modeling.

(1) In 2030, what is the total population of China?

(2) What will be the percentage of males and the percentage of urban population in China in 2030?

(3) As China is now entering a new stage of development, is it appropriate to analyze China's national census at an average frequency of once every 10 years, taking into account the country's current

development trend? Whether the answer is yes or no, give reasons. In particular, if the answer is not appropriate, give the appropriate number of years.

2. Problem Analysis

In order to improve the prediction accuracy, we select three individual prediction models to predict the total population of China in 2030, and compare them with the corresponding prediction data by combining a large number of statistics on the total population collected from the website of the National Bureau of Statistics. The prediction accuracy of these three methods is found respectively, and then the combined prediction model is built according to the principle of the smallest sum of squared errors, and finally the prediction value of the combined model is obtained. The individual forecasting models selected are: the gray forecasting model, the logistic model in accordance with the species growth trend, and the BP neural network forecasting model.[3]

3. Problem Assumptions and Notation

3.1 Problem Assumptions

(1). No consideration is given to the in-migration and out-migration of domestic and foreign populations.

(2). The effect of war, disease, and disaster on population size is not considered.

(3). The population distribution of different age groups does not change over time.

(4). Assume that the data obtained from the search are credible and reasonable.

3.2 Description of symbols

Table 1: Symbol Description

Serial number	Symbols	Description
1	$x^{(0)}$	Raw data sequence
2	λ_k	Raw data level ratio sequence
3	$x^{(1)}$	Raw data accumulation to generate sequences
4	a	Development coefficients of the whitening differential equation
5	b	The amount of gray action of the whitening differential equation
6	ε	Residuals of the grey prediction model
7	c	The posterior ratio difference of the gray prediction model

4. Problem Modeling and Solving

4.1 Population prediction based on gray prediction model

1) We take the total population from the fourth to the seventh census data as the original data (Table 2), and the total population units in this paper are: million people.[4]

Table 2: Total National Population, 1990-2020 Census

Year	1990	2000	2010	2020
Original data	116002	129533	137054	144350

(1) Then we can get the original data sequence

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4))$$

(1)

(2) cascade test on the original data columns

$$\lambda_k = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}$$

(2)

(3) If all falls λ_k in tolerable coverage $\left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}}\right)$ Interval.Then $x^{(0)}$ Can be used as a model $GM(1,1)$ of the data for gray prediction. Otherwise, the $x^{(0)}$ necessary transformations need to be made so that they fall within the allowable coverage.

2) $GM(1,1)$ Model building

(1) For the original reference data column $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n))$ Do 1 time

summation to get the cumulative series $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n))$.

(2) Construct the data matrix and data vector

$$B = \begin{pmatrix} -Z^{(1)}(2)1 \\ ... \\ -Z^{(1)}(4)1 \end{pmatrix}, Y = \begin{pmatrix} x^{(0)}(2) \\ ... \\ x^{(0)}(4) \end{pmatrix}$$

(3)

(3) Establish the grey differential equation.

$$x^{(0)}(k) + az^{(1)}(k) = b, k = 2, ..., 4$$

(4)

The corresponding whitening differential equation is

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b$$

(5)

3) Model solving

Based on matlab we can solve for a,b , So $u = (a,b)^T$. Generate series values

$$x^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a}, k = 2, 3, 4$$

(6)

4) Model accuracy check

Denote the residual equation as $\varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1, ..., n$, he variance of the $x^{(0)}$ original series and the residual series are s_1^2, s_2^2. The posterior test ratio difference is calculated as. Table 3 shows the reference table of accuracy inspection level.

Table 3: Reference table for accuracy inspection level

Model accuracy level	Mean Square Error Ratio
Level 1 (Good)	c<=0.35
Level 2 (pass)	0.35<c<=0.5
Level 3 (before)	0.5<c<=0.65
Level 4 (not qualified)	0.65<c

$GM(1,1)$ The accuracy test table shows that the model has high accuracy and credibility for future population trend prediction.

4.2 Solution based on gray prediction model

Using the gray prediction model we can obtain the predicted values of the total population between 1990 and 2020 at 10-year intervals and perform an error analysis with the true values (Table 3), here we can obtain a=-0.5447, b=117059.8713.[5]

$$u = \begin{pmatrix} -0.5447 \\ 117059.8713 \end{pmatrix}$$

Then the function can be found as where the posterior test ratio difference is, combined with the previous reference table of accuracy test level, it can be seen that the system has good prediction accuracy.

The relative errors between the population calculation and the actual data in 1990, 2000, 2010 and 2020 are calculated (Table 4), and the model prediction $GM(1,1)$ schematic for the time interval of 10 years.

Table 4: Table of model-based tests for total population projections

Year	True value	Model value	Relative error (%)
1990	113368	113368	0
2000	126583	126652	0.055
2010	133972	133743	0.171
2020	141178	141229	0.036

The results of this gray projection model to predict the future total population of the country are shown in Table 5 below

Table 5: Results of model-based projections of the total population of the country

Year	2030	2040
Total population	149136	157484

From the table, the model-based projection of the total population of the country in 2030 is 149.136 million.

4.3 Population projections based on logistic models

1) Logistic modeling

We assume that the population growth law is modeled according to the logistic growth law, and we take into account the natural factors and environmental capacity conditions that have a retarding effect on population growth. Let the equation be

$$\begin{cases} \dfrac{dx}{dt} = r(x)x \\ x(0) = x_0 \end{cases} \quad (7)$$

Where is the r(x) hysteresis function of natural factors and environmental and other conditions on population growth

Let the maximum number of people that can be accommodated by natural resources and environmental conditions be x_m, so $r(x_m) = 0$.

Then the equation can be found as

$$\begin{cases} \dfrac{dx}{dt} = rx\left(1 - \dfrac{x}{x_m}\right) \\ x(0) = x_0 \end{cases} \quad (8)$$

2) Solution of the logistic model

Table 6: Logistic model-based test for predicting the total population of China

Year	True value	Model value	Relative error (%)
2011	134735	134735	0
2012	135404	135316	0.065
2013	136072	135934	0.101
2014	136782	136596	0.136
2015	137462	137305	0.114
2016	138271	138063	0.150
2017	139008	138876	0.095
2018	139538	139747	0.150
2019	140005	140682	0.484
2020	141178	141686	0.360

According to the logistic model, we can get the predicted values from 2011 to 2020, and then compare them with the real values to get the following results, and calculate the relative error between the predicted and real data between 2011 and 2020 (Table 6).[6]

The table shows that the relative errors are within a small range, which indicates that this logistic-based population forecasting model is able to predict the total population of China in the short and medium term with a high degree of accuracy.

According to the logistic-based population projection model, we can obtain the projected values of our country-total population for the years 2021 to 2030 (Table 7)

Table 7: Prediction results of our total population based on logistic model

Year	Total population	Year	Total population)
2021	142765	2026	149556
2021	143926	2027	151261
2023	145177	2028	153109
2024	146526	2029	155117
2025	147982	2030	157302

From the table, it can be seen that the predicted value of the total population of China in 2030 based on the logistic model is 1,573,200,000.

5. Evaluation and promotion of the model

5.1 Advantages of the model

This paper uses multiple models when considering the problem. For example, to build a total population forecasting model for problem 1 to forecast the total population in 2030, this paper combines models: gray forecasting model, according to the principle of minimum sum of squared errors, and then forecasts the total population in 2030.[7]

Combined population forecasting models are used. Theoretical and empirical studies show that the combined model can obtain a better prediction value than any of the independent prediction values under different forecasting models and the same data sources, and the combined forecasting model combines various different types of forecasting models, which concentrates more information and forecasting techniques, and can reduce the forecasting error and significantly improve the forecasting effect.

5.2 Shortcomings of the model

This paper does not consider factors such as in-migration and out-migration, and only uses the existing population data to predict the future total, without taking into account the national conditions and characteristics of China's population growth.Due to the limitation of computer computing power, this paper cannot consider both the depth of data while considering the breadth of data.[8]

5.3 Model improvement and extension

In the subsequent research, in order to further balance the depth and breadth of the data, first we need to train the data set several times to continuously optimize the neural network model and enhance the

accuracy of optimizing our model.In order to make the population prediction more fully, we can further consider factors such as in-migration and out-migration, and combine our national conditions with population growth.

6. Conclusion

Logistic models, BP neural network models, and individual models were used to analyze the prediction accuracy under the assumptions and in accordance with the historical data of China, and the results showed that the accuracy of individual models were high. Then the models were used to predict the population size of China in 2030, and the corresponding results were obtained. The prediction accuracy is further improved and the results are closer to the real values.

References

[1] Tao X.X., Yang G.Y., Fang Y. Y. Prediction of total population in China based on combined prediction model [J]. Journal of Chongqing Institute of Science and Technology (Social Science Edition), 2013(07):99-103.

[2] Tian Zichen, Ji Gang, Liu Miao. Analysis and prediction of total population in Xinjiang based on improved gray GM (1, 1) model [J]. Mathematical Practice and Understanding, 2021, 51(05):258-264.

[3] Yang Zhenzhen, Liu Lin, Xie Yanqiu, Wang Miao. Research on the prediction of population aging development trend and coping strategy based on gray prediction model [J]. China Management Information Technology, 2021, 24(09):198-200.

[4] Chen Xia, Xiao Lan. Improvement of Logistic Model and Population Forecasting in China [J]. Journal of Chengdu University of Information Engineering, 2020, 35(02):239-243.

[5] Ma Xiaoxing. Census income prediction based on BP neural network [J]. Modern Computer, 2021(04):38-41.

[6] Xiao JC, Hong Han. Spatial distribution of population in urban clusters and urbanization evolution and development trend prediction [J]. Economic Journal, 2021(01):19-30+2.

[7] Zhang Shaoping. Prediction of population aging in China based on combined prediction model [J]. Party School of the CPC Qingdao Municipal Committee. Journal of Qingdao Administrative College, 2020(06):26-30+35.

[8] Zou DW, Ji JX, Zhang Y, Nie K. Correlation analysis of population influencing factors in China based on regression analysis and gray prediction [J]. Journal of Changchun Normal University, 2020, 39(06):10-17.