Health status assessment based on ARIMA time series and function fitting

Yadong Lu[1], Zimai Dong[2], June Huang[1]

[1]College of Science, China Three Gorges University, Yichang, Hubei 443002, China

[2]School of Transportation Engineering, Dalian Maritime University, Dalian, Liaoning 116026, China

Keywords: ARIMA Time Series Model, Function Fitting, GM (1, 1) gray model.

Abstract: This paper analyzes the health condition of China based on the data query and mining method. The ARIMA time series model is established and solved, and it was found that the aging population in China increased in the next three years, and the growth rate was still not decreased. The GM (1,1) gray model was solved, and it was found that the per capita disposable income of residents increased, but the per capita consumption in the proportion of available expenditure decreased. The incidence of male, female, and overall diseases in Beijing in the next three years was predicted by using the function fitting method.

1. Introduction

With the rapid development of China's economy and the aging population, the medical level is getting more and more attention. The increasing demand for medical care has also made it an urgent hope for improving the quality of medical care. At present, the private hospitals and the public hospitals have formed different levels of competition and cooperation. How to deal with the relationship between the private and public is related to the improvement of China's medical career. We should work together to promote the development of China's medical cause. This is not only related to the development of the national medical level, but also to the quality of our personal life.

2. Data Analysis

2.1 Analysis of China's Population Aging and Economic Situation

The following is an analysis of the income, age composition and economic development level collected by the National Bureau of Statistics [1].

The following table shows the population of elderly people over 65 years old in recent years, cf. Table 1.

Table 1. The Population of People over 65 (In thousands)

Year	2012	2013	2014	2015	2016	2017	2018
Population	127,140	131,610	137,550	143,860	150,030	158,310	166,580

Available from the data in the above table, the situation of China's aging population is severe. The data is visualized below, and the vertical line is drawn as shown below, cf. Figure 1.

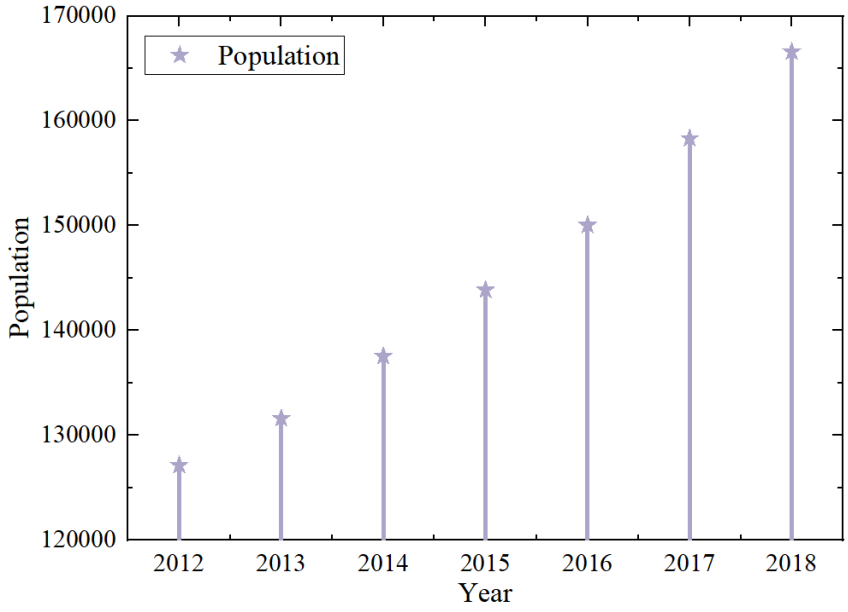

Figure 1. The Population of People over 65

From Figure 1, the following conclusions can be drawn:

●With the increase of time, the number of elderly people over 65 years old in China is increasing, which means that the population aging in China is grim.

●The growth trend is still rising at a constant rate and there is no saturation, which reminds us that aging is still a topic we need to pay attention to in the future.

In the next part, we will analyze the data of the per capita disposable income and GDP to understand the economic development of China in recent years. This is very helpful for the establishment of the model later, cf. Table 2.

Table 2. The Per Capita Disposable Income and GDP

Year	2013	2014	2015	2016	2017	2018
Per Capita Disposable Income	18,311	20,167	21,966	23,821	25,974	28,228
GDP (In billions)	59,296.3	64,128.1	68,599.3	74,006.1	82,075.4	90,031.0

The above table shows the changes in the number of the per capita disposable income and GDP in China from 2013 to 2018. The data source is the National Bureau of Statistics. From the data in the above table, we can roughly see that both the per capita disposable income and GDP are increasing. By 2018, the per capita disposable income reached 28,228 yuan. And GDP is up to 90,031.0 (In billions).

In order to get more information about the data, we visualize the data, draw a dotted line chart and a bar chart as shown below, cf. Figure 2.

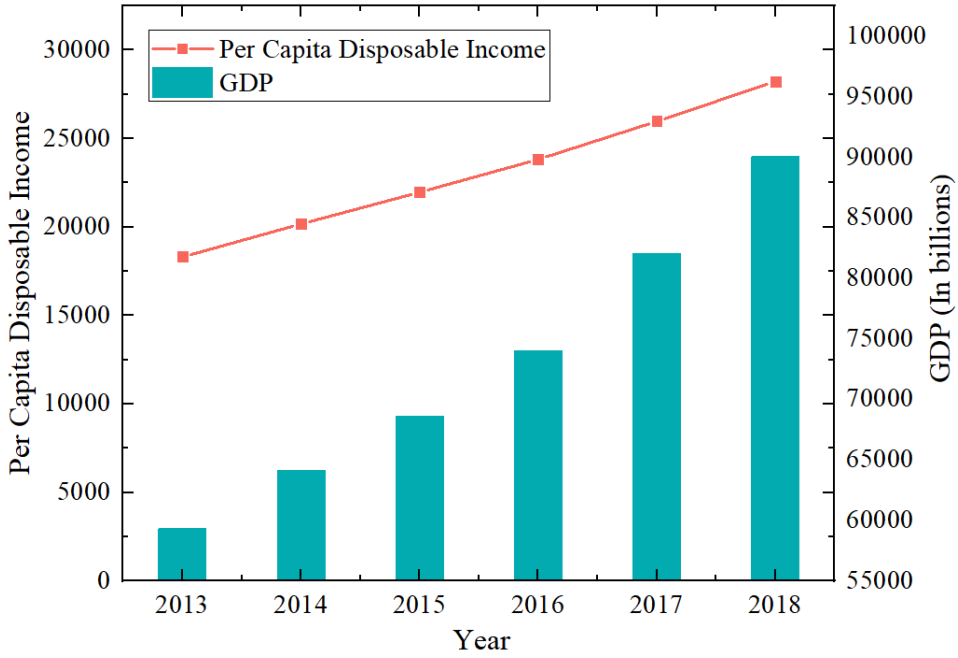

Figure 2. The Per Capita Disposable Income and GDP

From the above picture, we can draw the following conclusions:

●China's per capita income and GDP have been on an upward trend in recent years, especially the acceleration of GDP growth in the past two years, which means that China's economic strength has been further strengthened.

●The per capita income growth trend is relatively stable. GDP' s growth rate is faster than China's per capita income.

3. Predicting the Trend of Aging and Economic Needs of Residents

3.1 Model Establishment

In order to establish an ageing prediction model, data on the elderly population aged 65 and over from 2008 to 2015 were sought. In order to establish a model of economic demand forecast for residents, data on per capita disposable income, per capita consumption expenditure and per capita GDP of residents from 2013 to 2018 were collected.

3.2 Forecast of Aging Trends

In this paper, the autoregressive integral moving average model, *ARIMA* time series model [2], is used to predict the trend of population aging in China in recent years.

In the *ARIMA* (*p*, *d*, *q*) model in Question 1, *AR* refers to autoregression and p is an autoregressive term. *MA* is the moving average and q is the moving average. And d represents the number of times the original sequence needs to be differentiated in the process of converting the time series into a stationary sequence. The principle is to convert the non-stationary time series into a stationary time series and then regress the dependent variable only to its hysteresis value and the present value and hysteresis value of the random error term.

The mathematical expression of the model is as follows:

$$\left(1-\sum_{i=1}^{p}\varphi_i L_i\right)(1-L)^d X_t = \delta + \left(1+\sum_{i=1}^{q}\theta_i L_i\right)\xi_t \tag{1}$$

$(1-L)^d$ is the difference order, where L_i is the delay operator. $\left(1-\sum_{i=1}^{p}\varphi_i L_i\right)$ and $\left(1+\sum_{i=1}^{q}\theta_i L_i\right)$ representing the autoregressive coefficient and the moving average coefficient.

Since the number of aging populations is a typical stock indicator, it is usually second-order and monotonic, that is, after a second difference, the sequence is converted from a non-stationary sequence to a stationary sequence. Therefore, the order of the initial difference of the *ARIMA* model is second order.

Next, judge the sequence, The order of *p* and *q*. The appropriate *ARIMA* model [3] and the order of *p* and *q* can be selected by observing the autocorrelation coefficient of the second-order difference sequence and the truncation of the partial correlation coefficient. Using the *matlab* software to generate the population sequence of the second-order difference, the autocorrelation coefficient and the partial autocorrelation coefficient are visualized. It is easy to see that the autocorrelation coefficient and the partial autocorrelation coefficient of the sequence are rapidly reduced to zero near the first-order lag. Thus, it can be determined that the *p* and *q* value of the model is 1. That is, the *AR* term lag coefficient and the *MA* term lag coefficient of the sequence are both 1.

Therefore, using the *p, q* (1, 2, 1) model to predict China's aging trend, we can predict the population and total population of China over 65 years from 2017 to 2021. Comparing the actual data from 2017 to 2018 with the forecast data, we can see the prediction accuracy.

The population of China over 65 years old is represented by *E*, and the total population is represented by *T*. If the proportion of the population over 65 years old in the *i*-year is *e*, then the formula for *e* is:

$$e = \frac{E}{T} \tag{2}$$

In the *i*-year, the growth rate of China's population over 65 years old is *r*, then the formula is:

$$r_i = \frac{e_i - e_{i-1}}{e_{i-1}} \tag{3}$$

From the trend of growth rate, we can get the trend of aging.

3.3 Forecast of Residents' Economic Needs

The economic needs of residents are defined as the degree of residents' desire for economic income. If the country's economic development continues to rise, the degree of residents' desire for economic income will be higher. vice versa. The economic needs of residents can be measured by the economic demand index of residents.

The economic demand index of residents can be calculated through the per capita disposable income of residents and per capita consumption expenditure. Among them, per capita disposable income is expressed by *I*, per capita household consumption expenditure is represented by *O*.

The economic demand index for the residents in the *i* year is defined as the per capita household consumption expenditure divided by the per capita disposable income, expressed by the letter *X*, which is calculated as:

$$X_i = O_i / I_i \tag{4}$$

From the data from 2013 to 2018, the corresponding economic demand index of residents can be calculated, and then the gray GM (1,1) model can be used to predict the economic demand index of residents in the next few years.

Set the original resident economic demand index sequence as:

$$X^{(0)} = \left\{ X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), X^{(0)}(4), X^{(0)}(5), X^{(0)}(6) \right\} \tag{5}$$

The original data is accumulated to generate a new sequence, and the whitening differential equation of the GM (1, 1) gray model is established:

$$\frac{dX^{(1)}}{dt} + aX = u \tag{6}$$

Where *a* represents the development of gray scale, *u* represents the endogenous control coefficient, and thus is available, the approximate data sequence is:

$$\hat{X}^{(0)}(t+1) = X^{(1)}(t+1) + X^{(1)}(t) \tag{7}$$

The above formula is the fitting sequence, and thus the future economic demand index sequence of residents can be obtained.

3.4 Model Solution and Result Analysis

Using Matlab software to solve the time series model and gray prediction model established by problem 1, we can get the number of aging populations, the growth rate of aging population in the next three years, per capita consumption of residents and economic demand, cf. Table 3.

Table 3. Forecast Result

Category	2019	2020	2021
The Proportion of the Aging Population to the Chinese Population	0.12382	0.12907	0.13489
Growth Rate of an Aging Population	0.04245	0.04511	0.04587
Per Capita Consumption of Residents	21463	23203	25066
Economic Demand (Consumption/Expendable Consumption)	0.70020	0.69620	0.69230

3.5 Targeting the Aging of China's Population

Combining China's current situation: China's all-round development has improved people's living standards and health care levels, so the life expectancy of Chinese people has been extended, but the strength of China has led to increased social pressure. Young people are more likely to choose less and not to have children, which seriously affects the age composition of China has made China's population aging more and more serious.

Here are some solutions:

●Integrate the implementation of the healthy aging strategy into long-term planning, implement it in local development plans, and improve the social security system for the elderly.

●Improve the home care environment, accelerate the construction of community aged services, and moderately develop public welfare facilities.

3.6 For the Economic Needs of the Chinese People

China's economy is developing rapidly, and people's living standards have improved. At the same time, prices are rising. Faced with high consumption, such as social housing, adoptive children, and schooling, it is an inevitable problem that oppresses people's consumption patterns. Although the per capita consumption level is on the rise, the quality of consumption is still in a downturn.

4. Analysis of the Most Common Diseases in Beijing in the Future

4.1 Data Collection and Processing

Through the Information Center of the Beijing Municipal Health and Health Commission [3], collect the incidence of various diseases in Beijing from 2014 to 2018, separate them by gender and collect data, standardize them, and then draw them to obtain the incidence of three diseases. The trend is as follows, cf. Figure 3.

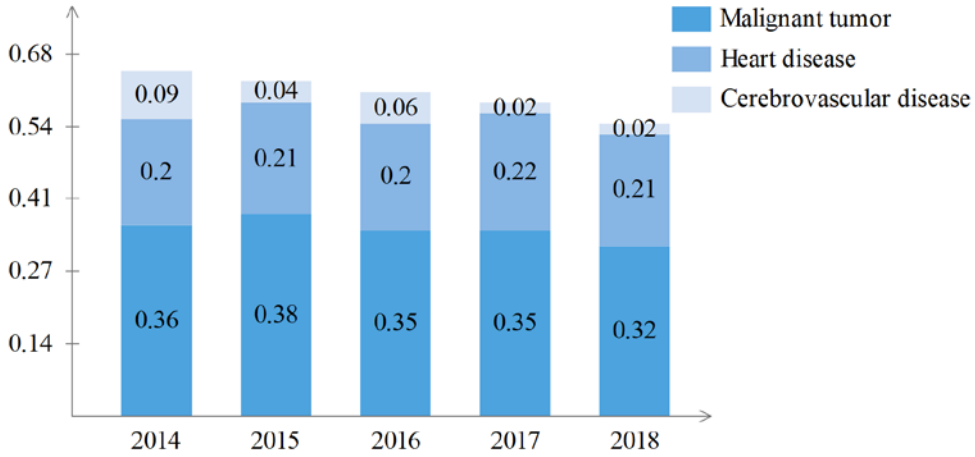

Figure 3. The Incidence of Three Diseases

From the above figure, the following conclusions can be drawn:

Malignant tumors, heart disease and cerebrovascular disease were among the top three male mortality rates in 2014-2018. The status of each of the three has not changed, and it is very stable, but the data is slightly different.

4.2 Finding the Relationship between the Incidence of Various Diseases

It is assumed that various diseases are represented by symbol x_i, the incidence rate is represented by y_i, and the year is represented by j (calculated from 2014, j=1 at this time), the following relationship can be obtained:

According to the actual situation, the sum of the incidence of these 9 diseases is not more than 1 each year. The formula is expressed as follows:

$$\sum_{i=1}^{9} y_{ij} < 1 \qquad (j = 1, 2, ...) \tag{8}$$

4.3 Predicting Future Disease Incidence

In this paper, for the incidence of various diseases, this paper adopts a polynomial function fitting method, taking the first type of disease x_1 heart disease as an example (the rest of the diseases are solved similarly), and the heart disease incidence function y_1 can be expressed as:

$$y_1 = a_1 x^m + a_2 x^{m-1} + \cdots + a_m x + a_{m+1} \tag{9}$$

Because the value of m is larger, the more accurate the data can be fitted, but the high-order polynomial fitting is not needed, which will destroy the numerical features. Therefore, the 4th, 5th, and 6th order polynomials are used for fitting, and finally the fitting is adopted. The most powerful function. Wherein, the fitting degree judgment function is:

$$K = 1 / \sum_{j}^{5} | \bar{y}_j - y_j | \tag{10}$$

That is, the reciprocal of the sum of the difference between the original data and the fitted data. The larger the *K* value, the better the fit.

4.4 Identify the Most Common Diseases

The most common disease in the future is the disease with the highest incidence. This paper defines the first type of disease that predicts the highest prevalence of various diseases in the next three years. The objective function is defined as:

$$MAX \quad \sum_{j=6}^{9} y_{ij} \qquad (i = 1,2,...,9) \tag{11}$$

4.5 Model Solution

The first question is that according to the statistics of the National Bureau of Statistics on the income of residents, the age structure of the population, and the level of economic development, it is reasonable to predict that the old papers in China will use the prevalence of various diseases in Beijing from 2014 to 2018 to classify the collected data. For male and female. A fitting function between the incidence and time of various diseases was obtained by using a function fitting method. According to the fitting function, the main diseases of male and female prevalence and their corresponding prevalence rates are as follows, cf. Table 4.

Table 4. Future prevalence

Prevalence	The first year of the future		The second year of the future		The third year of the future	
	Male	Female	Male	Female	Male	Female
Malignant tumor	27.531	24.023	27.208	23.804	26.885	23.585
Heart disease	24.478	29.189	24.568	29.59	24.68	29.991
Cerebrovascular disease	17.996	17.564	17.518	16.956	17.04	16.348

Subsequently, in order to select the most common diseases in the future, according to the definition of the objective function, the group with the largest prevalence rate in the next three years should be selected, and the most common diseases of men, women and groups in the future are calculated, cf. Table 5.

Table 5. Most common disease

Prevalence	Male	Female	Group
Most common disease	Malignant tumor	Heart disease	Cerebrovascular disease
Average prevalence	27.208	29.59	27.079

It can be seen from the table that in the future, the most common disease in men is malignant tumor, while women are heart disease. Although there are differences between the two, heart disease is the most common disease.

5. Result Analysis

In order to make the prevalence of major diseases of men and women clear in the next few years, the data will be visualized, and the average number of major diseases of male and female in the next three years will be calculated as follows, cf. Figure 4.

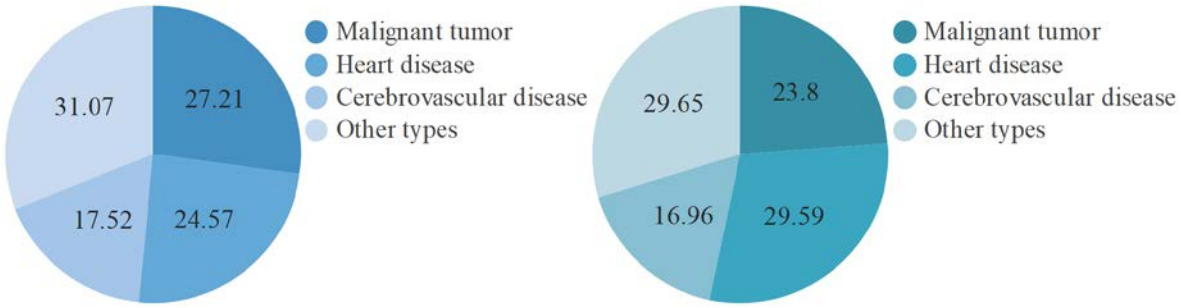

Figure 4. Average number of major diseases

It is evident from the figure that heart disease, malignant tumors and cerebrovascular diseases occupy most of the male or female. And heart disease and malignant tumors are basically the same, but heart disease is still slightly better.

References

[1] National Bureau of Statistics. http://www.stats.gov.cn/tjsj/. [Online; accessed 17-November-2019].

[2] Research and Application of Short Time Series Prediction Model Based on ARIMA Model [J]. Zhang Xiaofei, Tian Jinfang. Statistical Education. 2006 (10).

[3] Beijing Health and Health Commission Information Center. http://www.phic.org.cn/sy/. [Online; accessed 17-November-2019].