Prediction of Measles Epidemic Trend in China

Hui DUAN¹, Hui-rong YANG¹ and Yan PENG²,*

¹College of Mathematics and Science, Capital Normal University, Beijing, China 100048
²School of Management, Capital normal University, Beijing, China 100048

* Corresponding author

Keywords: Measles, Time series analysis, Prediction.

Abstract. Measles is an acute respiratory disease with high mortality and strong infectivity. Based on the national measles report from 1950 to 2014, this paper established a measles incidence prediction model to provide reference for measles prevention and control decision-making. We selected the incidence data from 1996 to 2014 as the training sample, while data of 2015 as the test sample. A forecasting model of annual measles incidence in China was established by time series analysis, and the incidence of measles in 2016-2018 in China was forecasted. Finally, the changes of measles incidence over time and its distribution in different regions were discussed. It was found that the incidence of measles was higher in western Region before 2004, while the incidence of measles in eastern and northeastern regions was lower, but the measles epidemic in eastern region began to recover in 2005, and the reasons of regional difference were excavated. The result showed that the ARIMA (1,2,1) model was feasible and effective in predicting the incidence rate of measles.

Introduction

Measles is an acute respiratory infectious disease caused by the measles virus. It is characterized by fever, inflammation of the upper respiratory tract, eye conjunctivitis and red papules on the skin and measles mucosal spots on the buccal mucosa. Retrograde pigmentation with bran-like desquamation is the main clinical feature with strong infectivity[1]. The elimination of measles is one of the most urgent tasks in infectious disease control before the Chinese government. At present, the incidence of measles is still very high in some parts of the country. There is no comprehensive systematic or authoritative research on measles immunization in different areas and people in China. If we can grasp the law of measles occurrence and development. To forecast the development trend of measles in the future and to take appropriate measures will be beneficial to the prevention and control of measles[2]. Therefore, according to the data of measles incidence and death report from 1950 to 2014, this paper carries on the time series analysis to forecast the measles incidence trend, and provides the reference basis for the measles prevention and control decision.

Method

Data Sources

The data are based on national measles reports from 1950 to 2014 provided by the Public Health Science Data Center, including statistics on reported morbidity and deaths across the country and provinces over the years. The data set mainly contains the administrative division code, year, region, and number of cases per year (representing the number of new measles cases in the population of the region within one year), the number of deaths, the number of deaths in one year, The number of new deaths from measles in the population of the region, the incidence rate (which represents the number of measles cases per one hundred thousand people in the region within a year) and the mortality rate (representing the number of deaths per one hundred thousand people in the region within a year) These variables are the number of deaths caused by measles.
Model Principle

In order to predict the trend of measles epidemic, RStudio software is used for time series analysis to establish the ARIMA prediction model and to process and analyze the data by using the relevant statistical methods.

General linear process \(\{Y_t\}\) can be expressed as a weighted linear combination of present and past white noise sequences.

\[
Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots.
\]

And satisfy \(\sum_{i=1}^{\infty} \varphi_i^2 < \infty\). Which \(Y_t\) is observations of the \(t\) time at the moment, \(\varepsilon_t\) is unobserved white noise of the \(t\) time, \(\varphi_t\) is white noise weigh at the \(t\) time. When the finite number of coefficients \(\varphi\) is not zero, we obtain the moving average process. At this point, change the symbol slightly and write down the weight of white noise at a certain time.

\[
Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}.
\]

We call this equation a q-order moving average process, brief account MA(q). Using one, \(-\theta_1\), \(-\theta_2\), \ldots, \(-\theta_q\) as the weight average of \(\varepsilon_t\), \(\varepsilon_{t-1}\), \(\varepsilon_{t-2}\), \ldots, \(\varepsilon_{t-q}\), to get \(Y_t\). Then slide the weights to \(\varepsilon_{t+1}\), \(\varepsilon_t\), \(\varepsilon_{t-1}\), \ldots, \(\varepsilon_{t-q+1}\) to get \(Y_{t+1}\). The term sliding average is derived from this. The moving average model was first studied by Slutskysky (1927) and Woldberg (1938). However, \(t\) time observation is not necessarily determined by white noise of several order delay, but also by several observations before \(t\) moment. The model is a regression of \(t\) time observation by using the observed value before \(t\) moment. It is called an autoregressive process. The p-order autoregressive process satisfies the equation:

\[
Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t.
\]

One of the \(\varepsilon_t\) independent from the \(Y_{t-1}, Y_{t-2}, Y_{t-3}\). Yulean (1926) has carried on the initial research to the autoregressive process. In some practical applications, the dynamic structure of data is described only by AR or MA models, which tend to be of higher order. With more parameters to estimate, the problem becomes cumbersome. Therefore, American statisticians Box, Jenn, Reinsel (1994) proposed a self-regressive moving average model, the basic idea of which is to combine the idea of the AR and MA models in a compact form so that the number of parameters used is kept to a minimum. The general form of the (p, q) models can be expressed as:

\[
Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}.
\]

The formula is a white noise sequence. Both \(p\) and \(q\) are nonnegative integers, the AR(p) and MA(q) models can be regarded as the special forms of ARMA(p,q). The ARMA model is aimed at stationary sequence, but it cannot be described directly by ARMA model for non-stationary time series. Only after logarithmic or differential processing a stationary sequence can be applied to the ARMA model. For this kind of non-stationary time series with short term trends, an autoregressive summation moving average model ARIMA(p,d,q) is introduced, where \(d\) is the number of difference. After the difference, it becomes a stationary ARMA(p,q) model. The simplest equation of this model is ARIMA(1,1,1), and equation of the model is:

\[
(1 - B)(1 - \varphi B)Y_t = (1 - \theta B)\varepsilon_t.
\]

Where \(B\) is the lag factor.
Data Preprocessing

The time series data of the observed system are obtained. It can be seen that the incidence of measles fluctuated greatly before 1970 and decreased significantly with time after 70 years. So, we consider using Arima model to fit and predict the incidence of measles. For getting overall situation of measles incidence in China, we selected several time periods for modeling and forecast. The incidence data from 1996 to 2014 is used as training data, and data from 2015 used as test data.

Identification and Order of Model

The sequence diagram shows that the incidence data show a certain fluctuation and a downward trend, and the statistical test also shows that the time series is non-stationary. After the second order difference of the time series, the sequence is shown to be stationary and non-white noise.

Figure 1. Measles incidence time series incidence in China from 1996 to 2014.

Figure 2. Fitting and forecasting map of measles incidence in China from 2016 to 2018.

In order to obtain the autocorrelation coefficient ACF and partial autocorrelation coefficient PACFs for stationary time series, the best order p and order q are obtained by analyzing autocorrelation graph and partial autocorrelation graph. The autocorrelation function and partial autocorrelation function graph are truncated after the first order. Combined with the previous data, the prediction model is determined to be ARIMA(1,2,1).

Figure 3. ACF function graphs.  
Figure 4. PACF function graphs.
Prediction and test of model

According to the established incidence model ARIMA(1,2,1), the incidence rate of the three years from 2016 to 2018 is forecasted. Using the RStudio software directly produces the following results. The asterisk is a sequence of observed values, the solid line is a fitting value, and the dashed line is a confidence line with a confidence level of 95%. Compared with the actual value, the incidence trend is basically consistent and reasonable, but there is still room for improvement to some extent.

The effects of the algorithms in RStudio shown as Table 1.

<table>
<thead>
<tr>
<th>function</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>meanf(train)</td>
<td>4.020202</td>
</tr>
<tr>
<td>naive(train)</td>
<td>1.969987</td>
</tr>
<tr>
<td>snaive(train)</td>
<td>0.9408461</td>
</tr>
<tr>
<td>rwf(train, drift=T)</td>
<td>2.720926</td>
</tr>
<tr>
<td>ses(train, initial='simple',alpha=0.2)</td>
<td>3.196338</td>
</tr>
<tr>
<td>holt(train, initial=&quot;simple&quot;,damped=F,beta=0.65)</td>
<td>7.425547</td>
</tr>
<tr>
<td>ets(train)</td>
<td>4.020202</td>
</tr>
<tr>
<td>auto.arima(train)</td>
<td>2.160718</td>
</tr>
<tr>
<td>arima(train, order = c(1, 2, 1))</td>
<td>0.6990417</td>
</tr>
</tbody>
</table>

The results in table 1 show that the standard error of ARIMA (1,2,1) model is the lowest. The accuracy of the predicted value and the real value is higher, and the overall fitting effect is better than other models. It shows that the model is reasonable and suitable for the practical problems in this paper.

Results

Forecast Result

The predicted values of measles incidence in China in 2016-2018 are as follows: 3.42 per one hundred thousand, 3.27 per one hundred thousand and 3.12 per one hundred thousand respectively. The incidence of measles in the country has been decreasing in the past three years, and it can be seen that the incidence of measles in our country is better.

Time Distribution of Measles Epidemic in China

As a whole, since the founding of the people's Republic of China, the first three years of disease incidence have been 1432.41 per one hundred thousand in 1959, 1265.74 per one hundred thousand in 1965, and 167.47 per one hundred thousand in 1964. From 1950 to 1965, due to population growth, economic backwardness and imperfection of medical technology, the incidence of disease was relatively high. During the three years of severe economic hardship from 1959 to 1961, the national
food shortage and famine caused by the policy of the Great Leap forward Movement to develop industry at the expense of agriculture led to a significant reduction in the number of infants and young children. After 1960, the economy began to improve, the population rebounded, morbidity and mortality rates began to rise, and peaked in 1965. Since the introduction of measles vaccine (MV) in 1965, especially after the planned immunization was carried out in 1979, with the improvement of the quality of MV, the production of liquid vaccine was stopped in 1985. All of them were changed to freeze-dried vaccine) and the national unified immunization program (the two-dose immunization program for 8 months and 7 years old was stipulated in 1985). The control of measles has made great progress and changed the situation of natural epidemic before the use of MV.

The incidence of measles decreased steadily from 1977 to 1990. After the end of the Cultural Revolution, productivity gradually recovered. With the advent of reform and opening up, the economy, science and technology of our country have developed rapidly, the reform of the medical and health care system has been gradually improved, and the level of medical technology has been greatly improved. Morbidity and mortality continue to decline, showing a good trend. Since September 11, 2010, China will carry out a large-scale measles vaccine intensified immunization. By comparison, we can see that the incidence of measles in China was 2.85 per one hundred thousand in 2010, 0.74 per one hundred thousand in 2011 and 0.45 per one hundred thousand in 2012, showing a decreasing trend year by year.

Regional Distribution of Measles Epidemic in China

The characteristics of measles were highly distributed and local eruption coexisted. In the early days of the founding of the people's Republic of China, the level of economic development in the whole country was generally low. Therefore, in Shanghai, Beijing, Tianjin and other densely populated developed areas, the incidence rate is relatively high, while in the sparsely populated backward areas, the incidence rate is relatively low. After 1965, the country began to widely use measles vaccines. As a result, the incidence of disease in all provinces has decreased significantly. Especially in Beijing, Shanghai and other regions, measles vaccine coverage, morbidity and mortality are extremely low. Since the reform and opening up, due to the development of economy, science and technology, medical treatment, the national morbidity and mortality are generally reduced, and gradually show regional differences. In Xinjiang, the relatively backward areas of Tibet, Yunnan, Guizhou, and so on, due to their special geographical, economic and ethnic characteristics, the western minority areas have a large span between the east and the south, the climate and geomorphology are complex and the people live in a scattered area, and the transportation is inconvenient. Health services are relatively poorly accessible due to sparse distribution of medical networks, and the economic level of people in the western region is very low, especially in rural areas, where most areas are still in poverty. This prevents residents in these areas from using health services such as vaccinations in a timely manner, resulting in higher morbidity and mortality. In economically developed areas such as Beijing and Shanghai, morbidity and mortality has been basically controlled.

Summary

The time series model ARIMA model used in this paper is very simple. It only needs endogenous variables and no other exogenous variables. It is easy to understand and has mature software support and is easy to popularize. However, the time series data are required to be stable or stable after the difference, and a large number of historical data are needed to make the prediction more accurate. In essence, the model can only capture linear relations, but not nonlinear relationships. Therefore, if we want to consider more relevant factors, we also need to explore other infectious disease prediction models. The elimination of measles is a process that requires long-term efforts. From the analysis of this article, we can see that the incidence of measles has decreased significantly since the start of the vaccine. But measles transmission is very strong, and with the development and progress of modern society and the frequent turnover of people, measles epidemic cannot be confined to a certain area.
Measles prevention and control also because of the combination of regional characteristics, measures targeted to local conditions, cannot be generalized. In cold weather, people often move indoors. In order to reduce the incidence of disease to some extent, we should ensure indoor sanitation and air circulation. There is also no guarantee that measles outbreaks or outbreaks will not occur in areas where measles is better controlled. Therefore, measles should not only strengthen prevention and control in various areas, but also need to unite all provinces and cities to fight measles. The research in this paper has its limitations, and there are more potential laws to be explored. I hope that interested friends can further analysis for the cause of measles prevention and control in China put forward reasonable suggestions.

Acknowledgement

This work is supported by Project of Social Science Innovation Platform of Beijing Education Commission (No. 025185305000/035).

References


