Prediction of China’s mortality rate under 5 in 2027 based on ARIMA model

Catherine Lin, Hannah Zhao, Sunny Sun, Tina Tian

G11 AP

Beijing National Day School

1. **Research question:**

The research will focus on the prediction of China’s mortality rate under 5 in 2027 based on ARIMA model. Our group is interested in this topic because child mortality rate can be an indicator of the overall development level of a region, including medical technology, economics, and public security. And the result will indicate the overall level of social development in China. Society will make further reductions in child mortality be possible through improvements such as health of expectant mothers and new moms.

1. **Background:**

There are three relevant pieces of information that help us better predict future mortality rates (under 5) based on related factors and test models. The economic level is the most important factor in the mortality rate. Higher values of GDP always results in lower mortality rate because GDP directly relates to the nation’s medical level and sanitation condition, which means people in the country have more resources and capital to take care of the kids, as well as the society can invest more in developing medical technologies. For infants, birth defects, preterm birth and low birth weight, sudden infant death syndrome, injuries, and maternal pregnancy complications are all main reasons for infant death, which contribute to children's mortality rate. As the medical level improves, sick children can receive a cure in time, and children with certain disabilities can be diagnosed earlier and treated under better care. Malnutrition reduces the effectiveness of health interventions, tainted water sources result in diarrheal illnesses, and unsanitary behaviors make mothers and children more susceptible to illness. The second resource used the autoregressive integrated moving average (ARIMA) model to forecast the infant mortality rate. They examined the accuracy of the model and concluded that there would be a decreasing trend of IMR (2017-2025). The third related resource specified is the ARIMA model. ARIMA model analyzes and forecasts non-stationary time series data, and it applies to abundant real-life situations when having a data set with non-linear patterns. These credible resources offer a comprehensive exploration of the accurate forecast for China’s mortality rate under 5 and implement the result to reduce the burden of the infant mortality rate.

1. **Sampling and Experimental Design**

**Variables:**

*Explanatory variable:* time in years

*Response variable:* Mortality rate, number of children under-5 died per 1,000 live births, in China.

**Type of study:**

Our study is Observational study, and we decided to use ARIMA model to analyze data and seek general trend.

**Data Collection and inference:**

The World Bank gathered the data from the UN Inter-agency Group for Child Mortality Estimation. This UN group collected data from a large majority of the countries in the world, which includes up to 266 countries. And the original data consists of data from 1960 to 2021. However, because the data from 1960 to 1989 are fragmentary and do not include absolutely all the countries, we decided to use the year 1990 to 2021 in China to keep the data intact.

These data can be used to establish relationships between time and mortality rate in China only. It can also be used to seek the general trend and be used to predict the future of Chinese children mortality rate.

1. **Exploratory Data Analysis**

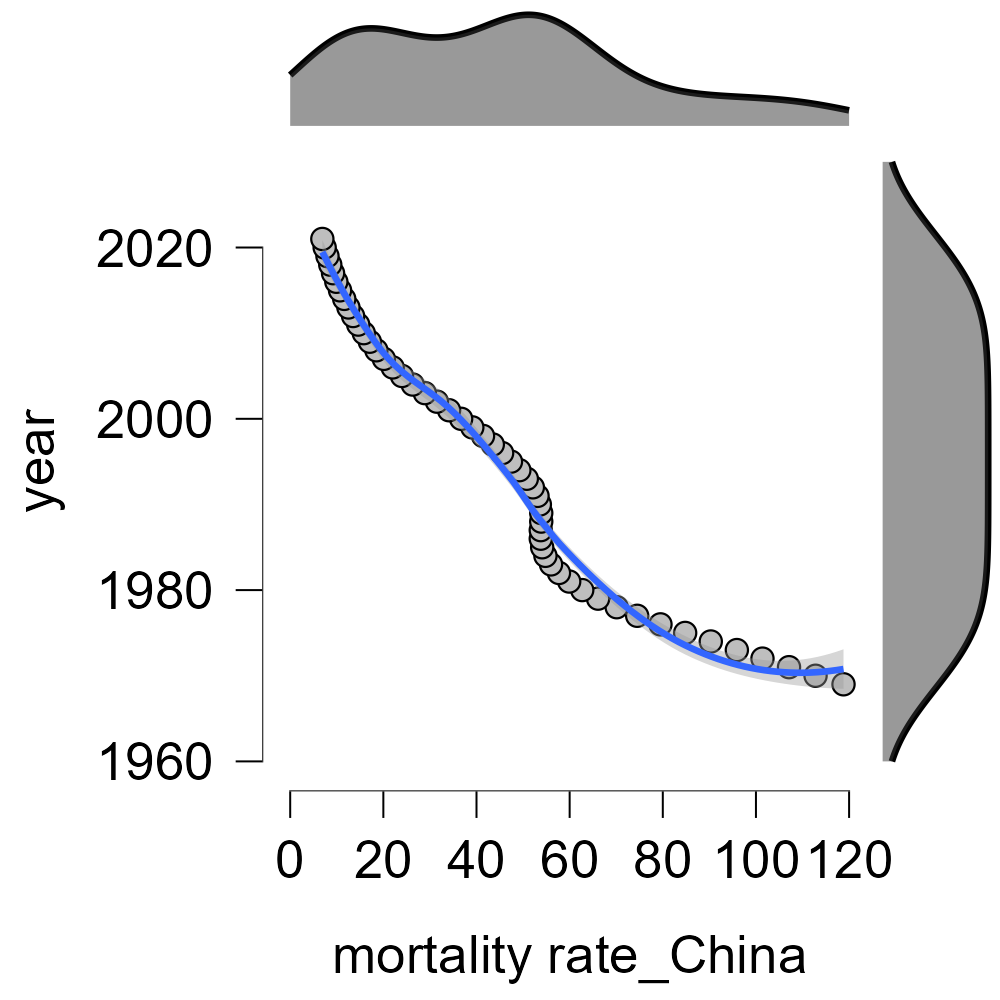
We used data on child mortality in China from 1969 to 2021 for subsequent predictive analysis. Table 1 below shows a one-way table of the data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | Std. Deviation | Minimum | Maximum |
| mortality rate\_China | 45.764 | 29.826 | 6.900 | 118.800 |

*Data source: World Bank*

**Table 1: descriptive statistic of child mortality in China from 1969 to 2021**

Next, the data is visualized into a scatter plot, where the horizontal coordinate is the death rate and the vertical coordinate is the year, and the images on both sides of the horizontal and vertical axes represent the density of the data. You can see that the mortality rate in China from 1969 to 2021 is generally declining.



*Data source: World Bank*

**Figure 1: Scatter plot of child mortality rate of China from 1969 to 2021**

Based on these preliminary results, we will use the ARIMA model for future predictive analysis. To meet the conditions for carrying out the ARIMA model, we have to make sure that the stochastic process is stationary stochastic, which means the value u must be a constant independent of time t, the variance must be a constant independent of time t, and the covariance rk  has to be a constant only related to the time interval K, independent of time t.

1. **Group Task Assignments and Timeline**

May 15th-17th: finish proposal

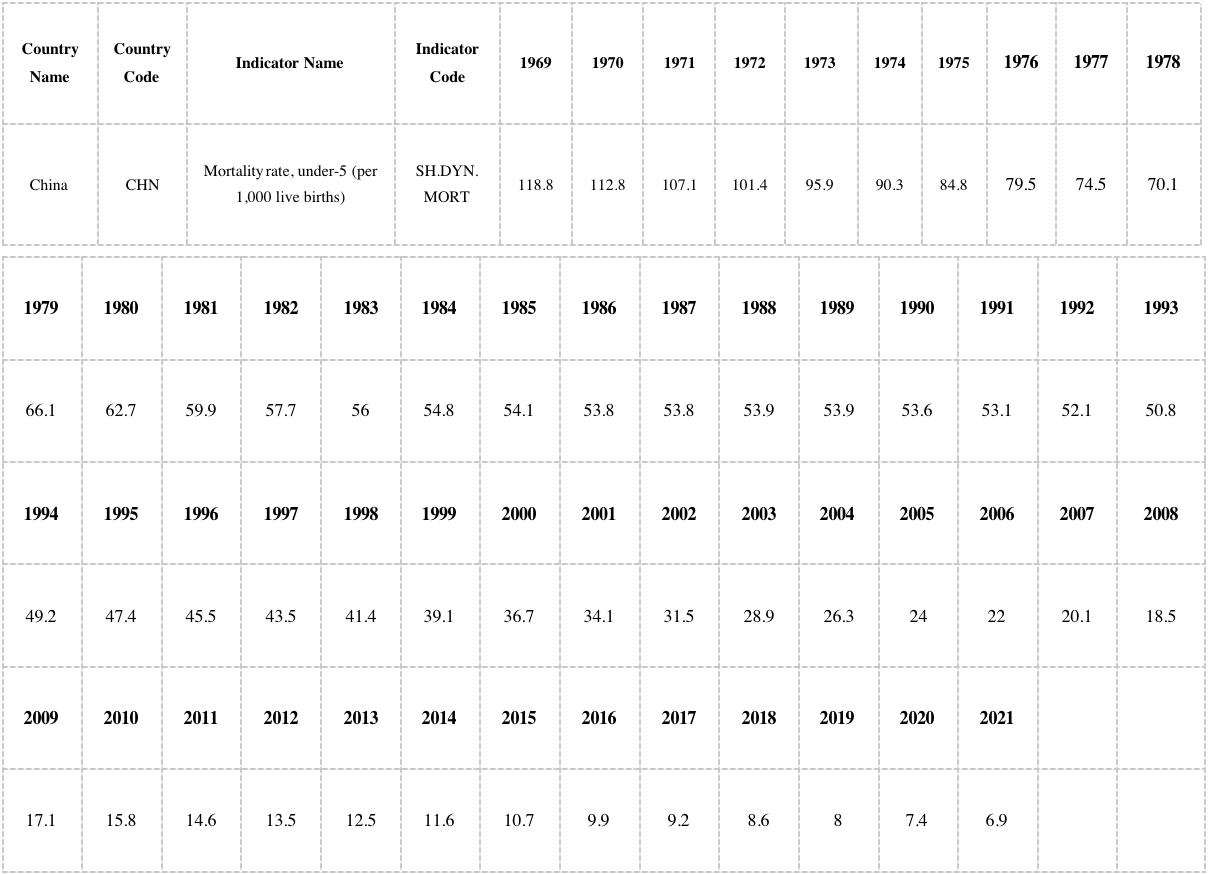
May 17th—20th: study ARIMA model to evaluate our data

May 21st—22nd: group discussion of “Introduction, problem, purpose”, analyze data

May 23rd—24th: finish the corresponding parts for each teammate

May 25th—26th: finish report, final check

1. **Data**

**

*Data source: World Bank*

**Table 2: original data about mortality rate under 5**

1. **Reference**

[1] United Nations. (n.d.). Education Is Key to Reducing Child Mortality: The Link Between Maternal Health and Education | United Nations. https://www.un.org/en/chronicle/article/education-key-reducing-child-mortality-link-between-maternal-health-and-education

[2] Infant Mortality | Maternal and Infant Health | Reproductive Health | CDC. (n.d.). https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm

[3] Mishra, A., Sahanaa, C., & Manikandan, M. (2019). Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model. PubMed, 26(2), 123–126. https://doi.org/10.4103/jfcm.jfcm\_51\_18

[4] Brwnlee, J. (2020). How to Create an ARIMA Model for Time Series Forecasting in Python. MachineLearningMastery.com. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/#:~:text=A%20popular%20and%20widely%20used,structures%20in%20time%20series%20data.

[5] Mishra, A., Sahanaa, C., & Manikandan, M. (2019b). Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model. PubMed, 26(2), 123–126. https://doi.org/10.4103/jfcm.jfcm\_51\_18

[6] Adhikari, R., & Agrawal, R. K. (2013). An Introductory Study on Time Series Modeling and Forecasting. In arXiv (Cornell University). Cornell University. <https://arxiv.org/abs/1302.6613>

[7] CME Info - Child Mortality Estimates. (n.d.). https://childmortality.org/data