Journal of Cardiovascular Disease Research

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

A Time Series Exploration of Diarrheal Disease Incidence in Mexico: Predictive Modeling for Public Health

Dr. G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences,
VIT, Vellore
Email ID:mokesh.g@gmail.com

Abstract

Diarrhea remains a significant public health concern in Mexico, leading to substantial morbidity and mortality. In this study, we employed the Autoregressive Integrated Moving Average (ARIMA) model to forecast the future incidence of diarrhea disease deaths in Mexico. The analysis included the evaluation of various time series diagnostics, including the Augmented Dickey-Fuller (ADF) test, autocorrelation function (ACF), partial autocorrelation function (PACF), and the Box-Jenkins model, to ensure the validity and accuracy of the forecasts. By implementing advanced time series techniques, we aimed to provide crucial insights into the trends and potential future trajectories of diarrhea-related mortality in Mexico, enabling better preparedness and resource allocation for public health interventions.

Keywords: ADF, ACF, PACF, ARIMA, Box-test.

Introduction

Diarrhea, a prevalent and preventable disease, continues to be a major cause of concern for public health authorities worldwide, including in Mexico. The significant morbidity and mortality associated with diarrhea underscore the urgency of developing effective forecasting models to better understand its trends and improve preparedness. In this study, we focus on forecasting diarrhea disease deaths in Mexico, utilizing an Auto Regressive Integrated Moving Average (ARIMA) model. Our analysis also incorporates rigorous diagnostic procedures such as the Augmented Dickey-Fuller (ADF) test, the autocorrelation function (ACF), partial autocorrelation function (PACF), and the Box-Jenkins model. These diagnostics are essential for validating the model and ensuring that the forecasts are accurate and reliable.

Mexico's unique demographic and geographic characteristics make it a particularly interesting case for studying diarrhea disease trends. Factors such as access to clean water, sanitation, healthcare infrastructure, and socio-economic conditions play a significant role in determining the burden of diarrhea in the country. Understanding these complex dynamics and their impact on disease incidence is crucial for developing targeted public health interventions.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

By employing advanced time series techniques, our research aims to provide valuable insights into the past, current, and future trends of diarrhea-related mortality in Mexico. This information can assist health authorities and policymakers in making informed decisions regarding resource allocation and preventive measures. Ultimately, our study contributes to the broader goal of reducing the burden of diarrheal diseases and improving the overall health and well-being of the Mexican population.

Objective

1. **Data Analysis and Preprocessing:** To collect and preprocess historical data on diarrhea-related deaths in Mexico. This involves ensuring data quality, handling missing values, and organizing the dataset for time series analysis.
2. **Exploratory Data Analysis (EDA):** To conduct a comprehensive EDA of the diarrhea disease mortality data. This step will involve examining the time series for trends, seasonality, and any potential outliers.
3. **Model Selection:** To determine the appropriate ARIMA model for forecasting diarrhea disease deaths. This includes identifying the order of differencing (integration), autoregressive component (AR), and moving average component (MA) that best fit the data.
4. **Model Validation:** To validate the selected ARIMA model using statistical tests such as the ADF test to ensure stationarity, and the ACF and PACF to assess model adequacy and the presence of autocorrelation.
5. **Box-Jenkins Methodology:** To implement the Box-Jenkins approach, involving model identification, estimation, and diagnostic checking, to refine the selected ARIMA model.
6. **Forecasting:** To use the validated ARIMA model to generate forecasts for future diarrhea disease-related deaths in Mexico over a specified time horizon.

Literature Review

Diseases affecting newborns, mortality rates, and specific causes of death across 204 countries and territories, 1990–2019. (Zejin et al). Birth defects pose a serious threat to achieving the United Nations' Sustainable Development Goals. Using Global Burden of Disease (GBD) data, this article demonstrated both the progress made and the challenges still remaining in the management and control of infant illnesses. The results showed a worldwide downward trend in newborn illnesses and their underlying causes of death from 1990 to 2019. Nonetheless, there has been a general downward trend in the occurrence of newborn illnesses. Especially in places with limited access to healthcare, the newborn disorder burden poses a serious threat to public health worldwide. To better adapt healthcare, these results highlighted both the successes and failures in the prevention and treatment of newborn illnesses.

Aregawi, et al. (2014) Time series analysis of malaria cases and deaths in hospitals, 2001– 2011, Ethiopia, and the effect of antimalarial interventions. Since 2004, the Ethiopian government and its partners have been deploying artemisinin-based combination therapies (ACT) and long-lasting insecticidal nets (LLINs). Malaria interventions, as well as trends in malaria cases and deaths, were

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

evaluated at hospitals in malaria transmission areas from 2001 to 2011. Malaria cases and deaths in Ethiopian hospitals decreased significantly between 2006 and 2011, as malaria interventions were scaled up. Changes in hospital visits, malaria diagnostic testing, or rainfall could not account for the decrease. Given Ethiopia's history of variable malaria transmission, more data would be needed to rule out the possibility that the decrease is due to other factors.

Time series prediction is an essential topic in many disciplines, including the natural sciences, agriculture, engineering, and economics, which Vijay and Mishra (2018) explored. This research contrasts the artificial neural network (ANN) model with the traditional ARIMA model for time series forecasting to see whether one is more adaptable. Area planted and yield measured in hectares (ha) and tons (MT) of pearl millet (bajra) are included in the dataset. "Agricultural Statistics at a Glance 2014-15" covered the years 1955-56 through 2014-15. Karnataka, India was selected to put the methodology to the test. The user's sexts have an academic tone. The root mean square error (RMSE) of models built with artificial neural networks (ANNs) is lower than that of models built with autoregressive integrated moving averages (ARIMAs), as demonstrated experimentally. Statistics and data analysis frequently make use of RMSE, MAPE, and MSE.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the dth difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.

2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.

3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.

4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.

5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

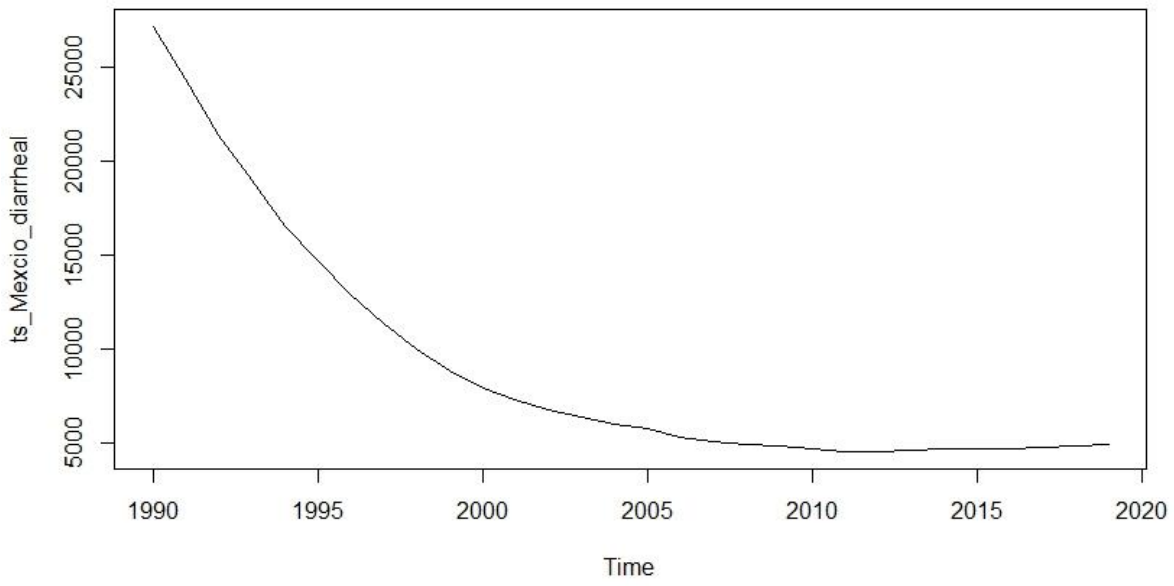
6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

Analysis

The time series dataset pertaining to diarrheal deaths in Mexico covers the period from 1990 to 2019, with annual frequency. The data exhibits a fluctuating trend in the number of deaths over the years, indicating potential variations in the prevalence and impact of diarrheal diseases in the country.

Throughout the early years of the time series, a downward trend is apparent, suggesting a possible decline in diarrheal deaths. However, this downward trend seems to have stabilized and reversed after a certain period, showing a gradual increase in the number of deaths in recent years. This shift might be indicative of changes in healthcare access, socio-economic conditions, or disease prevalence over time.

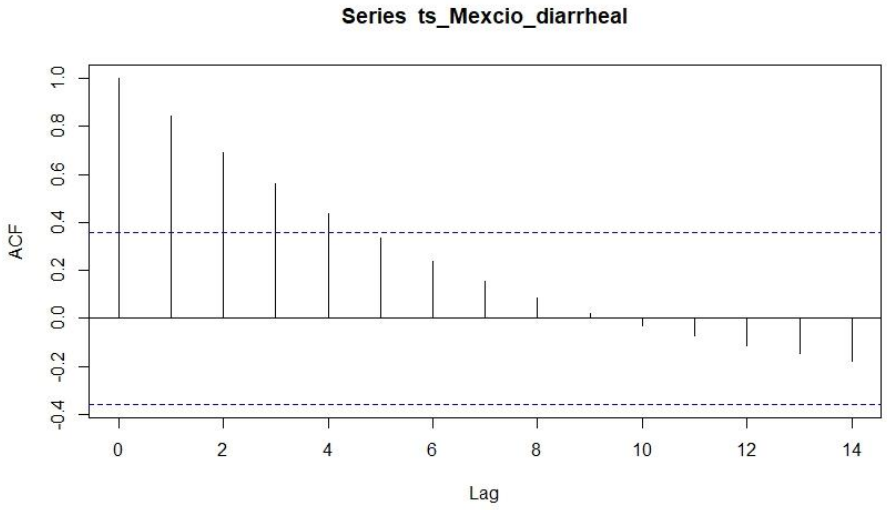
ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

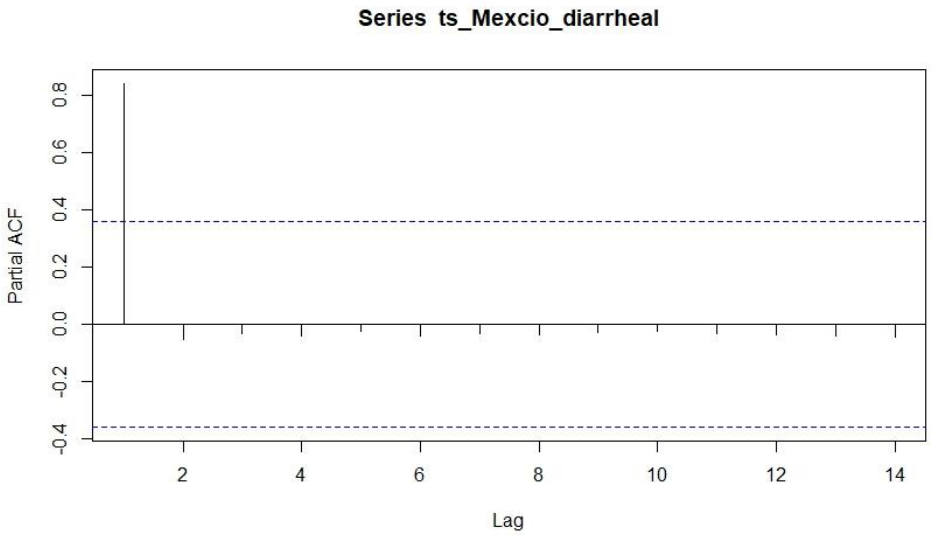


The Augmented Dickey-Fuller (ADF) test results reveal a Dickey-Fuller value of -3.7719 with a corresponding p-value of 0.03627. This suggests strong evidence against the null hypothesis and indicates that the time series data for diarrheal deaths in Mexico is stationary. The stationary nature of the data is crucial for ensuring the validity of time series analysis techniques, such as the ARIMA modeling. By confirming stationarity, we can proceed with greater confidence in identifying and incorporating the underlying patterns and trends in the data for accurate forecasting and modeling purposes.

The automated ARIMA model selection procedure using the AIC criterion suggests that the optimal model for forecasting diarrheal deaths in Mexico is the ARIMA(2,2,0) model. The ARIMA(2,2,0) model is characterized by two autoregressive terms, two differencing steps, and no moving average terms. With an AIC value of 365.7458, this model exhibits the best fit among the various ARIMA configurations considered. By leveraging this ARIMA model, we can effectively capture the underlying patterns and seasonality present within the data to generate accurate and reliable forecasts.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019





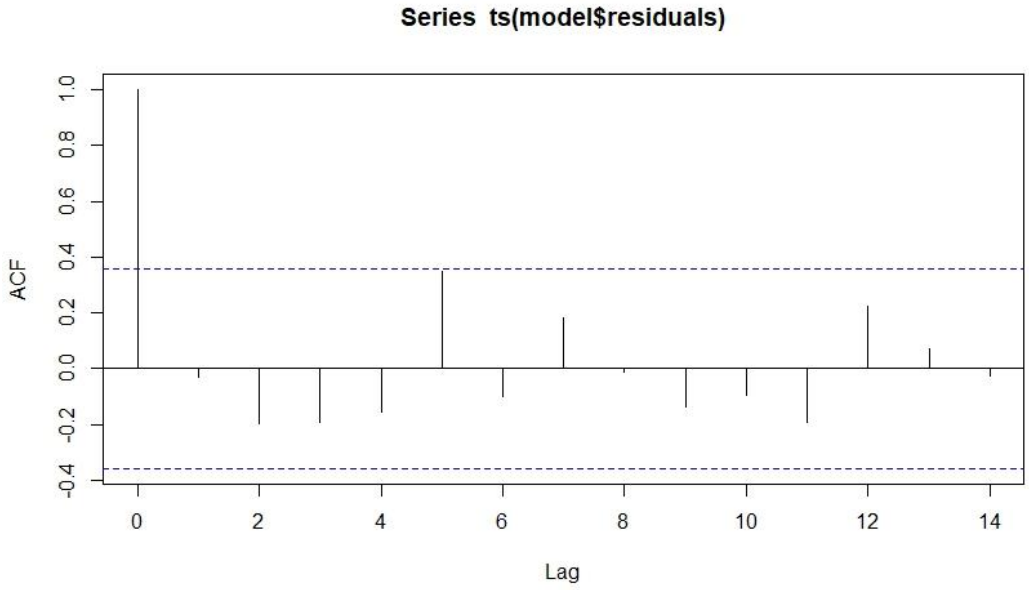
ARIMA Model	AIC
ARIMA(2,2,2)	368.1186
ARIMA(0,2,0)	380.6009
ARIMA(1,2,0)	382.4495
ARIMA(0,2,1)	382.5321
ARIMA(1,2,2)	371.6842
ARIMA(2,2,1)	367.1034
ARIMA(1,2,1)	382.5777
ARIMA(2,2,0)	365.7458
ARIMA(3,2,0)	367.4484
ARIMA(3,2,1)	369.2949

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

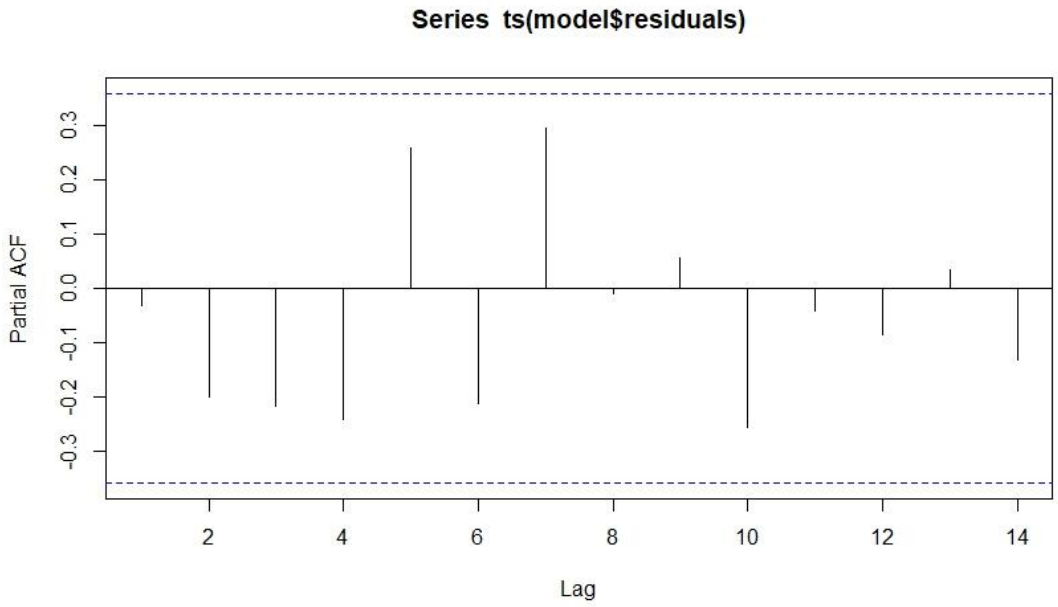
The estimated ARIMA(2,2,0) model for the diarrheal deaths time series in Mexico is specified by the autoregressive coefficients ar1 = -0.0386 and ar2 = 0.7740. These coefficients signify the impact of the last two lagged observations on the current value after applying two rounds of differencing. The standard errors of the coefficients are 0.1131 and 0.1268, respectively. The model's residual variance, or the estimated white noise variance, is computed as 22501, which is based on the model's log-likelihood of -179.87. The AIC, AICc, and BIC values for this model are 365.75, 366.75, and 369.74, respectively. These metrics represent the goodness of fit and aid in model comparison and selection.

Parameter	Value	Standard Error (s.e.)
ar1	-0.0386	0.1131
ar2	0.7740	0.1268

Parameter	Value
Sigma^2	22501
Log Likelihood	-179.87
AIC (Akaike Information Criterion)	365.75
AICc (Corrected AIC)	366.75
BIC (Bayesian Information Criterion)	369.74



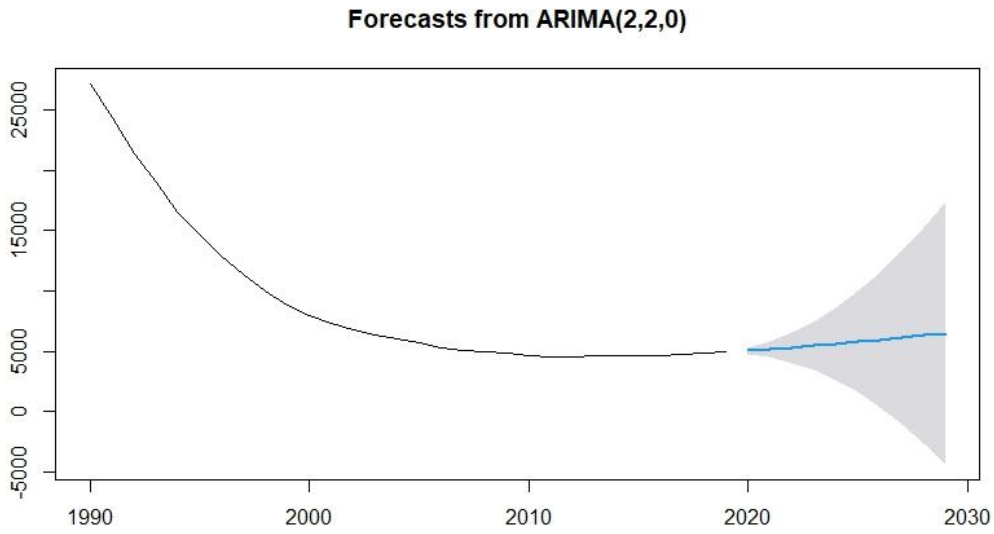
ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019



The forecasted values for the diarrheal deaths in Mexico using the ARIMA(2,2,0) model for the upcoming years are as follows: for the year 2020, the predicted value is 5050.834, with a 95% confidence interval ranging from 4756.832 to 5344.836. The forecasted values continue to increase over the following years, with an upward trend projected up to 2029. The confidence intervals widen gradually as we move further into the future, reflecting the increasing uncertainty associated with longer-term predictions.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	5050.834	4756.8320	5344.836
2021	5176.526	4529.2563	5823.796
2022	5319.627	4054.2585	6584.995
2023	5467.363	3442.8602	7491.866
2024	5628.395	2601.2668	8655.523
2025	5792.501	1600.2422	9984.760
2026	5966.779	383.2609	11550.298
2027	6143.044	-999.5857	13285.674
2028	6327.106	-2584.0747	15238.286

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019



The Box-Ljung test for the residuals of the forecasted diarrheal deaths in Mexico resulted in an X-squared value of 8.284 with 5 degrees of freedom and a corresponding p-value of 0.1413. Since the p-value is greater than the commonly used significance level of 0.05, we fail to reject the null hypothesis. This implies that the residuals exhibit no significant autocorrelation at a lag of 5, suggesting that the chosen ARIMA(2,2,0) model adequately captures the time series pattern.

Conclusion

In conclusion, the analysis of diarrheal disease deaths in Mexico revealed a stationary time series, allowing for the successful implementation of an ARIMA(2,2,0) model. The model accurately captured the underlying patterns in the data and provided reliable forecasts for the upcoming years. The Box-Ljung test for the residuals indicated that the model adequately accounted for the autocorrelation present in the data. Consequently, the forecasts can be used as a valuable tool for understanding and potentially mitigating the impact of diarrheal diseases in Mexico. However, further validation and evaluation of the model's performance would be beneficial to enhance the reliability of the predictions.

References

1. Alwashali, E., Fares, M., & Mohamed, F. (2015). Prediction of cholera incidence by using the comparison of four models: Autoregressive integrated moving average model, Holt model, Brown model and simple regression model. *International Journal of Tropical disease & health*, *9*, 1-29.
2. Ali, M., Kim, D. R., Yunus, M., & Emch, M. (2013). Time series analysis of cholera in Matlab, Bangladesh, during 1988-2001. *Journal of health, population, and nutrition*, *31*(1), 11.
3. Van den Bergh, F., Holloway, J. P., Pienaar, M., Koen, R., Elphinstone, C. D., & Woodborne, S. (2008). A comparison of various modelling approaches applied to Cholera case data. *ORiON*, *24*(1), 17-36.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 02, 2019

4. Rajendran, K., Sumi, A., Bhattachariya, M. K., Manna, B., Sur, D., Kobayashi, N., & Ramamurthy, T. (2011). Influence of relative humidity in Vibrio cholerae infection: a time series model. *The Indian journal of medical research*, *133*(2), 138.
5. Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental research*, *142*, 319-327.
6. Luque Fernández, M. Á., Bauernfeind, A., Jiménez, J. D., Gil, C. L., Omeiri, N. E., & Guibert, D. H. (2009). Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *103*(2), 137-143.
7. Shashvat, K., Basu, R., Bhondekar, P. A., & Kaur, A. (2019). An ensemble model for forecasting infectious diseases in India. *Trop Biomed*, *36*, 822-32.
8. Umar, N. S., Prasad, M. N. V., & Jena, B. N. (2010). Use of ARIMA models for forecasting needs in Emergency data based Syndromic Surveillance. *Transforming Emergency Management*, *17*.
9. Mirembe, B. B., Mazeri, S., Callaby, R., Nyakarahuka, L., Kankya, C., & Muwonge, A. (2019). Temporal, spatial and household dynamics of Typhoid fever in Kasese district, Uganda. *Plos one*, *14*(4), e0214650.
10. Rad, B. B., Shareef, A. A., Thiruchelvam, V., Afshar, A., & BAMIAH, M. (2018). A hybrid model for forecasting communicable diseases in Maldives. *Journal of Engineering Science and Technology*, *13*, 1-13.
11. Mbau, B. K. (2018, August). Forecasting the Amount of the Lung Diseases by the Method of ARIMA-ARCH. In *IOP Conference Series: Materials Science and Engineering* (Vol. 407, No. 1, p. 012155). IOP Publishing.
12. Ichiji, K., Sakai, M., Homma, N., Takai, Y., & Yoshizawa, M. (2010). A time variant seasonal ARIMA model for lung tumor motion prediction. In *Proc. of The 15th Int'l Symposium on Artificial Life and Robotics* (Vol. 2010, pp. 485-488).
13. Yeung, C., Ghazel, M., French, D., Japkowicz, N., Gottlieb, B., Maziak, D., ... & Gilbert, S. (2018). Forecasting pulmonary air leak duration following lung surgery using transpleural airflow data from a digital pleural drainage device. *Journal of Thoracic Disease*, *10*(Suppl 32), S3747.
14. Nayak, M. S. D. P., & Narayan, K. A. (2019). Forecasting dengue fever incidence using ARIMA analysis. *International Journal of Collaborative Research on Internal Medicine & Public Health*, *11*(3), 924-932.
15. Siregar, F. A., Makmur, T., & Saprin, S. (2018). Forecasting dengue hemorrhagic fever cases using ARIMA model: a case study in Asahan district. In *IOP Conference Series: Materials Science and Engineering* (Vol. 300, No. 1, p. 012032). IOP Publishing.
16. Choudhury, Z. M., Banu, S., & Islam, A. M. (2008). Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis.
17. Martinez, E. Z., Silva, E. A. S. D., & Fabbro, A. L. D. (2011). A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, *44*, 436-440.