

Forecasting Measles Immunization Coverage Using ARIMA Model

Rachel T. Alegado^{1,2}, Gilbert M. Tumibay³

¹Nueva Ecija University of Science and Technology, Cabanatuan, Philippines

²Graduate School, Angeles University Foundation, Angeles, Philippines

³Angeles University Foundation, Angeles, Philippines

Email: rachelalegado@neust.edu.ph, tumibay.gibo@auf.edu.ph

How to cite this paper: Alegado, R.T. and Tumibay, G.M. (2019) Forecasting Measles Immunization Coverage Using ARIMA Model. *Journal of Computer and Communications*, 7, 157-168.

<https://doi.org/10.4236/jcc.2019.710015>

Received: September 8, 2019

Accepted: October 22, 2019

Published: October 25, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study aimed to find a model to forecast monthly measles immunization coverage using Autoregressive Integrated Moving Average (ARIMA). The monthly registered data for measles immunization coverage from January 2014 to December 2018 were used for the development of the model. The best model with the smallest Normalized Bayesian Information Criterion (BIC) of 8.673 is ARIMA (0, 1, 0). ARIMA (0, 1, 0) was used to forecast the monthly measles immunization coverage for the next 36 months from January 2018 to December 2020. The results obtained prove that this model can be used for forecasting future immunization coverage and will help decision-makers to establish strategies, priorities, and proper use of immunization resources.

Keywords

Forecasting, Measles, Immunization Coverage, ARIMA, Modeling

1. Introduction

Measles is a highly contagious respiratory disease caused by a virus which belongs to the paramyxovirus family. It is an airborne disease which spreads quickly through the coughs and sneezes of infected people. It may also be spread through direct contact with the mouth or nasal secretions. Even though the safe and effective vaccine is available, it remains a notable cause of death among young children worldwide. In 2017, there were about 110,000 global deaths related to measles, and mostly in children under the age of five [1]. In the Philippines, the rate of measles vaccination has declined steadily from 80 percent in 2008 to below 70 percent in 2017. As a result, many children have become susceptible to measles infection. World Health Organization estimates that 2.6 mil-

lion Filipino children under the age of five years are vulnerable to measles [2].

In the Philippines, Expanded Program on Immunization (EPI) was established in 1976, one dose of measles vaccine given at nine months of age has been included in the program, then later on introduced nationwide in 1983 [3]. The measles vaccine is available free of cost in government health centers, with the first dose is given when a child is nine months old, and the second dose is given when a child is 12 months old [4].

This study aimed to propose a forecasting model for monthly measles infant immunization coverage. The forecasting result generated by the model can serve as the basis for preparing the accurate monthly volume of measles vaccine. By doing so, this will help to achieve a high level of coverage and improve measles immunization planning and program.

A time-series approach with autoregressive integrated moving average (ARIMA) modeling is used to provide a monthly forecast of immunization coverage with precision. It is a combination of three statistical models. It uses the Autoregressive, Integrated, and Moving Average (ARIMA) model for statistical information. ARIMA model is commonly used in analysis and forecasting. It is the most efficient forecasting technique in social science and used extensively for time series [5]. In medical applications, time series forecasting models have been successfully applied to predict the progress of the disease, estimate the mortality rate, and assess the time-dependent risk [6].

Several studies were conducted to forecast immunization coverage. Authors in [7] utilized ARIMA and Back-Propagation Neural Networks (BPNN) to forecast the annual vaccine demand of a specific vaccine. Similarly, researcher in [8] use ARIMA analysis in predicting vaccine demand in advance through analysis of progress of birth of newborn babies in Korea, while [9] use Artificial Neural Network (ANN) to expose the trend and proposing the forecasting model for monthly pentavalent infant immunization coverage and [10] use ARIMA and Neural Network to propose computer-based forecast model to build a decision support system aiming to forecast the annual vaccine demand for specific vaccines of Taiwan. However, none of them are focused on forecasting measles immunization coverage and provided three-year forecasts. Moreover, this study will contribute to existing knowledge of applying the ARIMA model in forecasting immunization coverage.

2. Methodology

The research methodology used in this study is summarized below. The study used monthly registered data for measles immunization coverage from January 2014 to December 2018 total time series 60 was taken from the City Health Office of Cabanatuan. ARIMA is used to uncover the measles immunization coverage trend and the development of the forecasting model using IBM SPSS.

2.1. Data Source

The dataset for this study was obtained from Cabanatuan City Health Office,

Immunization Department. The dataset consists of the number of infant per month receiving the measles immunization for 60 month period from January 2014 to December 2018 in Cabanatuan City, Nueva Ecija, Philippines.

2.2. ARIMA Model

Autoregressive Integrated Moving Average model is popular and widely used stochastic time series model. It is a combination of three statistical models. An ARIMA model is denoted as an ARIMA model (p, d, q) , where p is the number of autoregressive terms, d is the degree of differencing involve, and q is the number of moving-average terms [11] [12] [13]. This study follows the Box and Jenkins methodology [14], which composed of four main steps as shown in **Figure 1**.

1) Model Identification. In this step, the order of respective time series model is determined, the order of differencing “ d ”, the order of AR process which is “ p ” and order of MA process is “ q ”. This mainly leads to the use of graphical procedures such as plotting the series, the ACF and PACF.

2) Estimation of Model Parameters. Involves the estimation of the parameters of the different models using previously available data and proceeds to the first selection of models using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.

3) Diagnostic checks (model checking). This step determines whether the model(s) specified and estimated is valid. Thus, the residual values are normally distributed, no significant periodicity is present in the residual, and the residual series is uncorrelated. The process will start again if this step is not fulfilled.

4) Forecasting. In this step, the model is now ready to be tested on real data to forecast future values of immunization coverage.

3. Results and Discussions

This study applies Box-Jenkins methodology for the identification, estimation, diagnostic checking, and forecasting of the univariate time-series data on monthly measles immunization coverage. The graph of the raw data is shown in **Figure 2**.

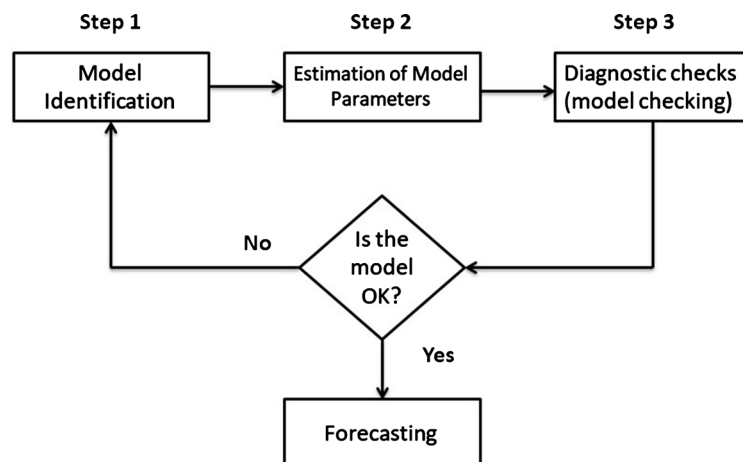


Figure 1. The box and Jenkins methodology.

3.1. Model Identification

The first stage in developing an ARIMA model is to determine if the series is stationary. The graph below (**Figure 2**) depicts that the series has a large variance. To reduce the variance, the “raw data” was transformed into its natural logarithmic equivalent. The graph of the log (immunization) is shown below.

Figure 2 and **Figure 3** show that the mean of the series is a function of time

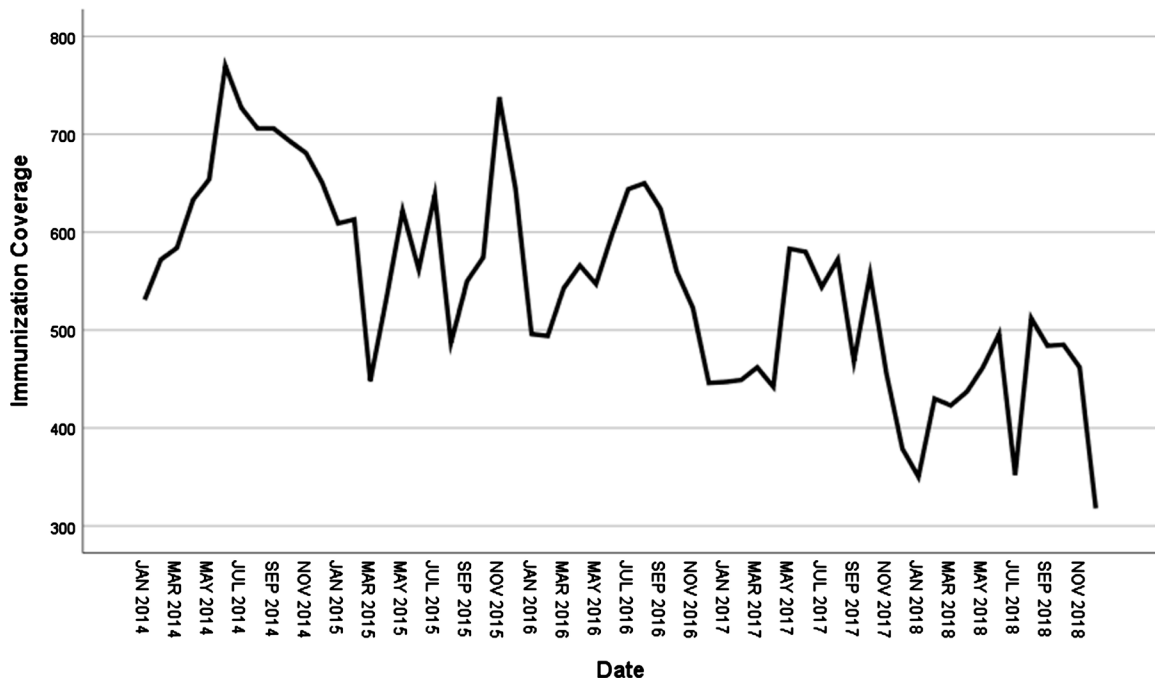


Figure 2. Monthly measles immunization coverage from January 2014 to December 2018.

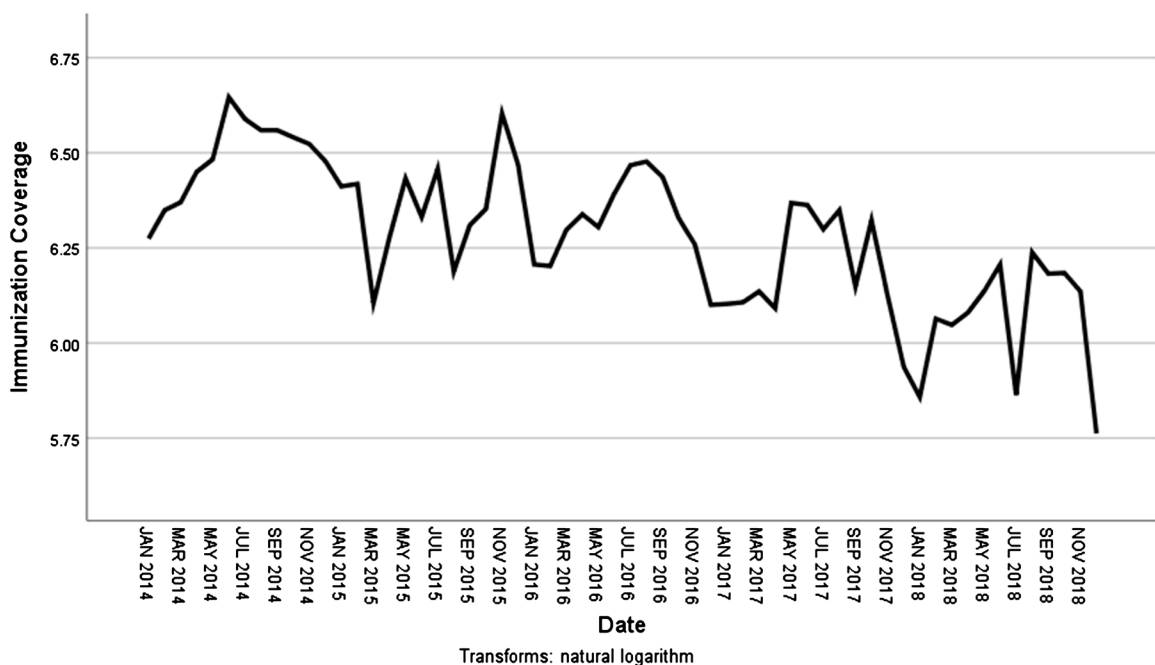


Figure 3. Log (Immunization).

(*i.e.*, the mean of the series decreases as time increases). Also, there are some points in time where the variances are smaller and larger (*i.e.*, the variances are not constant over time). Visually, both sequence charts show that the monthly measles immunization data is non-stationary.

Augmented Dickey-Fuller (ADF) test was conducted to test the hypothesis statistically that the natural log-transformed series has a unit root or not. The series is non-stationary if there is a unit root and the series is stationary if there is no unit root. Results of the ADF test is summarized in **Table 1(a)** below. The results of the three test equations are inconsistent, an indication that the series is non-stationary. The finding using the ADF test is consistent of the sequence charts above.

Differencing the natural log-transformed series can transform the series from non-stationary series to stationary. The graph below shows that the “first difference of the log-transformed series” is stationary (**Figure 4**).

The ADF test for the first difference of the natural log-transformed data confirms that a stationary series can be achieved after the first differencing.

The ACF and PACF of the first difference natural log-transformed series are shown below. The height of bars in the ACF represents the autocorrelation coefficients. The two parallel lines represent the confidence interval. All bars fall inside the confidence interval, indicating that the autocorrelations are statistically non-significant at 5% level. This means that an ARIMA with Moving Average component of order 0 can be tentatively considered, that is, MA (0) (**Figure 5**).

The height of bars in the PACF represents the partial autocorrelation coefficients. Though the partial autocorrelation coefficients fall within the confidence level, that is, the partial autocorrelation coefficients are statistically nonsignificant at 5%, some are larger than the others. The PACF suggests an ARIMA model with AR of order 0 can be considered. AR of orders 1, 4, 8, and 9 can also

Table 1. (a) Unit root test of the log-transformed data using the ADF test; (b) Unit root test of the first difference of the log transformed data using the ADF test.

(a)			
Test equation	T-value	p-value	Remarks
Intercept Only	-2.4495	0.1330	There is a unit root.
Trend and Intercept	-4.7055	0.0018	There is no unit root
No Constant, No Intercept	-0.5265	0.4845	There is a unit root.
(b)			
Test equation	T-value	p-value	Remarks
Intercept Only	-9.2437	0.0000	There is no unit root.
Trend and Intercept	-9.2414	0.0000	There is no unit root
No Constant, No Intercept	-9.2900	0.0000	There is no unit root.

Note: No unit root indicates a stationary series. To be stationary, the three test equations should show no unit root.

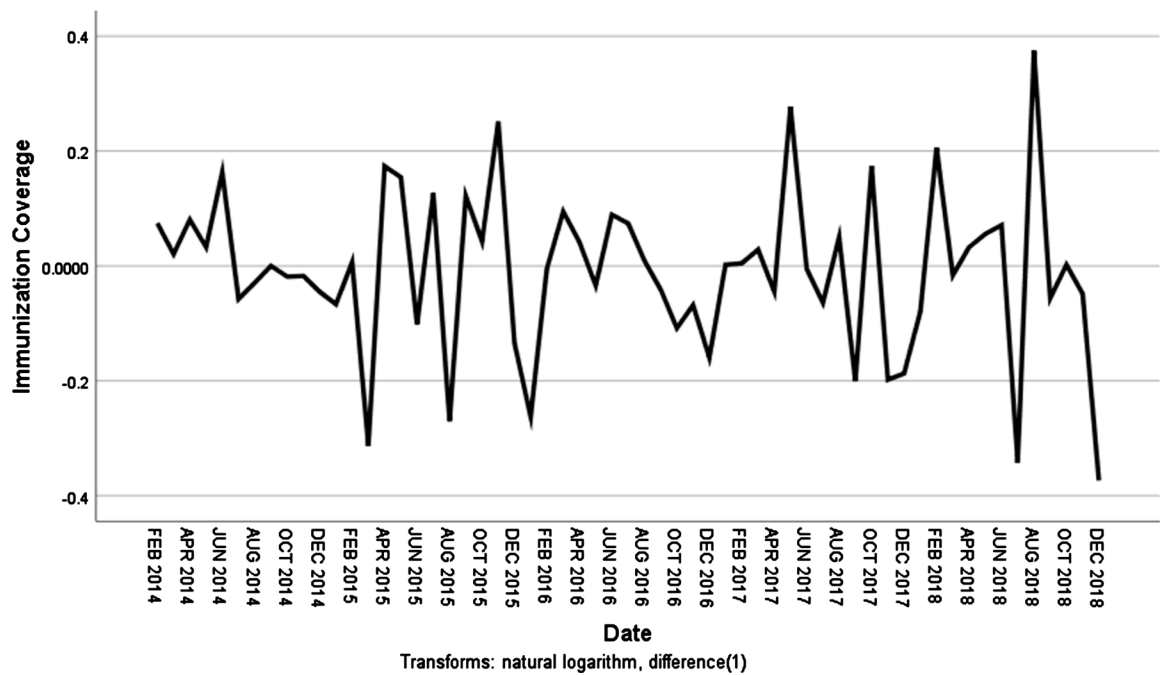


Figure 4. First difference of the log transformed series.

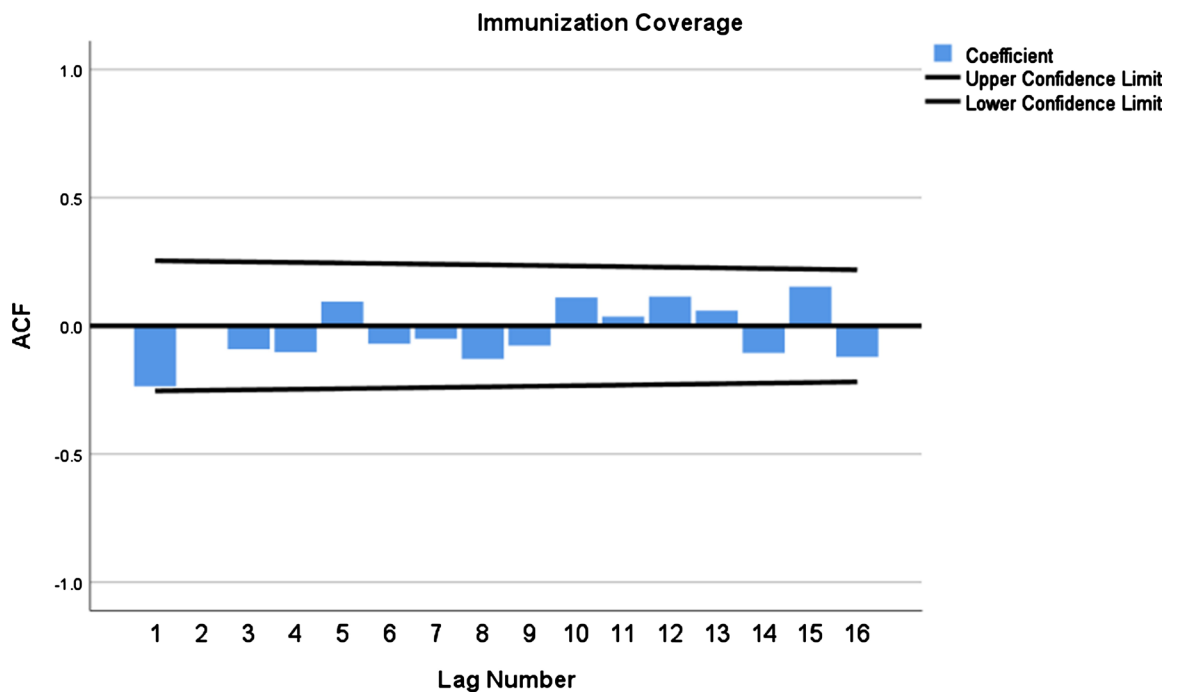


Figure 5. Plot of autocorrelation (ACF) of the first difference natural log-transformed series.

be considered. In short, ARIMA with AR (0), AR (1), AR (4), AR (8), and AR (9) can be considered.

Based on the results of the ACF and PACF, the following ARIMA Models can be considered:

ARIMA Model 1: ARIMA (0, 1, 0)

ARIMA Model 2: ARIMA (1, 1, 0)

ARIMA Model 3: ARIMA (4, 1, 0)

ARIMA Model 4: ARIMA (8, 1, 0)

ARIMA Model 5: ARIMA (9, 1, 0)

3.2. Estimation of Model Parameters

After the estimation of the parameters, the Root Mean Square Errors (RMSE) and Normalized BIC are summarized as shown in **Table 2**. The smaller the RMSE and Normalized BIC, the better the model. In this study, the normalized BIC is used because its value is not a function of the number of parameters.

Using the Normalized BIC as the basis for the selection of the best ARIMA Model, model #1, the ARIMA (0, 1, 0), emerged as the best model with the smallest Normalized BIC of 8.673. The estimates of the model parameters generated by the IBM SPSS Forecasting module are presented below. In equation form, the ARIMA (0, 1, 0) can be written as: $D\log(\text{Immunization}) = -0.009$ (**Figure 6**).

3.3. Diagnostic Checks

3.3.1. Normality Distributed Residuals

Shapiro-Wilk test was used to test if the “residuals” of the ARIMA (0, 1, 0) is normally distributed. Shapiro-Wilk statistic is statistically nonsignificant ($SW = 0.973$, $df = 59$, $p = 0.219$), indicating that the residuals have no departure from normality. **Figure 7** Histogram and **Figure 8** Normal Q-Q plot support the normality.

3.3.2. No Autocorrelations in the Residuals

The Ljung-Box Q was used to test the absence of autocorrelations as a whole. The Ljung-Box Q is statistically nonsignificant (Ljung-Box $Q = 13.778$, $p = 0.743$), indicating the absence of autocorrelations. The bars of the PACF and ACF plots are within the 95% confidence limits, indicating the absence of autocorrelations (**Figure 9**).

Table 2. RMSE and normalized BIC of the alternative ARIMA models.

Models		RMSE	Normalized BIC
1. ARIMA (0, 1, 0)		73.857	8.673
2. ARIMA (1, 1, 0)		72.569	8.707
3. ARIMA (4, 1, 0)		72.240	8.909
4. ARIMA (8, 1, 0)		71.272	8.938
5. ARIMA (9, 1, 0)		68.517	9.145

ARIMA Model Parameters				Estimate	SE	t	Sig.
Immunization	Immunization	Natural	Constant	-0.009	-0.019	-0.459	-0.648
Coverage-Model_1	Coverage	Logarithm	Difference	1			

Figure 6. ARIMA model parameters.

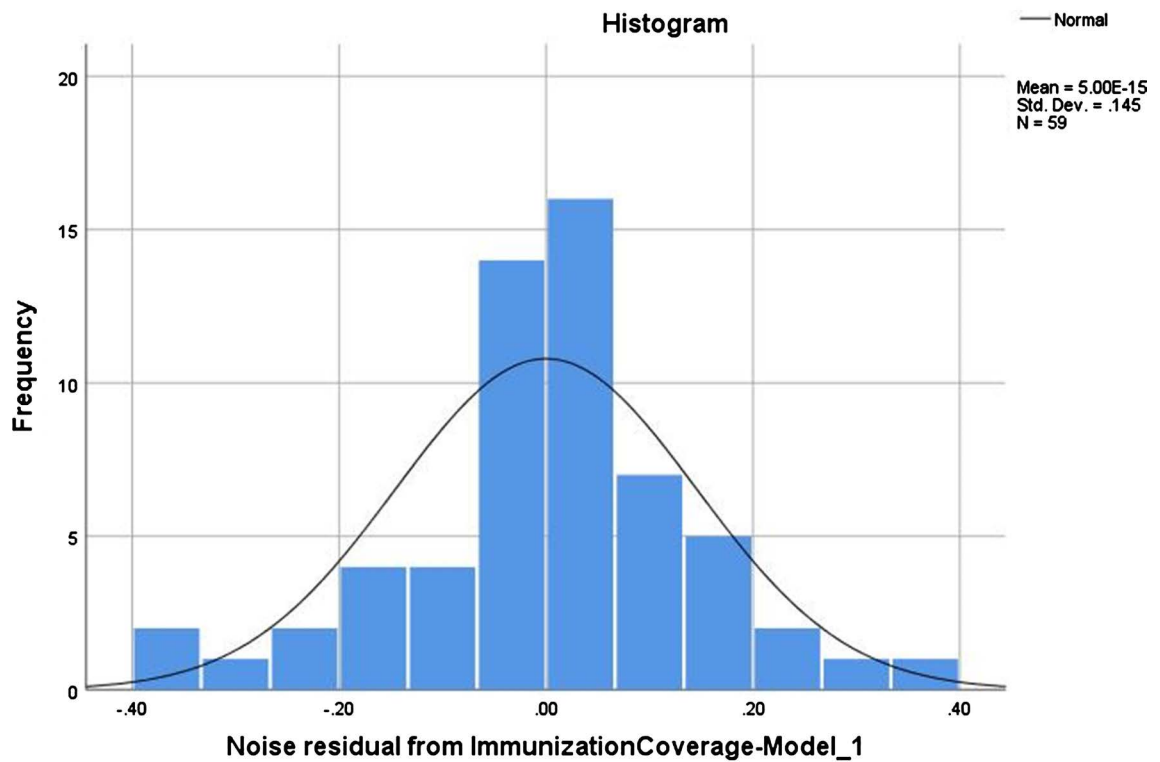


Figure 7. Histogram.

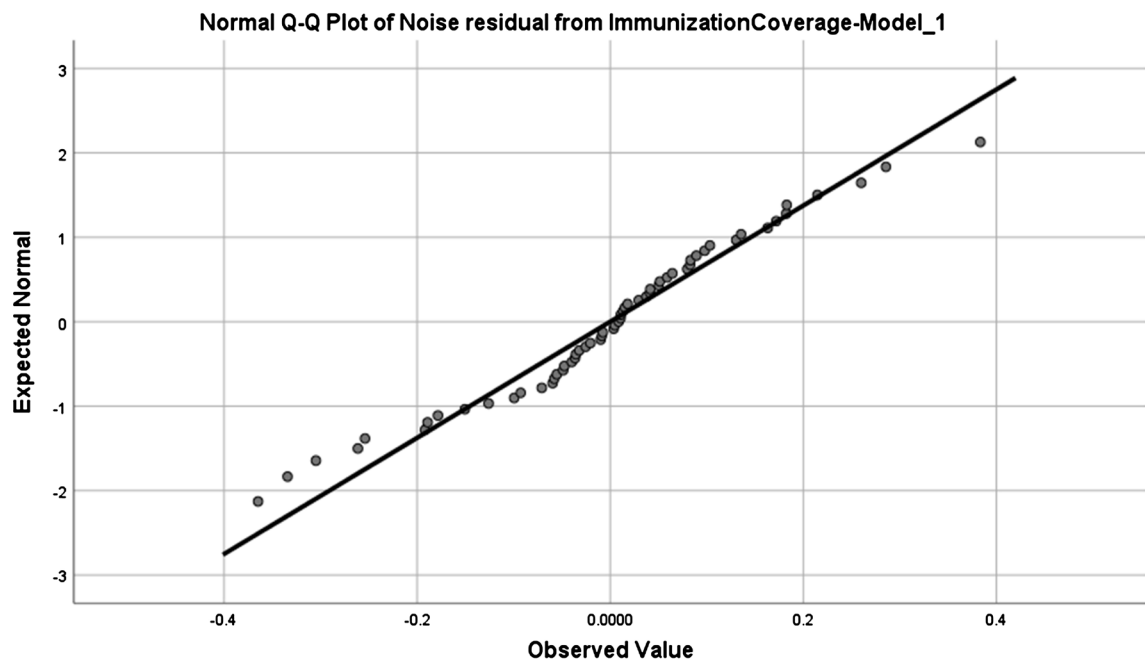


Figure 8. Normal Q-Q plot of noise residual.

3.4. Forecast

ARIMA (0, 1, 0) was used to forecast the monthly measles immunization coverage for the next 36 months from January 2018 to December 2020. **Figure 10** shows the actual data (blue line) and fitted/forecasted values (red line) using the

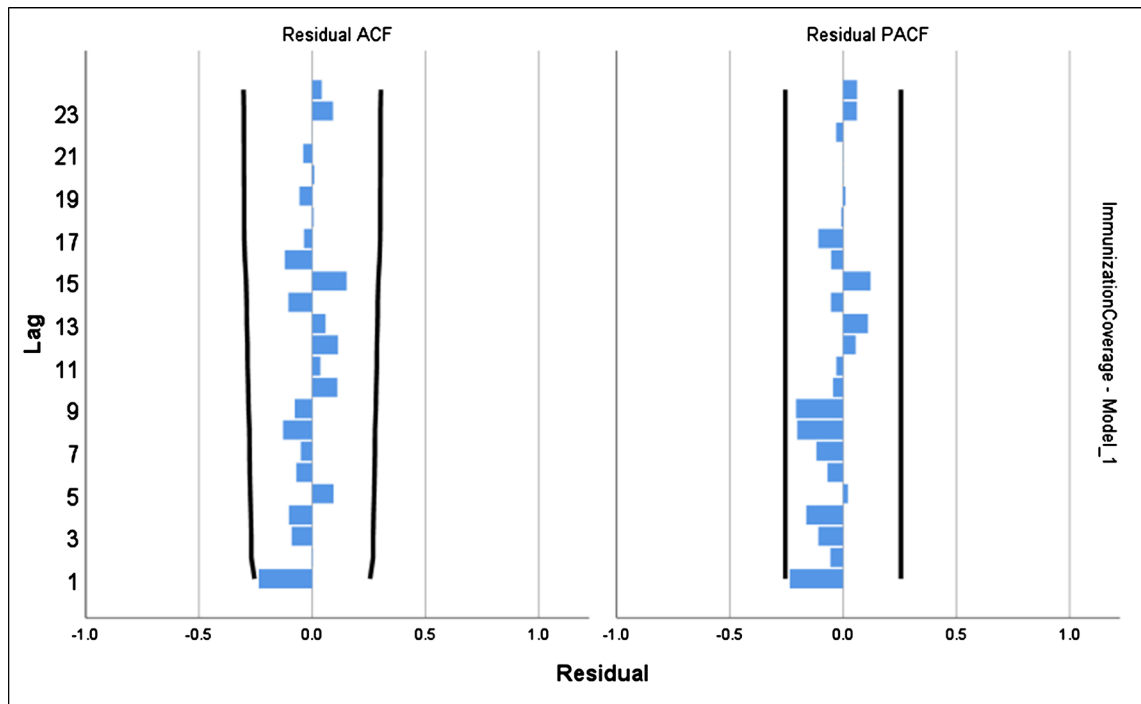


Figure 9. ACP and PACF of residuals.

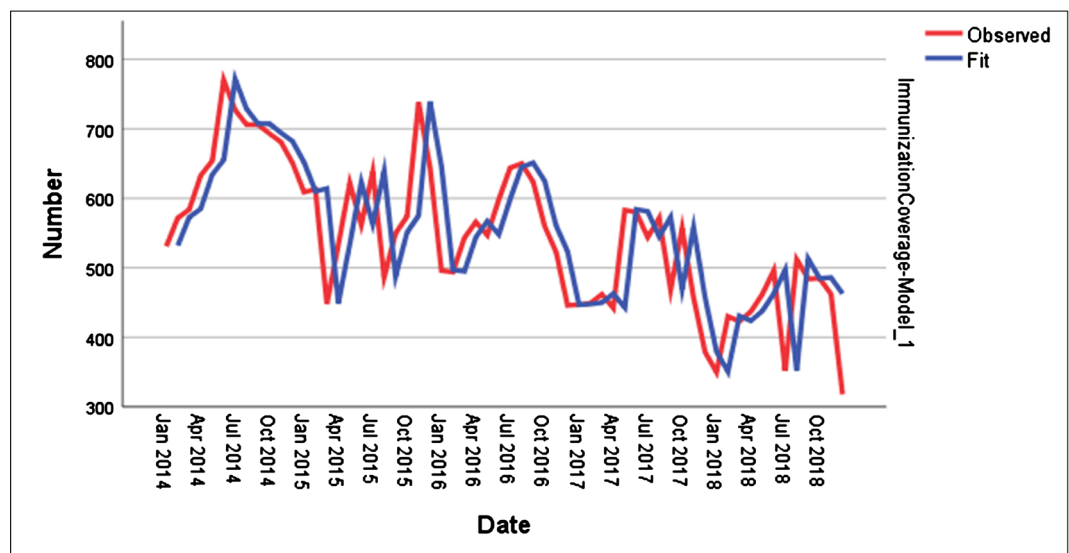


Figure 10. The observed and fit values for measles immunization coverage.

equation. Table 3 shows forecast values of immunization coverage from January 2018 to December 2021 using ARIMA (0, 1, 0).

4. Conclusion

In this study, the ARIMA model was developed to forecast monthly measles immunization for the next 36 months from January 2019 to December 2021. The monthly registered data for measles immunization coverage from January 2014 to December 2018 were used for the development of the model. The best model

Table 3. Forecasted measles immunization coverage for the next three years (January 2019 to December 2021).

Year	Lo 95	Hi 95
Jan 2019	174.02	461.98
Feb 2019	114.38	521.62
Mar 2019	68.61	567.39
Apr 2019	30.03	605.97
May 2019	-3.96	639.96
Jun 2019	-34.69	670.69
Jul 2019	-62.94	698.94
Aug 2019	-89.25	725.25
Sep 2019	-113.95	749.95
Oct 2019	-137.32	773.32
Nov 2019	-159.54	795.54
Dec 2019	-180.77	816.77
Jan 2020	-201.14	837.14
Feb 2020	-220.74	856.74
Mar 2020	-239.64	875.64
Apr 2020	-257.93	893.93
May 2020	-275.66	911.66
Jun 2020	-292.87	928.87
Jul 2020	-309.61	945.61
Aug 2020	-325.91	961.91
Sep 2020	-341.81	977.81
Oct 2020	-357.34	993.34
Nov 2020	-372.52	1008.52
Dec 2020	-387.37	1023.37
Jan 2021	-401.92	1037.92
Feb 2021	-416.17	1052.17
Mar 2021	-430.16	1066.16
Apr 2021	-443.89	1079.89
May 2021	-457.37	1093.37
Jun 2021	-470.63	1106.63
Jul 2021	-483.66	1119.66
Aug 2021	-496.49	1132.49
Sep 2021	-509.12	1145.12
Oct 2021	-521.56	1157.56
Nov 2021	-533.82	1169.82
Dec 2021	-545.90	1181.90

with the smallest Normalized BIC of 8.673 is ARIMA (0, 1, 0). The results obtained prove that this model can be used for forecasting future immunization coverage and will help decision-makers to establish strategies, priorities, and proper use of immunization resources in Cabanatuan City. As future work, the researcher will develop other models by using the neural network approach to compare it with the ARIMA model.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] WHO (2019) Measles, Fact Sheet No. 90. <https://www.who.int/news-room/fact-sheets/detail/measles>
- [2] WHO (2019) News, Feature Stories No. 5. <https://www.who.int/philippines/news/feature-stories/detail/questions-and-answers-on-the-measles-outbreak-in-the-philippines>
- [3] Ylade, M.C. (2018) Epidemiology of Measles in the Philippines. *Acta Medica Philippina*, **52**, 380. <https://www.actamedicaphilippina.org/article/5144>
- [4] DOH Philippines (2019) Expanded Program on Immunization. <https://www.doh.gov.ph/expanded-program-on-immunization>
- [5] Adebiyi, A.A., Adewumi, A.O. and Ayo, C.K. (2014) Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, **2014**, Article ID: 614342. <https://doi.org/10.1155/2014/614342>
- [6] Bui, C., Pham, N., Vo, A., Tran, A., Nguyen, A. and Le, T. (2017) Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases. In: *International Conference on the Development of Biomedical Engineering in Vietnam*, Springer, Singapore, 809-818. https://doi.org/10.1007/978-981-10-4361-1_138
- [7] Chiu, R., Chang, C. and Chang, Y. (2008) A Forecasting Model for Deciding Annual Vaccine Demand. *Fourth International Conference on Natural Computation*, Jinan, 107-111. <https://doi.org/10.1109/ICNC.2008.862>
- [8] Kim, K.-W., Li, G.Z., Park, S.-T. and Ko, M.-H. (2016) A Study on Birth Prediction and BCG Vaccine Demand Prediction Using ARIMA Analysis. *Indian Journal of Science and Technology*, **9**, 1-7. <https://doi.org/10.17485/ijst/2016/v9i24/96151>
- [9] Imran, M., Abdul Nasir, J. and Chohan, A. (2016) Modeling towards Pentavalent Vaccine Coverage in Pakistan. *Journal of University Medical & Dental College*, **7**, 41-46.
- [10] Chiu, R., Chan, C.-L. and Feng, C. (2007) A Forecast System for Making Decision of Vaccine Demand. 2008 *Fourth International Conference on Natural Computation*, Jinan, 18-20 October 2008, 23-27.
- [11] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H. and Lachhab, A. (2018) Forecasting of Demand Using ARIMA Model. *International Journal of Engineering Business Management*, **10**. <https://doi.org/10.1177/1847979018808673>
<https://journals.sagepub.com/doi/full/10.1177/1847979018808673>
- [12] Mahsin, M.D. (2011) Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series Analysis. *Journal of Mathematical Modelling and Application*, **1**, 67-73.

- [13] Chase, C.W. (2013) Demand-Driven Forecasting: A Structured Approach to Forecasting. John Wiley & Sons, Hoboken. <https://doi.org/10.1002/9781118691861>
- [14] Ariyo, A.A., Adewumi, A.O. and Ayo, C.K. (2014) Stock Price Prediction Using the ARIMA Model. *UKSim-AMSS 16th International Conference on Computer Modeling and Simulation*, Cambridge, 26-28 March 2014, 106-112. <https://doi.org/10.1109/UKSim.2014.67>