ANALYSIS AND MODELING OF PREVALENCE OF MEASLES IN THE ASHANTI REGION OF GHANA

Kwame Asare Gyasi-Agyei 1, William Obeng Denteh 2 *, Amoakoh Gyasi-Agyei 3

1 Department of Applied Mathematics, Koforidua Polytechnic, Koforidua, Ghana. gaakus@yahoo.com. Mobile: 0203086648 / 0244186326.

2 Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. obengdentew@yahoo.com. Mobile: 0233737737

3 Faculty of Engineering, Science & Computing, Pentecost University College, Accra, Ghana info@gictrac-centre.org. Mobile: 0203612828

ABSTRACT

In this paper, autoregressive integrated moving average (ARIMA) model is used to predict the prevalence and incidence of measles in the Ashanti Region of Ghana. The Mean Absolute Error (MAE) and the Mean Squared Error (MSE) are used to compare the in-sample forecasting performance of four selected competing models. The working data from the Ashanti Health Services spans from 2001 to 2011. It is evident from the analysis that measles data in the Ashanti Region of Ghana could best be modeled with ARIMA (2, 1, 1) and that measles prevalence in the Ashanti Region is expected to increase if no preventative measures are taken. The forecasting accuracy using MAE for ARIMA (2, 1, 1) is calculated as 28.1141 and the forecasting accuracy using MSE for ARIMA (2, 1, 1) is calculated as 2947.15.

Keywords: Measles, prevalence, models, forecasting, autoregressive, virus, vaccine, symptoms

1.0 INTRODUCTION

Measles is one of the most contagious but vaccine-preventable diseases which is caused by the measles virus. It is one of the most readily communicable diseases and probably the best known, and most deadly of all childhood rash/fever illnesses. It is a childhood disease that rarely occurs in adults (Wikipedia, 2008).

Measles is an illness that is spread through the coughs and sneezes of infected people. One can catch measles by breathing in these droplets or, if the droplets have settled on a surface, by touching the surface and then placing the hands near the nose or mouth. The measles virus can survive on surfaces for a few hours. In fact, the virus is one of the most contagious viruses known to man. As a result, it can spread rapidly in a susceptible population. Infected people carry the virus in their respiratory tract before they get sick, so they can spread the disease without being aware of it (Nettleman, 2008).

When a person becomes infected with the measles virus, it begins to multiply within the cells that line the lungs and the back of the throat. The virus can also spread to the lymph glands, bone marrow, liver, eyes, thymus, tonsils, spleen, skin and brain. The symptoms typically appear ten to fourteen days after a person is infected with the measles virus. The period between measles transmission and the start of symptoms is called the incubation period. During this period, the virus is multiplying. It includes fever, sore throat, cough, sore eyes, red watery eyes, vomiting, runny nose, loss of appetite and fatigue. Koplik’s spot is the characteristic symptom of measles. The majority of people infected with the virus recover, but measles complications can be dangerous (Schoenstadt, 2006).

Complications of measles includes ear infections, diarrhea, pneumonia, seizures and encephalitis (inflammation of the brain) – this is rare, but can cause permanent brain damage or death. Up to 30 percent of people with measles will develop complications – usually children under five and adults over the age of 20. Measles during pregnancy increases the risk of miscarriage, premature labour and low birth-weight babies (WHO, 2011).
The skin rash appears within three to five days of the onset of symptoms. The rash often begins on the face and spreads downward all over the body. A very high fever may develop with the rash. The rash starts to disappear after a few days, and the fever resolves. It is better to leave the rashes alone as scratching leaves the patient in worse condition (Nettleman, 2008).

Unvaccinated young children are at highest risk of contracting measles and its complications, including death. Any non-immune person (who has not been vaccinated or previously recovered from the disease) can become infected. The risk factors include lack of immunization with the measles vaccine, travel to, or residence in, a country where measles is still prevalent, vitamin A deficiency (WHO, 2011).

There is currently no proven treatment that can kill the virus that causes this illness. The best way to prevent it is to get the measles vaccine. Remember never to give aspirin to children or teenagers because it may cause a disease known as Reye syndrome (Nettleman, 2008).

The major problem of measles is that Measles weakens the immune system and opens the door to secondary health problems, such as pneumonia, blindness, diarrhea, encephalitis etc. When one person has measles, 90 percent of the people they come into close contact with will become infected, if they are not already immune to it (Schoenstadt, 2006).

A total of 770,909 children, aged under-five years in the Ashanti Region, were immunized against measles in November, last year representing 95.5 per cent coverage. Dr Joseph Oduro, the Deputy Regional Director of Health Service, said the figure fell short of the target of 807,622 set. He said the children were also given Vitamin “A” supplement during the three-day exercise. Dr Oduro told the Ghana News Agency (GNA) in Kumasi that since year 2003, no child had died from measles in the region. This should be heart-warming news as it shows that Ghana is on course to achieving the goal of reducing infant mortality. He praised the various stakeholders who had over the years worked tirelessly to ensure consolidation of the gains made. The Deputy Regional Director spoke of the need for more public education and social mobilization to promote child health and development.

Source: GNA

1.1 Objectives of the paper are:
1. to observe the pattern of measles infections in the Ashanti Region of Ghana from January 2001 to November 2011.
2. to model the prevalence of measles in the Ashanti Region using Box-Jenkins ARIMA process.
3. to evaluate the impact of AIC on in-sample forecasting performance of the selected ARIMA models.

1.2 Significance of the paper
i. This paper provides a method for assessing the prevalence of measles in the Ashanti Region of Ghana and its effects in the near future.

ii. This paper contributes to the research information on measles in the country, so that it can help in further work in the area of research to investigate the effects of the disease on Ghanaians.

iii. The paper attempts to present both application and theory at a level accessible to a wide variety of students and researchers.

From the revealed analysis no research work has been extended to model cases of prevalence of measles in the Ashanti Region of Ghana. This paper therefore seeks to study the prevalence of measles cases in the Ashanti Region of Ghana using time series analysis.

This paper is organized in four sections. Section 1 reviews the background of the subject matter of the paper. Section 2 consists of the discussion of the methods used for the paper. Section 3 deals with data
Allen et al. (1993) stated that an epidemic of rubella occurred on the campus of Texas Tech University in January, February and March of 1989. A vaccination programme was initiated as soon as the epidemic was confirmed. Extensive case histories of all confirmed cases were collected by the Lubbock City Health Department and given an exhaustive statistical analysis by a group from the Department of Mathematics at Texas Tech University. The data and statistical analysis were used to formulate stochastic and deterministic models of the measles epidemic based on the standard SEIR model. The analysis and the simulations indicated that in order to prevent measles outbreak on a university campus a high rate of immunity above 98 per cent might be required.

A mathematical model of the dynamics of measles in New Zealand was developed in 1996. The model successfully predicted an epidemic in 1997 and was instrumental in the decision to carry out an intensive MMR (measles-mumps rubella) immunization campaign in that year. While the epidemic began some months earlier than anticipated, it was rapidly brought under control, and its impact on the population was much reduced. In order to prevent the occurrence of further epidemics in New Zealand, an extended version of the model had since been developed and applied to the critical question of the optimal timing of MMR immunization (Roberts and Tobias, 2000).

Stamp et al. (1990) presented a mathematical model for the simulation of a localized measles epidemic. Their work was presented along with a computer simulation based on this model. The simulation results were compared with the results of a measles outbreak which occurred at Texas Tech University. The effectiveness of the vaccination programme undertaken during the Texas Tech epidemic and the effect of altering the level of herd immunity were also considered.

A simple stochastic mathematical model was developed and investigated for the dynamics of measles epidemic by Kassem and Ndam (2010). Their model, which was a multi-dimensional diffusion process, included susceptible individuals, latent (exposed), infected and removed individuals. Stochastic effects were assumed to arise in the process of infection of susceptible individuals. Using the best currently available parameter values, the intrinsic variability in response to a given initial infection was examined by solving the stochastic system numerically. The results of the simulation seemed to agree with the historical pattern of measles in Nigeria.

Cliff and Haggett (1993) reviewed the application of statistical models to the outbreaks of measles epidemic. They looked first at its epidemiological characteristics and assessed the extent to which those either aid or hinder modeling. They then turned to the models that had been developed to simulate geographical spread. A distinction was drawn between process-based and time series models. They provided applications from work, by using Icelandic data. Finally they considered the forecasting potential of the models described.

Souza (1982) proposed a new approach to forecasting based on the Bayesian principles of information theory and called the Poisson - gamma single - state model. In his paper, a two-state version of the Poisson - gamma model was formulated by considering the uncertainty not only in the parameters but also in the model itself. That model was particularly useful for modeling epidemic data such as measles by considering two different situations of the generating process at each time point.

According to Commey and Richardson (1984), admissions of children with measles constituted 8.8% of all admissions to the paediatric medical service of the Korle Bu Teaching Hospital, Accra, Ghana, over the ten-year-period 1973-1982. Measles remains endemic in urban Accra as in the towns of other developing nations. The peak of admissions occurred in the age range seven to 12 months. Complications were frequent, with a high mortality (16.86%). Bronchopneumonia, the commonest complication (63.9%) was also the commonest cause of death in 51.5% of cases. Comparative national case-mortality rates were, however, surprisingly low and should be accepted with caution. There is an urgent need for intensification of immunization efforts through amalgamation of the preventive and curative services of Ghana, especially
for children attending health centres for medical care. Vaccination should be administered before the peak age of admission, preferably at six months of age, with a second dose administered as soon after one year of age as possible.

2.0 MATERIALS AND METHODS

The data used for the modeling and analysis was obtained from the Ministry of Health in the Ashanti Region of Ghana. It was a secondary data. The data on monthly bases consist of the measles cases from various hospitals in the Ashanti Region for the period of January 2001 to December 2011.

Ashanti region is used as our case study area since the region is a cosmopolitan region in Ghana. All sorts of people are living in the region. The region is the third largest of 10 administrative regions in Ghana, occupying a total land surface of 24,389 square kilometers or 10.2 per cent of the total land area of Ghana. In terms of population, however, it is the most populated region with a population of 3,612,950 in 2000, accounting for 19.1 per cent of Ghana’s total population; however, its density (148.1 per square km) is lower than those of the Greater Accra (895.5/km$^2$) and Central (162.2/km$^2$) Region.

The region is centrally located in the middle belt of Ghana. It lies between longitudes 0.15W and 2.25W, and latitudes 5.50N and 7.46N. The region shares boundaries with four of the ten political regions, Brong-Ahafo Region in the north, Eastern region in the east, Central region in the south and Western region in the South west. The region is divided into 27 districts, each headed by a district chief executive.

Majority of the region’s population are Ghanaians by birth (87.3%) with about five per cent naturalized Ghanaians. A smaller proportion (5.8%) of the population originate from outside Ghana, made up of 3.7 per cent mainly from the five English-speaking countries of ECOWAS and 2.1 per cent from other African countries. The non-African population living in the region is 1.8 per cent of the total population. Akans are the predominant ethnic group in the region, representing 77.9% of Ghanaians by birth. A high proportion (78.9%) of the Akan population is Asante. The non-Akan population in the region comprises the Mole-Dagbon (9.0%), the Ewe (3.2%), the Grusi (2.4%), the Mande-Busanga (1.8%) and the Ga-Dangme (1.4%). The other smaller ethnic groups form about 1.3 per cent of the population of the region.

The model used to analyze the collected data is the Auto-Regressive Integrated Moving Average (ARIMA) which was developed by Box and Jenkins in 1970, simply because the ARIMA model fits the collected data very well as compared to the other models. The R software package has been used in addition to manual calculations to model the given data. The Internet and Kwame Nkrumah University of Science and Technology school library are among the resources exploited.

The basis of the Box-Jenkins approach to modeling time series is summarized below and consists of three phases: identification, estimation and testing, and application (Spyros et al., 1998).

2.1 Identification Stage
Transform data to stabilize variance.

2.1.1 Difference data to obtain stationary series:
For non-seasonal data, first differencing is usually sufficient to attain stationarity. The first-order difference is denoted by equation (1).

\[
\nabla X_t = X_t - X_{t-1}
\]

2.1.2 Model selection:
Examine data, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify potential models. The autocorrelation function (ACF), $\rho_{t,s}$, is given by the formula in equation (2)
\[
\rho_{t,s} = \text{Corr}(X_t, X_s) = \frac{\text{Cov}(X_t, X_s)}{\sqrt{\text{Var}(X_t)\text{Var}(X_s)}} \quad \forall \ t, s \in \{0, \pm 1, \pm 2, \pm 3, \ldots \}
\]  

(2)

where \(\text{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] = E(X_t, X_s) - \mu_t\mu_s\)

The autocorrelation coefficient estimated from sample observations at lag \(k\) is given by equation (3).

\[
\gamma_k = \frac{\sum_{t=k+1}^{n}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{(Spyros, 1998)}
\]  

(3)

2.2 Testing Stage

2.2.1 Augmented Dickey-Fuller (ADF) Test

The hypotheses \(H_0: \ X_t\ \text{is non–Stationary and} \ H_1: \ X_t\ \text{is Stationary}\) can be tested in the regression equation (4).

\[
\Delta X_t = \beta_0 + \alpha + \beta_t X_{t-1} + \sum_{i=1}^{n} \gamma_i \Delta X_{t-i} + \varepsilon_t
\]  

(4)

Accept \(H_0\) if \(P-value > 0.05\), else accept \(H_1\).

2.2.2 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

An alternative approach to the ADF test is the KPSS test. A hypotheses of \(H_0: \ X_t\ \text{is level or trend stationary}\) is tested against \(H_1: \ X_t\ \text{is non- stationary}\) in the regression equation in (5).

\[
X_t = \alpha_t + \beta_t + \mu_t, \text{ where a random walk, } \alpha_t = \alpha_{t-1} + \varepsilon_t \text{ is allowed.}
\]  

(5)

Accept \(H_0\) if \(P-value > 0.05\), else accept \(H_1\).

2.3 Select best model using suitable criterion:

2.3.1 The Akaike’s Information Criteria (AIC)

The AIC is equal to twice the number of parameters in the model minus twice the logarithm of the likelihood function. Mathematically, AIC is calculated by equation (6).

\[
\text{AIC}(p, q) = 2k - 2\log(\text{Maximum Likelihood})
\]  

(6)

where \(k = p + q + 1\) if the model contains an intercept or constant term and \(k = p + q\) otherwise.

Given two or more competing models, the one with the smaller AIC value will be deemed more appropriate (Spyros et al., 1998).

2.3.2 The Schwartz’s Bayesian Information Criteria (BIC)

Like the AIC, the BIC is an order selection criterion for ARIMA models. It is defined mathematically by equation (7).

\[
\text{BIC}(p, q) = k \times \log(n) - 2 \times \log(\text{Max. Likelihood})
\]  

(7)
2.4 Diagnostics

2.4.1 Testing the Model for Adequacy (Portmanteau Test)
After identifying an appropriate model for a time series data, it is very important to check that the model is adequate. Ljung and Box (1978) provided a modified portmanteau test statistic for checking the randomness of the error terms. Their statistic is given by equation (8),

\[ Q^* = n(n+2) \sum_{k=1}^{h} \left( \frac{r_k^2}{n-k} \right) \]

which is approximately distributed as a \( \chi^2 \) with \( h - p - q \) degrees of freedom, where \( n \) is the length of the time series, \( h \) is the first \( h \) autocorrelations being checked, \( p \) is the order of the AR process and \( q \) is the order of the MA process and \( r \) is the estimated autocorrelation coefficient of the \( k^{th} \) residual term. If the calculated value of \( Q^* \) is greater than \( \chi^2 \) for \( h - p - q \) degrees of freedom, then the model is considered inadequate and the model is adequate if \( Q^* \) calculated is less than \( \chi^2 \) for \( h - p - q \) degrees of freedom. If the model is tested inadequate, then the forecaster should select an alternative model and test for the adequate of the model (Spyros et al., 1998).

2.5 Measuring Forecasting Accuracy

2.5.1 Mean Absolute Error (MAE)
The Mean Absolute Error (MAE) is defined mathematically by equation (9).

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i| \]

(9)

2.5.2 Mean Squared Error (MSE)
Symbolically, the Mean Squared Error (MSE) is defined by equation (10).

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \]

(10)

where \( X_t \) is the actual observation for time period \( t \), \( F_t \) is the forecast value for the same period and \( e_i = X_t - F_t \) is the error term and \( n \) is the number of forecasting values.

3 DATA ANALYSIS AND RESULTS

3.1 Time Plot of Prevalence of Measles in the Ashanti Region of Ghana
In general, the trend in measles prevalence in the Ashanti Region of Ghana seems to be decreasing, but not always the case. The annual measles time plot in Figure 1 does not exhibit seasonal variation. Most of the data points are a little bit far apart from the mean. This indicates that there is a clear case of non-stationarity in the mean. It follows that the measles series is non-stationary in the mean.
3.1.1 Stationarity checks using the ACF and PACF

Figure 1: Time plot of measles prevalence in the Ashanti-Region of Ghana from January 2001 to November 2011.

Figure 2: The Autocorrelation Function and Partial Auto-Correlation Function of prevalence of measles

Figure 2 depicts the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the measles data. The lags from 1 to 18 autocorrelation exceed two standard errors above zero (they are significantly far from zero). The autocorrelation function is decreasing gradually with time and that shows that there is a non-stationarity in the measles data. The second part of Figure 2 exhibits the partial autocorrelation function (PACF) of the measles data. The first lag is almost unity which confirms that the measles time series is non-stationary.
Table 1: ADF and KPSS Tests

<table>
<thead>
<tr>
<th>Augmented Dickey – Fuller Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dickey-Fuller Lag Order P-Value</td>
<td></td>
</tr>
<tr>
<td>-2.7315 5 0.2723</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KPSS Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KPSS Level Lag Parameter P-Value</td>
<td></td>
</tr>
<tr>
<td>2.0023 2 0.01</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows both Augmented Dickey-Fuller and KPSS test results. There was no stationarity in the original measles data; since p-value for ADF test was greater than 0.05 and that of KPSS was less than 0.05 using a 5% significant level. The two tests confirmed that there was non-stationarity in the original measles data which needs to be differenced to achieve mean stationarity.

3.2 First-Order Difference of Prevalence of Measles in the Ashanti Region from January 2001 to November 2011.

From figure 3 a transformation of the measles data using first-order difference was performed to remove the non-stationarity in the original measles data. The data fluctuate around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time. There was not any seasonal behavior in the time plot, and the measles data now looks to be approximately stable for further investigations.

![Figure 3: First-Order Differencing of Prevalence of Measles in the Ashanti Region](image)

3.2.1 Objective Test for Stationarity for the First-Order Differenced Series

From Table 2 we can observe that there is a stationarity in the measles data, since p-value for ADF test was less than 0.05 and that of KPSS was greater than 0.05. The measles data now looks to be approximately stationary in the mean for further investigations.
Table 2: Augmented Dickey-Fuller (ADF) and KPSS Tests

<table>
<thead>
<tr>
<th>Augmented Dickey – Fuller Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dickey-Fuller Lag Order</td>
<td>P-Value</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-7.003</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KPSS TEST</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KPSS Level Lag Parameter</td>
<td>P-Value</td>
</tr>
<tr>
<td>0.0273</td>
<td>2</td>
</tr>
</tbody>
</table>

3.3 Selecting Competing Models Using ACF and PACF of the First-Order Differencing of Measles Prevalence

The first part of Figure 4 shows the sample Auto-Correlation Function (ACF) of the first order differencing of the prevalence of measles. Except for marginal significance at lags 2, 6 and 8, the model seems to have captured the essence of the dependence in the series. Inspecting the sample ACF, we see that PACF is tailing off and the ACF is cutting off at lags 2, 6 and 8. This would suggest that the prevalence of measles follows an MA (2) model, MA (6) model or MA (8) model. The second part of Figure 4 shows the sample Partial Auto-Correlation Function (PACF) of the first order differencing of the prevalence of measles in the Ashanti Region of Ghana at different lags. Inspecting the sample PACF, we see that the ACF is tailing off and the PACF is cutting off at lags 2 and 6.

Figure 4: ACF and PACF of first-order differencing of prevalence of measles in the Ashanti Region of Ghana

Except for marginal significance at lags 2 and 6, the model seems to have captured the essence of the dependence in the series. This suggests an AR (2) or AR(6) model for the measles prevalence. As a preliminary analysis, we will fit both models. It follows that, in both the ACF and the PACF of the first order differencing of the measles data, the following models were suggested:

ARIMA (0, 1, 2), ARIMA (2, 1, 0), ARIMA (1, 1, 0) and ARIMA (2, 1, 1)
3.4 Estimation of Tentative Models

3.4.1 Parameter estimate and diagnostics of ARIMA (0, 1, 2) model

Table 3: Parameter estimate for ARIMA (0, 1, 2) with non-zero mean.

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>Estimate</th>
<th>St. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ma 1</td>
<td>0.1311</td>
<td>0.0947</td>
<td>1.3844</td>
</tr>
<tr>
<td>ma 2</td>
<td>-0.2968</td>
<td>0.1122</td>
<td>2.6453</td>
</tr>
</tbody>
</table>

The coefficients of the estimated MA (2) parameters are within the invertibility condition bounds as shown in Table 3. From Table 3, the estimated ARMA (0, 2) model can be written as shown in equation (11):

\[ X_t = 0.1311 \omega_{t-1} - 0.2968 \omega_{t-2} - 1.8115 + \omega_t \]  

(11)

Table 4: Box-Ljung test of ARIMA (0, 1, 2) with non-zero mean

<table>
<thead>
<tr>
<th>Box - Ljung Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Squared</td>
</tr>
<tr>
<td>17.0893</td>
</tr>
</tbody>
</table>

Results from Table 4 showed that the model’s residuals were non-significant with Ljung Box test statistic of 17.0893 and a p-value of 0.6472. Hence the model was adequate for forecasting.

3.4.2 Parameter estimate and diagnostics of ARIMA (2, 1, 0) model

Table 5: Parameter estimate for ARIMA (2, 1, 0) with non-zero mean

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>Estimate</th>
<th>St. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar 1</td>
<td>0.1703</td>
<td>0.0852</td>
<td>1.9988</td>
</tr>
<tr>
<td>ar 2</td>
<td>-0.2311</td>
<td>0.0861</td>
<td>2.6841</td>
</tr>
</tbody>
</table>

The coefficients of the estimated AR(2) parameters are within the causality condition bounds as shown in Table 5. The estimated model for ARMA (2) was given by equation (12):

\[ X_t = 0.1703X_{t-1} - 0.2311X_{t-2} - 1.7820 + \epsilon_t \]  

(12)

Table 6: Box-Ljung test of ARIMA (2, 1, 0) with non-zero mean

<table>
<thead>
<tr>
<th>Box - Ljung Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Squared</td>
</tr>
<tr>
<td>18.4227</td>
</tr>
</tbody>
</table>
Results from Table 6 shows that the model’s residuals were non-significant with Ljung Box test statistic of 18.4227 and a p-value of 0.5596. Hence the model was adequate for forecasting.

3.4.3 Parameter estimates and diagnostics of ARIMA (1, 1, 0) model

Table 7: Parameter estimate for ARIMA (1, 1, 0) with non-zero mean

| Coeff. | Estimate | St. Error | |t – Value| |
|--------|----------|-----------|-----------------|
| ar1    | 0.1399   | 0.0867    | 1.6136          |

<table>
<thead>
<tr>
<th>AIC</th>
<th>AICc</th>
<th>BIC</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1426.28</td>
<td>1426.47</td>
<td>1434.88</td>
<td>-2.0022</td>
</tr>
</tbody>
</table>

The coefficient of the estimated AR (1) parameter is within the causality condition bounds as shown in Table 7. The estimated ARMA (1, 0) model can be written as shown in equation (13).

\[ X_t = 0.1399 X_{t-1} - 2.0022 + \varepsilon_t \]  

(13)

Table 8: Box-Ljung test and Forecasts from ARIMA (1, 1, 0) with non-zero mean

<table>
<thead>
<tr>
<th>Box - Ljung Test</th>
<th>X Squared</th>
<th>d f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.0519</td>
<td>20</td>
<td>0.0864</td>
</tr>
</tbody>
</table>

Results from Table 8 shows that the model’s residuals were non-significant with Ljung Box test statistic of 29.0519 and a p-value of 0.0864. Hence the model was adequate for forecasting.

3.4.4 Parameter estimate and diagnostics of ARIMA (2, 1, 1) model

Table 9: Parameter estimate for ARIMA (2, 1, 1) with non-zero mean

| Coeff. | Estimate | St. Error | |t-value| |
|--------|----------|-----------|-----------------|
| ar1    | 0.9801   | 0.1051    | 9.32450         |
| ar2    | -0.2693  | 0.089     | 3.02580         |
| ma1    | -0.9107  | 0.0781    | 11.6607         |

<table>
<thead>
<tr>
<th>AIC</th>
<th>AICc</th>
<th>BIC</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1417.05</td>
<td>1417.53</td>
<td>1431.39</td>
<td>-1.9980</td>
</tr>
</tbody>
</table>

The coefficients of the estimated ARMA (2, 1) parameters are within the causality and invertibility condition bounds. The estimated ARMA (2, 1) model can be written as shown in equation (14).

\[ X_t = 0.98X_{t-1} - 0.2693X_{t-2} - 0.9107\omega_{t-1} - 1.9980 + \omega_t \]  

(14)

Table 10: Box-Ljung test and Forecasts from ARIMA (2, 1, 1) with non-zero mean

<table>
<thead>
<tr>
<th>Box - Ljung Test</th>
<th>X Squared</th>
<th>d f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.0093</td>
<td>20</td>
<td>0.5868</td>
</tr>
</tbody>
</table>
Results from Table 10 showed that the model’s residuals were non-significant with Ljung Box test statistic of 18.0093 and a p-value of 0.5868. Thus the model was adequate for forecasting.

### 3.5 Forecasting From ARIMA (2, 1, 1)

**Table 11:** *Point Forecasts from ARIMA (2, 1, 1) with non-zero mean*

<table>
<thead>
<tr>
<th>Point Forecast From ARIMA (2, 1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec-11</td>
</tr>
<tr>
<td>15.9037</td>
</tr>
</tbody>
</table>

From Figure 6, the yellow line depicts the 95% confidence interval, the red line is the 85% confidence interval and the blue line is the forecasting points. The model was used to forecast six months ahead and showed that the measles prevalence in the Ashanti Region will be increased from December, 2011 to May, 2012 as shown in the blue line as well as Table 11, its forecast points. The original measles data from December, 2011 to May, 2012 is given in Table 11 (ii) below.

**Figure 6:** *Forecasts from ARIMA (2, 1, 1) with non-zero mean*

**Table 11 (ii):** *Original Measles Data from December, 2011 to August, 2012*

<table>
<thead>
<tr>
<th></th>
<th>Dec-11</th>
<th>Jan-12</th>
<th>Feb-12</th>
<th>Mar-12</th>
<th>Apr-12</th>
<th>May-12</th>
<th>Jun-12</th>
<th>Jul-12</th>
<th>Aug-12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
<td>17</td>
<td>11</td>
<td>16</td>
<td>13</td>
<td>15</td>
<td>9</td>
<td>8</td>
<td>18</td>
</tr>
</tbody>
</table>
3.6 The Error Metrics
Forecasting accuracy based on the Mean Absolute Error (MAE) of the forecasted values was checked for each fitted model as shown in Table 12. It highly favored the forecasted value of ARIMA (0, 1, 2). This means that, the ARIMA (0, 1, 2) forecast error of 27.98298 out-performed all the forecast errors so far as the MAE is concerned.

Similarly, the forecasting accuracy based on the Mean Squared Error (MSE) of the forecasted values also favored ARIMA (2, 1, 1), the best selected model. This means that, the ARIMA (2, 1, 1) forecast error of 2947.151 out-performed all the forecast errors so far as the MSE is concerned. Hence ARIMA (2, 1, 1) was confirmed to be the best model.

Table 12: The Mean Absolute Error (MAE) and the Mean Squared Error (MSE)

<table>
<thead>
<tr>
<th>MODEL</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1, 1, 0)</td>
<td>29.3936</td>
<td>3277.67</td>
</tr>
<tr>
<td>ARIMA(2, 1, 0)</td>
<td>28.4420</td>
<td>3103.25</td>
</tr>
<tr>
<td>ARIMA(2, 1, 1)</td>
<td>28.1141</td>
<td>3069.11</td>
</tr>
<tr>
<td>ARIMA(0, 1, 2)</td>
<td>27.9830</td>
<td>2947.15</td>
</tr>
</tbody>
</table>

4.0 CONCLUSION, FINDINGS AND RECOMMENDATION

4.1 Conclusion

From Table 13 ARIMA (2, 1, 1) had the lowest AIC value of 1417.05. It was the best model so far as the AIC was concerned.

Table 13: Summary of Diagnostics Test

<table>
<thead>
<tr>
<th>MODEL</th>
<th>AIC</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA (0, 1, 2)</td>
<td>1419.93</td>
<td>1420.25</td>
</tr>
<tr>
<td>ARIMA (2, 1, 1)</td>
<td>1421.28</td>
<td>1421.60</td>
</tr>
<tr>
<td>ARIMA (1, 2, 0)</td>
<td>1426.28</td>
<td>1426.47</td>
</tr>
<tr>
<td>ARIMA (2, 1, 1)</td>
<td>1417.05</td>
<td>1417.53</td>
</tr>
</tbody>
</table>

Hence the best selected model was

\[ X_t = 0.98X_{t-1} - 0.2693X_{t-2} - 0.9107\omega_{t-1} - 1.9980 + \omega_t \]

However, the forecasting accuracy based on the Mean Squared Error (MSE) for ARIMA (2, 1, 1), the best selected model, was the lowest. Its forecast error of 2947.151 out-performed all the forecast errors. In conclusion, the research study reported in this monograph has found that measles data in the Ashanti Region of Ghana could best be modeled with ARIMA (2, 1, 1). The study again found out that measles prevalence in the Ashanti Region is expected to increase if no preventative measures are taken. In other words, there is going to be an irregular pattern in measles (i.e increasing and decreasing movement of measles) from December, 2011 to May, 2012 as shown in Figure 6.
This model did not consider mass vaccination as one of the methods to prevent the prevalence of measles in the region. The results of this paper can be used as a tool to facilitate the introduction of measles vaccine and improve measles vaccination in the country as a whole.

4.2 Findings:

Results from this paper shows that increasing use of measles vaccine is having a significant impact on the rate of measles transmission and its related complications in the Ashanti region of Ghana.

Increasing the measles vaccination coverage rate in the region will further decrease the prevalence of measles in the region and decreasing the vaccination coverage will increase the rate of transmission of measles in the region which will affect the development of human resources in the country.

From definitions, the measles data in the Ashanti Region of Ghana from 2001 to 2011 depicts a Stochastic Time Series and its model is linear.

4.3 Recommendation

The following recommendation is made to the stakeholders of Ashanti Region:
Ghana Health Service should continue the mass measles vaccination in the region to possibly eradicate the disease or reduce it to the barest minimum.

REFERENCES


Wikipedia, the free encyclopedia (2008), en.wikipedia.org/wiki/Measles

Wikipedia, the free encyclopedia (2011), en.wikipedia.org/wiki/Ashanti_Region

