KWAME NKRUMAH UNIVERSITY OF SCIENCE AND

TECHNOLOGY, KUMASI

TIME SERIES ANALYSIS OF MALARIA CASES IN THE BEREKUM MUNICIPALITY

BY

KISSIWAA VIVIAN

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MSC. APPLIED STATISTICS

SEPTEMBER, 2019

Declaration

I hereby declare that this submission is my own work towards the award of the MSc. Applied Statistics degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Kissiwaa Vivian ....

Signature Date

Certified by:

Mr. Isaac Akpor Adjei ....

(Supervisor) Signature Date

Certified by:

Prof Mrs. Atinuke Olusola Adebanji ....

(Head of Department) Signature Date

Dedication

I dedicate this work to Almighty Father in Heaven for the life and great care, knowledge and understanding of the situations in the universe and my family.

Abstract

Health is a very essential aspect of human life. It is however imperative to know that one of the endemic diseases which deteriorates our health, Malaria poses a threat to 40% of the world's population. In Ghana, it has become one of the leading causes of mortality and morbility particularly among pregnant women and children under five years. This research, therefore, sought to model the trend of Malaria cases from 2011 to 2015 and to forecast the incidence of Malaria cases for the years 2016 to 2018. The time series methods were used to explore the historical pattern of the variable of interest. The study adopted quantitative research approach and design. The pragmatic worldview as an interpretive framework which made use of quantitative methods in collecting data from the Berekum Municipal Hospital. Data collected during the study showed three trends from 2011 to 2017. From 2011-2013, a similar trend was seen where two peaks were recorded in a year (the middle and towards the end of the year) being the highest cases of malaria recorded. It further declined from 2014 to 2015 and continued for the predicted time series cases from 2016 to 2017. Therefore the community should expect a decline in malaria cases in subsequent years but there should be a system to monitor the trend in case there is any change. The trend of malaria cases should be modeled for different localities since climate factors may affect the prevalence and incidence of malaria.

Acknowledgements

My profound gratitude goes to the Almighty God for the resilience and strength He gave me throughout this study. My heartfelt gratitude goes to my husband for the selfless effort towards the program of study. In connection with the study, I wish to sincerely acknowledge the contribution of my supervisor, Mr. Isaac Akpor Adjei, a researcher who put me right on various errors throughout the writing of my thesis and also for his professional encouragement. I would also like to express my sincere

gratitude to the entire teaching staff of the Department of Mathematics, KNUST-Kumasi for their wonderful support and encouragement.

May God bless all of you.

Contents

List of Abbreviation

AR .....Autoregressive

MA .....Moving Average ARMA .....

Mixed Autoregressive and Moving Average

AIC ..... Akaike's Information Criteria SBC

..... Schwarz's Bayesian Criteria

BIC .....Bayesian Information Criteria

OPD .....Out Patient Department

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Overview of Study

According to Organization (2014), Malaria is a common and life-threatening disease in many parts of Africa. For the past century after discovering the mechanism with which it is transmitted, the cases of malaria remains prevalent in about 100 countries posing a threat to approximately 3.3 billion people.

Among vector-borne diseases, Malaria disease causes one of the biggest peril affecting about 40% of the population globally and making them prone to many infections. A survey conducted in 2013 indicated that, 97 countries had ongoing transmission of Malaria. According to the World Health Organization (2014) and Oduro et al. (2015), the disease was responsible for 20% to 40% of the outpatients visits and 10% to 15% of the patients admitted at the hospital. According to Organization (2014), in 2013, it was estimated that 3.4 billion people are prone to Malaria. In the year 2012 alone, an estimated 207 million cases of Malaria was brought fore and an estimate of 627000 death cases from the menace were recorded worldwide. In 2010, according to the World Health Organization (WHO) (Organization et al., 2012b), 219 million cases of Malaria were recorded globally.

According to research conducted by various health organizations, Sub-Saharan Africa (SSA) accounts for 90% of the world's 300 to 500 million malaria cases and 1.5 to 2.7 million deaths annually. The Malaria pandemic poses a great burden to the health system in Africa. According to Asenso-Okyere et al. (2009), there are about 300 million cases of Malaria which brings about acute illness yearly which cost more than US$ 12 million every year in Africa and it retard economic growth by 1.3% per annum.

According to World Health Organization et al. (2012b), the disease was liable for 9% of all the deaths in Africa.

Malaria continues to be among one of the leading causes of morbidity and mortality especially with x children who are aged under five and women who are pregnant according to Worrall et al. (2005). According to Orkoh and Annim (2017), the cases of deaths related to Malaria has increased from 79.7 per 1000 and 3,378 deaths in the year 2009 to 108.3 per 1,00 and 3, 859 deaths in the year 2010 respectively.

The Health Management Information System (HMIS) run by the Ministry of Health in Ghana provides a weekly record of the number of patients treated for Malaria found in different hospitals across the country. According to Takyi Appiah et al. (2015), Malaria poses a serious health problem to Ghana. On an average 30,300 cases of Malaria are attended to dairly in the country's health facilities. Also, according to National Malaria Control Program (NMCP), Ghana in 2013 recorded about 11.3 million Malaria cases at Outpatients Department (OPD) (Takyi Appiah et al., 2015). Malaria massively affects all the segments of the society. The impact of Malaria is not seen only in the health sector, but also in the social and economic life of a country (Guinovart et al., 2006). Even though Malaria is treatable and preventable, it causes a significant mortality and mobility especially in regions which are poorly resourced.

Malaria is predominantly referred to as a disease which affect the poor or a disease of poverty. According to Sachs and Malaney (2002), the growth of the economy every year of countries with massive Malaria cases was 1.3% lower than that of the countries which does not have Malaria. Malaria's prevalent rate is very high in the tropics as a result, it is considered the disease of the poor (Sachs and Malaney, 2002).

1.2 Problem Statement

Globally, one of the biggest threat of risk of disease infection is Malaria which affects about 40% of the world's population. Malaria decreases economic growth by more than one percentage point per year in endemic countries. In spite of various recent

advancements in diagnostic and treatment modalities, the menace of Malaria still remains a case of public health in Ghana. In Ghana Malaria continues to remain the leading cause of mortality and morbidity with pregnant women and children who fall under five years. An incidence of deaths associated with malaria has increased according to statistics shown by various researchers in Malaria cases. Despite the devastating effects of the disease on the economy of Ghana, the important malaria-free environment concept has not been fully appreciated by authorities in Ghana. Malaria affects all segments of society. While its burden is seen on people of all ages, the most serious impact of Malaria is on children and women who are pregnant because of their weak immunity. In Africa, most of the deaths are reported in infants, children and pregnant women. According to Takyi Appiah et al. (2015), the lead cause of death in children who are under five years in Africa is Malaria. Malaria research in Ghana has shown that the disease is not homogeneous and uniform in Ghana. It is a localized problem with great differences from place to place. Prevalence shows diversities and variations to the extent that neighboring communities can display complete difference in transmission patterns. This is due to a combination of factors: meteorological and ecology of both human host and the vectors and proximity between human habitat and the breeding places.

Studies conducted by Asenso-Okyere et al. (2009), showed that in Ghana, the raining season is the uttermost time for malaria transmission and it's mostly happen together at the peak period of agricultural activities such as planting and harvesting. Also the number of Malaria cases reported in various the various hospital is inconsistent with time. The variation in time is susceptible to creating an artificial drifts in the observed data hence the need for a time series analyses on Malaria cases to decompose the observed variation into trends, cyclic effect or irregular fluctuations. Perhaps the impact of Malaria cases has not been quantitatively validated to persuade policy makers, managers of program, politicians and development partners to dedicate the

looked-for attention and resources to combating this dreadful disease. This research makes an attempt to fill the gap with an appropriate information.

1.3 Objective of the Study

The objectives of the research are;

- To model the trend of malaria cases in the Berekum Municipality
- To forecast the trend of the incidence of malaria cases in the Berekum Municipality

1.4 Methodology

Time series as employed by Takyi Appiah et al. (2015) will be used for this survey. The data for monthly reported cases of Malaria fever between the years 2010 to 2015 was obtained from the Statistical Department of Holy Family Hospital of Berekum Municipality. Data from January 2010 to December 2015 was used to comparing the mean monthly cases. The data sourced from Malaria cases from the year 2010 is used as the base year and the mean reported cases is compared with the mean reported cases of the other years (2011 to 2015) to see the trend of occurrence of the disease over the years under consideration. The data from January 2010 to December 2015 would be for the development of a time series ARIMA model in forecasting the cases reported in the Berekum Municipality. The Box-Jenkins method was used in the model identification,estimation of the model parameters, model adequacy and forecasting

1.5 Scope

1.5.1 Contextual Scope

The study is confined to Malaria fever cases reported at the Holy Family Hospital in the Berekum Municipality. The project sought information on monthly outpatient cases for five (5) consecutive years (2010-2015) of the disease.

1.5.2 Geographical Scope

The geographical scope covers the Berekum Municipality which is located in the Brong Ahafo Region. The Wenchi Municipality and the Jaman District which are found on the Northeastern and Northwestern respectively are the boundaries that the Municipality shares borders with. It shares borders in the South with the Dormaa Municipality and the East with Sunyani West District.

1.6 Justification of the Research

Malaria is one of the greatest public health issue but currently, there are few research works which have been carried out on the incidence of Malaria in the Municipality of Berekum. It seems that must studies has not been done on malaria in the Municipality. Based on this, it is therefore expedient to furnish interest groups and decision makers with an important information concerning the incidence of Malaria found in the Municipality for possible policy alternative and to stimulate further research.

1.7 Limitation of the Study

This project encountered a lot of constraints for its successful accomplishment. Officers of the statistical department of the hospital had to search through bulky documents for the needed information.

1.8 Research Organization

The study is categorized into five sections. Chapter one gives a general introduction of the study with popular emphasis on the overview of the study, problem statements, the objectives of the study, justification of the problem, the research methodology and the limitations of the study. Chapter two focuses on the literature review on Malaria with emphasis on conceptual and theoretical clarifications which are spelt out to know the precise time series modeling to be used. The method used in the analysis is seen in Chapter three which comprises the introduction, describing basic statistics, method and the time series concept and the methodology associated with Box- Jenkins. Chapter four deals with an analysis and discussion section where the features of the area together with the data used in the study are made known in order to draw inferences or read meanings into the data collected. Chapter five ends the study by emphasizing on the summary of findings, the recommendations and the conclusions associated with the implication of the study.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter reviewed the necessary literature on the objectives of the research. It therefore considered issues related to time series analysis of malaria from the Global, African and in the Ghanaian perspective.

2.2 Background of Malaria

Historically, malaria has been a major cause of deaths all over the world. Carrington (2001) has noted that malaria is no longer a burden to the developed world and has also been controlled by most developing countries, Malaria is still endemic in sub-Saharan Africa and continues to be the number one killer, despite several attempts to eradicate malaria in sub Saharan Africa in which Ghana is a part.

Although malaria is preventable, a lot of people continually die from such a preventable disease. Rapidly growing cities like Berekum are also characterized by environmental problems such as poor refuse management, stagnant pools of water, bushes around houses, chocked gutters and other unsanitary conditions which provide breading grounds for mosquitoes. In this context, eradication of malaria is likely to be difficult despite numerous protective measures which might be available. This is because the prevalence of malaria is dependent on the abundance of the female anopheles' mosquito. This is where a particular environment provides enough breading grounds for mosquitoes. The rate at which it bites is the number of times it is able to bit its host while the person is asleep or awake. The plasmodium parasite in the mosquito determines the rate of Malaria development. With all these conditions prevailing in Berekum. It therefore becomes important for this research to be taken in ascertaining the series of time that the cases of Malaria have been reported in contemporary times so that policy intervention on this menace can be meted out and also to know the successes that can be chalked when an efficient time series analysis is used as well as the challenges that this analysis faced and the way forward for the Ministry of Health as far as malaria continues to be among the Top Ten Disease chart in Ghana.

2.3 The Causes of Malaria

Carrington (2001), posits that the spread of Malaria can be from one person to another by a bite of a mosquito which is infected. When a human is bitten by a mosquito, the malaria parasite is passed into the blood stream and swiftly makes its way into the liver. When it is found in the liver it multiplies and invades the bloodstream eventually. The mosquito which is infected can bite at any time in the day but most of its attacks are in the night at down. Sometimes the parasite of Malaria can be handed onto a baby from its mother but this is however very rare. Mosquitoes are insects who have two wings with their females having long proboscis to suck the blood by piercing the skin of humans and animals as a vector of malaria. Any agent (person, animal or microorganism) who carries and transmits a disease is known as a vector. The vector responsible for the transmission of the plasmodium parasite (a parasite is an organism that feeds on it host to benefit and at the same time causes harm to the host) is the female mosquito known as the anopheles gambiae mosquito.

According to Carrington (2001), four different protozoa in the plasmodium genus: the Plasmodium vivax, Plasmodium ovale, Plasmodium malaria, or the Plasmodium falciparum. The Plasmodium falciparum is the most dangerous among the four responsible for causing malaria. In the low endemic areas, the Plasmodium vivax is the most predominant. Low endemic areas are areas where malaria is not so prevalent and the high endemic areas are where the malaria is prevalent. The Plasmodium falciparum is the dangerous of the four which is more prevalent in high endemic areas like sub-Saharan Africa.

Carrington (2001) further states that, a person only gets infected with the malaria parasite when he or she has been bitten by the female mosquito that has already bitten and sucked the blood of an infected person. The incubation of Malaria is developed under favorable conditions grow and become strong in the human host

for about eight to ten days. According to the Ghana Association of Teachers Biology book (GAST), mosquitoes and the plasmodium parasite needs certain favorable conditions for their survival. Temperatures of approximately 21 to 32 degrees Celsius at a relative humidity of at least 60% are conducive for their breeding and survival within the mosquito. This document further states that, the development of the malaria parasite increases more rapidly as temperature rises and ceases entirely below 15.6 degree Celsius. Increased rainfall which results in stagnant pools of water or water from the surface of the earth provides the breeding grounds for the mosquitoes which results in an increase in an increasing number of mosquitoes relative to the population of human which intend increases the tendency of mosquito's bit on human host and hence causing an increase in the population affected with Malaria.

Carrington (2001) affirms that, over 300 million new cases of malaria cases are detected every year with roughly two to three million deaths as a result of the contraction of the disease. In the tropical African region malaria is very pandemic with an estimated 90% of the malaria cases and with deaths among women who are pregnant and children.

2.4 Symtoms of Malaria

Previous knowledge gathered from the Biology GAST gives these symptoms as being most common with malaria disease. These are as follows:

- Fever and high temperature: This has the sign of a rise in the body temperature.
- Chills: This comes with a sensation of cold which often marks the beginning of an infection and fever development;
- Headaches: A dilation of cerebral arteries or contractions of the muscle or reaction to drugs which causes pains in the head.

- Muscle aches: Pain in the contractile organs of the body;
- Confusion: A feeling of embarrassment that leaves you confused;
- Dizziness: A reeling sensation; a feeling that you are about to fall;
- Vomiting: lasting for several hours- This is a reflex act where the content of the stomach through the mouth is ejected;
- Sweating: The process of the sweat glands of the skin secreting a salty fluid; and
- Tiredness: Temporary loss of strength and energy.

According to the same sources, all persons do not develop the same symptoms when they have malaria, everyone has different symptoms and different rates of development. Malaria is developed within 6 -8 days after bite from a mosquito that is infected or as late as several months after the preliminary infection. Although it is uncommon, it is possible for a person to develop malaria some months and sometimes years after departing from a country with malaria parasite.

2.5 Effects of Malaria

Along with Malaria morbidity and mortality comes with economic and social losses. Morbidity refers to the rate at which people fall sick from a particular disease, for instance the rate at which people contract malaria. On the other hand, mortality refers to the rate at which people die from contracting a disease. The severity of malaria is directly associated with the economic and social consequences as it increases mortality and morbidity. Malaria has caused able men and women to spend days or weeks out of work which tend to reduce their productivity. The adults lose workdays whiles the chidren spend several days away from school. Also, money is spent on buying drugs and paying for hospital bills which could have been used on other profitable activities. The worse scenario is loss of life of the productive population through malaria.

According to Worley (2006), the older population have developed some collective immunity to malaria in high endemic areas. Those that are higher risks are women who are pregnant, children and those who have not been immune. The same source further states that, pregnancy aggravates malaria through a nonspecific hormone-dependent depression of the immune system. The defensive antiplasmodium activity is suppressed during pregnancy. Also children who are under the age of five easily contract malaria because their immune system does not have antiplasmodium to fight against the plasmodium parasite. Antiplasmodium is the human immune system that helps the body to fight against plasmodium that causes malaria.

2.6 Early Attempt of Controlling Malaria

During the 1950s a programme was carried out by the World Health Organization in an attempt to exterminate malaria from some parts of the world. The population of mosquito were controlled by the use of insecticides mainly DDT.

Malaria was eradicated from the United States, Southern Europe and the former Soviet Union. But it was clearer in hotter and more humid climates like SubSaharan Africa where mosquitoes are found all year round and are resistant to the insecticides being used.

The other approaches employed in an attempt to control malaria including the use of drug to treat the whole populations failed because the parasites of the malaria became resistant to the drugs.

There was a confirmation that by 1970, it became obvious that the program had failed because the number of cases of malaria was still increasing.

According to Organization (2015), the global and regional strategies for Malaria control were established in Africa in 1991 and 1992 respectively. In the year 1995 the World Bank and the World Health Organization Regional Officer for Africa developed a long lasting collaboration on malaria control. In the year 1997, the Organization of

African Unity Summit led to the establishment in April 1998 of African Initiative for Malaria Control (AIM) and in July 1998 as a result of the political support given to Malaria control. The Roll Back Malaria (RBM) project was established by the Director General of World Health Organization (WHO). Because the concepts and the goals of the two programmes were alike, there was an agreement that AIM would be called RBM in the African Region.

2.7 Malaria Transmission

Studies have revealed that, Malaria in humans is known to be instigated by the four species of Plasmodium parasites; Plasmodium vivax, Plasmodium malariae, Plasmodium ovale and Plasmodium falciparum

The genus Plasmodium which is a unicellular protozoan parasites causes malaria transmitted between humans by the female anopheles mosquito. The lifecycle of the plasmodium is made up of multiple stages in both mosquito and human hosts.

When a mosquito takes a blood meal from an infected human, male and female Plasmodium gametocytes may be ingested. The sexual reproduction of the parasites takes place within a vector mosquito's stomach which results in the development of sporozoites which migrate to the salivary glands of the mosquito. The sporozoites can contaminate another human entering the bloodstream upon a subsequent blood meal which then become sequestered in the liver, reemerging into the blood, reproducing asexually in the erythrocytes stages, and finally producing gametocytes to complete the cycle Wikipedia contributors (2019).

According to Artavanis-Tsakonas et al. (2003), the Plasmodium does battle with the complex human immune system during the stages in human host by attacking the gametocyte and erythrocyte stages. How effective the immune system response depends on the system memory from previous exposure to related parasites. The

manner in which malaria is transmitted depends upon the availability of susceptible humans, infected humans and the mosquitoes having sufficient longevity. The longevity of mosquito is precarious for the viability of the Plasmodium lifecycle since the time taken for the Plasmodium to undergo ingestion, reproduction, development and retransmission to a human host is analogous to the mean lifespan of the mosquito Gilles and Warrell (2002). The Plasmodium falciparum causes the most severe and fatal human malaria infection.

The other widespread species is the Plasmodium vivax.

2.8 Malaria in Ghana

According to Akpalu and Codjoe (2013), in Ghana there are three malaria epidemiologic zones: the northern savannah, the coastal savannah or mangrove swamps and the tropical rainforest. It is in these three zones that the country's entire population is spread across and subsequently it is at risk of getting malaria.

According to Dadzie et al. (2013), there are two dominant species of the malaria vector species in Ghana: Anopheles funestus and Anopheles gambiae and these mosquitoes usually bite late at night and are commonly found in areas where favourable breeding sites exist. Research conducted by President's Malaria Initiative in Ghana (2013) found out that in the northern Savannah over 50% of the mosquito bite occur outdoor.

According to Akpalu and Codjoe (2013), the transmission of malaria in the country therefore occurs in all year round with variations between the dry and wet seasons. The normal period of the intense malaria transmission season is about seven months, beginning in April or May which last until September.

In Ghana an estimated 3.5 million people get malaria each year according to Organization et al. (2012a). For instance in the year 2006, the prevalence of malaria

per thousand people was estimated at 171. As at the same year there were 2835 deaths attributed to the malaria disease representing 19% of all recorded deaths according to Manda (2017). In addition to this malaria accounted for 38.6% of all outpatient illnesses and 36.9% of all admissions according to Akpalu and Codjoe (2013).

There are several researches on trend of malaria in Ghana. A survey conducted by Ankamah et al. (2018) highlighted what remain the lead cause of mortality and morbidity among women who are pregnant and children who fall under five is malaria. The study also found out that Incidence of malaria and deaths associated with the menace of malaria increased from 79.7 per 1000 and 3,378 deaths in the year 2009 to 198.3 per 1,000 and 3,859 deaths as at 2010 respectively. Asamoah et al in the year 2008 in their work emphasized on SARIMA (1,1,0) (1,0,0) for Total OPD reported cases, ARIMA (1,1,0) for Admission reported cases, SARIMA (1,0,0) (1,0,0) for female OPD reported cases and SARIMA (1,1,0) (1,0,0) cases of malaria reported for OPD pregnant cases. Akpalu and Codjoe (2013) examined Economic Analysis of Climate Variability Impact on Malaria. This study found out that malaria prevalence increases with rainfall

2.9 Global Analysis of Malaria

Various study has been done on different aspects of malaria. Most of them centers on Parasitology to chemotherapy (getting a cure with drugs) to the eradication of the disease. Many extensive research has been conducted on different aspect of the disease which is a genetic diversity of the parasite (McKenzie et al., 2008) prior to the mosquito and parasite lifecycle (McKenzie and Bossert, 2005), the development of the drug resistance (Okosun et al., 2011); (Klein et al., 2012), and the exposure-dependent partial immunity (Dietz et al., 1974; Aron and May,1982; Filipe et al., 2007).

Analysis of Malaria cases in Bangladesh with remote Sensing Data was examined in (Abd, 2007). In this study, malaria cases with respect to epidemiologic data were correlated with satellite- based Vegetation Health (VH) indices to explore if they can be used as substitution for monitoring malaria epidemics in Bangladesh. The research study brought to the fore that during the cooler months, that is November to March, when the mosquitoes are less active, the correlation was low. During the wet and dry season, in April to October, there exist a considerable increase reaching 0.7 for the TCI in early October and -0.66 for the VCI in midSeptember.

Cao et al. (2000), piloted a research aimed at determining the effect of Permethrin Insecticide Treated Nets (PITN) use on the incidence of febrile episodes and nonhousehold malaria expenses in Benin. The survey found out that the PITNs decreased the risk associated with malaria by 34% in the rural areas' children. Meanwhile PITN use did not reduce treatment and prevention expenses.

In a laboratory of parasitology malaria was noted to be the major killer of paediatric illness and death in Kinshasa (Tio et al. (1991)). As a matter of fact the treatment of fevers as malaria with chloroquine is no longer satisfactory because the plasmodium falciparum had a resistance to Chloroquine. With reference to the study the differences in endemicity of malaria existed between the various parts of town which had to be taken into consideration alongside the socioeconomic and ecological factors that underlie when the estimate of potential control methods are planned.

2.10 Time Series and Malaria Studies

According to Freckleton (2009), the main reason for a time series is to develop a statistical model explaining the behavior of a random variable changing over time which allows the making of future estimation of random variable possible.

There are a number of purposes that the analysis of time series of malaria may serve.

A survey conducted by Zinszer et al. (2013) disclosed that the prediction of malaria has been piloted in about 13 different countries with China having the most frequent site for malaria forecasting. Geographically the size of the region of study ranges from the municipal level to a larger administrative divisions such as the provinces and country.

Most of the research studies of about 97% use health clinic records of malaria infections from the general population as their source of data for malaria infections

Dan et al. (2014) observed the modeling and forecasting of malaria mortality rate using SARIMA Models. The study employed the Box-Jenkins methodology to build ARIMA model for malaria mortality rate from January 1996 to December 2013 with a total of 2016 data points. The study found out that the most effective methods for analyzing time series data is the method propounded by Box and Jenkins which is the Autoregressive Integrated Moving Average (ARIMA).

Ostovar et al. (2016) examined the Time Series Analysis of Meteorological Factors which influences Malaria in South Eastern Iran. In this study, a total of 2002 locally transferred microscopically confirmed malaria cases which occurs in the Minab district of Hormozgan Province in Iran over a period of 6 years from the month of March 2003 to March 2009 were analyzed. Meteorological variables (the rainfall, relative humidity and temperature in the district) were also assessed. The study concluded that, the incidence of malaria with conventional precision in an early-warning system can be predicted using the Time-series models.

Over the last century, in many parts of the world, an analyses of time-series malaria cases shows that rainfall excess is correlated with changing incidence of malaria which occurs in certain eco-epidemiologic settings (He et al., 2005).

Kristan et al. (2008) examined the prevalence of malaria from a historical morbidity patterns in epidemic-prone areas of Ethiopia using time series. The purpose of the study was to assess the accuracy of different methods of forecasting malaria cases

from a pattern of historical morbidity areas with an unstable transmission. The survey tried five approaches using incidence data reported from the health facilities in the 20 areas of the central and north-western Ethiopia. The study showed that there is the need to have a forecasting incidence from a historical morbidity patterns alone with limitations and the indications of the need to improve epidemic early warning by incorporating external predictors such as meteorological factors.

Lin et al. (2009) used time series analysis to probe the relationship between the imported malaria in the non-endemic provinces of China and the falciparum malaria in the endemic provinces. An autoregressive integrated moving average model was first fit to the predictor variable. When the seasonal ARIMA (1, 1, 1) and (0, 1, 1) model for malaria incident was tested among all the models, it was the one that best fit the data according to the AIC and the criteria for the goodness-of-fit.

CHAPTER 3

Methodology

This brings to the fore the methodology; data collection method and analysis procedures.

3.1 Source of Data

A secondary data was collected from the Statistical Division of Berekum Municipal Hospital. The completion of this thesis is dependent on the relevant information which was sourced.

3.2 Data Analysis Procedure

There are two forms that the procedure for analyzing the data comes with; these are the quantitative. The quantitative data analysis technique was used when the design is quantitative. With the quantitative data analysis it converts data statistically to help the researchers the more exactly and make a fortitude about the features of the population on the basics of data from samples. Statistical forecasting methods which fall under Quantitative forecasting methods was used. Generally, the Quantitative methods are subjective in nature. This means in formulating the relationship they rely on experts' opinions in formulating relationship.

There are two sub-types that these methods can be further classified which are the Time Series methods and the Causal or Explanatory methods. The Causal methods investigate the occurrence of other variables which affect the variable of interest. The inputs which are the other variables are then analyzed and a suitable relationship between the variable of interest and the inputs is formulated. The Time Series is also called "stand alone" methods.

The methods used in Time Series explore the pattern of history present in the variable of interest and it further assumes this will continue in the near future. The prediction of future values makes use of this association. The availability of data and the specific cases that occurs determines the choice of the method to be used.

3.3 The Concept of Time Series

The set of observations on a variable of interest that has been collected in time order could be daily, weekly, monthly, et.c is termed as time series. In a situation where the time series is predicted exactly it is said to be deterministic for instance the salary of a person may be determined according to the number of years he or she has worked but more importantly the time series are stochastic in nature in that the future values are resolute based on the past values.

3.3.1 Components of Time Series

The traditional time series are mainly concerned with decomposition into the following four components: cyclical, irregular, seasonal and trend variations

3.3.2 Trend

The general tendency of a time series to increase, decrease or stagnate over a long period of time is referred to as trend. In other words, it is also seen as a long-term decay or growth.

A deterministic trend model together with a seasonal effect can take either an additive form

$$X_t = m_t + s_t + Y_t \tag{3.1}$$

Or a multiplicative form, such as

$$X_t = m_t s_t Y_t \qquad t = 0,1,\ldots,n \tag{3.2}$$

Or mixed form,

$$X_t = m_t s_t + Y_t \qquad t = 0,1,\ldots,n \qquad m_t = m(t), \qquad s_t = s(t) \tag{3.3}$$

Where $m_t = m(t)$ is usually a slowing changing function of time, so called "trend component". $s_t = s(t)$ is a periodic function of time and Y_t is a random noise component.

Model (3.2) can be easily transformed to the additive form by taking a logarithm of both sides.

Model (3.3) is often referred to as a multiplicative one.

3.3.3 Cyclic

This is the long-term oscillations or swing, about a trend line curves which may be periodic over an equal interval of time. It is a wave-like fluctuation about trend. The length and the amplitude of the cycle are not constant as in the seasonal component but may vary from one to the next.

3.3.4 Seasonality

This is an identical or almost identical patterns which a time series seem to follow. It occurs in a periodic change usually in the year's cycles. In other words it is a regular recurring variation or fluctuation.

3.3.5 Irregular Variations

The irregular constituent of a time series is the residual factor that accounts for the abnormalities of the actual time series value from what we would expect. This is caused by a short term unanticipated and non-recurring factors such as earthquakes, wars, floods that affect the time series. Since this component accounts for the random variability in the time series, it is unpredictable in its impact on the time series in advance.

3.4 Box-Jenkins ARIMA Models

Box-Jenkins ARIMA models make use of historical values of a single variable to predict its future values; hence they are classified as univariate methods. The variable of interest must be separated by equally spaced time intervals to apply Box-Jenkins methodology. This can be done by considering a discrete time series of equally spaced observations in time:

$$X_t = x_1, x_2, x_3, x_4, \ldots, x_{n-1}, x_n \tag{3.4}$$

The basic essence of Box-Jenkins methodology is that it considers the perceived time series, to be the outputs of an unobservable "black box" process. The inputs to this black box are a series of independent random shocks.

For statistical purposes, the random shocks are assumed to be normally distributed with zero mean and a constant variance. This sequence is typically referred to as "white noise". This means that the Box-Jenkins method views a time series as the result of a white noise process using a black box which is nothing more than a linear filter.

In essence, the ARIMA model assumes that the outputs (the values of the observed time series) may depend on the following:

- The previous and current inputs (white noise or random shocks); and
- The previous output values of the time series under study, $y_{t-1}, y_{t-2}, \ldots$ in varying proportion.

How much each of these will determine the future output will depend on their associated coefficients.

Specifically the Box-Jenkins approach proposes a simple linear form for the observed time series value:

$$Y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \cdots - \theta_q e_{t-q} \quad (3.5)$$

or

$$\Phi(B)(1-B)^d y_t = \Theta(B) e_t \quad (3.6)$$

Where $\Phi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p)$,

$\Theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$

B is the backward shift operator. i.e $By_t = y_{t-1}$, $B\ d =$

order of differencing.

Equation (3.5) depicts the current output as a linear weighted sum of previous inputs and outputs. Note that only "p" nonzero output terms and the "q" nonzero input terms are required to produce the current output.

The general notation of ARIMA models is ARIMA (p,d,q), where "p" is the order of Autoregressive component, "d" is the order of differencing used and "q" is the order of Moving Average component in the model. The Autoregressive and Moving Average components are described below and the concept of differencing is labelled in the next session. Depending on the above definition, the ARIMA models can be classified into:

- Autoregressive (AR) Models: This is where the current output y_t depends solely on p prior outputs and current input (random shock) e_t. In this case the Box-Jenkins model takes the form

$$Y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t \quad (3.7)$$

or

$$\varphi(B) y_t = e_t \quad (3.8)$$

and is called an Autoregressive model of order p, denoted by AR (p) or ARIMA $(p,0,0)$.

- Moving Average (MA) models: This is the situation where the current output rest solely on the current input and q prior inputs. The Box-Jenkins model takes the form

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (3.9)$$

or

$$y_t = \Theta(B) e_t \quad (3.10)$$

and is called a Moving Average model of order q, denoted by MA (q) or ARIMA $(0,0,q)$.

- Mixed Autoregressive and Moving Average (ARMA) models: When the current output y_t depends on both the AR and MA processes, the Box-Jenkins model takes the form of equation (3.5) and is called an Autoregressive Moving Average model, denoted by ARMA (p,q) or ARIMA $(p,0,q)$.

3.5 Stationary Time Series

ARIMA time series models are designed for stationary time series. The BoxJenkins methodology requires that the time series under analysis be "stationary" in both mean and variance.

In simplest non-statistical terms, the concept of Stationarity can be explained as follows:

- If the mean of the plotted series varies over time, the series is considered non-stationary in mean. If there is no evidence of a change in mean level over time, then the series is considered mean-stationary; and

- If there is no obvious change in the plotted series in the variance over time, then the series is considered to be stationary in variance over time, then the series is considered to be stationary in variance, otherwise it is considered to be non-stationary in variance. One of the other advantages of Box-Jenkins model is that it can be applied to non-stationary series after making them stationary using some sort of transformation. In order to induce mean Stationarity in the (mean non-stationary) data, typically, a concept called "differencing" is used. A difference of order one (or second or higher orders) is all that is required to achieve mean Stationarity in the majority of cases. A Difference of order one shows that each value of the time series is subtracted from the previous value which is immediate.

$$w_t = 5y_t = y_t - y_{t-1} \tag{3.11}$$

Note that $By_t = y_{t-1}$. Hence

$$y_t - y_{t-1} = (1 - B)y_t \tag{3.12}$$

Therefore, $5 = (1 - B)$ is the difference operator.

The result is a new time series w_t having one less observation than the original series. The difference of order two implies that the first order differenced series is differenced again resulting in

$$k_t = w_t - w_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \tag{3.13}$$

The result again is a new time series k_t, having two less observations than the original series.

This can be generalized to d^{th} order differencing when d is the order of differencing required to achieve mean stationarity. After modelling the order differenced series with an appropriate ARMA modelled values corresponding to the original undifferenced series, it therefore become essential to reverse the differencing transformation and "integrate" the times. This is represented by "I" in the acronym ARIMA and the order of integration is same as the order of differencing ("d" in this case). Next, in order to make the series stationary in

variance, if required, a different class of transformations can be carried out like the logarithmic transformation (taking log of the original data), square root, cubic root etc.

$W_t = \ln(y_t)$ (Natural logarithmic transformation)

$W_t = \sqrt{y_t}$ (Square root transformation), etc.

If the series is not positive throughout, it can be made positive by adding a suitable constant *c* to each observation of the series before the transformation is carried out. One of the best methods available to detect the proper transformation required to reduce heteroscadicity in data is using the Box-Cox transformation methodology, a general class of transformation which includes all other transformations mentioned earlier as special cases. The Box-Cox transformation also makes the data more normal distributionlike. Once the series of interest has achieved stationarity in both mean and variance using transformations, appropriate AR(p), MA(q), ARMA (p,q) or ARIMA(p,d,q) models can be fitted to the series.

3.6 Box-Jenkins Methodology

In the modelling of the Time Series, the Box-Jenkins methodology is used. The forerunners who propagated the approach which combines the autoregressive models were Box and Jenkins. In a Box-Jenkins series model there are four primary stages. These are model identification, estimation of the model parameters, diagnostic check of the residuals and model adequacy and forecasting.

3.6.1 Steps in Analyzing Data and Identifying ARIMA Models

Step 1- Identification of the order of the ARIMA model At the identification stage, the historical data of the time series of interest is statistically analyzed and an appropriate

subclass of models from the general ARIMA (p, d, q) family is selected. The approach can be summarized as follows:

a) Suitably transform the time series to remove the non-stationarity in variance (if present);

b) Difference the time series as many times as is needed to produce mean stationarity (if required), hopefully of reducing the process under study to the mixed Autoregressive Moving Average ARMA (p,q) process; and

c) Identify the order of the ARMA model. That is, identify the autoregressive order "p" and moving average order "q" present in the transformed and differenced data.

The basic tools for model identification steps (b) and (c) are the graphs of estimated sample. The estimated Sample of Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) obtained from the series. These graphs are used not only to help guess the form of the model but also to obtain approximate estimates of the parameters (using Yule-Walker equations), which are useful at the estimation stage to provide starting values for iterative procedures employed during the estimation of final parameters.

For a time series $y_t, t \geq 1$ the autocorrelation coefficient at lag k is:

$$\rho_k = \frac{cov(y_t, y_{t+k})}{var(y_t)} \tag{3.14}$$

The sample k^{th}

$$r_k = \frac{\sum_{t=1}^{n-k}(z_t - \dot{z}(z_{t+k} - \dot{z})}{\sum_{t=1}^{n}(z_t - \dot{z})^2} \tag{3.15}$$

order autocorrelation is,

y_t and y_{t+k}

$$y_{t+k}, y_{t+2}, \ldots, y_{t+k-1}$$

The theoretical partial autocorrelation at lag k, r_{kk} may be seen as the autocorrelation between, disconnected by a lag of k time intervals, with the effects of the intervening variables excluded.

The sample partial autocorrelation coefficients can be computed as:

$$r_{11} = r_1,$$

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1}(r_{k-1} j^r k - j)}{1 - \sum_{j=1}^{k-1}(r_{k-1,} j^r j)}, \quad k = 2, 3, \ldots \tag{3.16}$$

Theoretically, it can be shown that an Autoregressive (AR) process of order p has an autocorrelation function of infinite extent dominated by sine waves and damped exponentials, and a partial autocorrelation function that is zero after lag p. On the other hand, the partial autocorrelation function of a Moving Average (MA) process of any order q is infinite in extent and its autocorrelation function is zero beyond lag q. For ARMA processes, the identification of the process order gets somewhat complicated by the fact that both the autocorrelation function are infinite in extent.

Table 3.1: Distinguishing characteristics of theoretical PACF and ACF

Process	ACF	PACF
AR(p)	Trails off towards zero (exponential decay or damped sine wave)	Cuts off zero after lag p
MA(q)	Cuts off to zero after lag q	Trails off towards zero (exponential decay or damped sine wave)
ARMA(p, q)	Trails off towards zero (exponential decay or damped sine wave)	Trails off towards zero (exponential decay or damped sine wave)

These opposite characteristics are used to ascertain the type and order of AR, MA or ARMA processes in the data. In addition, other patterns may also be present in the ACF and PACF plots which can help to identify the true orders of AR and MA coefficients. In practice the flawless procedure of the significant spikes is confounded by a sampling error in the estimated PACF and ACF and the proper identification can become reasonably difficult depending on specific cases. Thus it requires some experience and judgment to identify a proper tentative form of the model.

Step 2- Estimation of the model parameters

After a tentative form of the model is identified, the AR and MA parameters need to be estimated in the best possible manner. Fundamentally, there are two ways of getting final estimates:

a. Trial and Error- This type of estimates examines many different values of parameters and choose that value (or values, if more than one parameter is to be estimated) that minimizes the sum of squared residuals of fitting the model. The residual at each time series observation and the model output value at the same time step

b. Iterative improvement – Choose a preliminary estimate obtained from the identification procedure (Yule-Walker equations) and use an efficient nonlinear least-squares algorithm, called Marquardt algorithm, to refine the estimate iteratively.

In solving for the MA and AR coefficient, several methods exist which include the use of nonlinear square estimation. These methods are: the Maximum Likelihood Method, Unconditional Least Squares Method and the Conditional Least Squares Method.

In the present work, Conditional least squares method was employed since is computationally faster and under the assumption of normally distributed random shocks in the model, the Least squares parameter estimates are either exactly equal

to or very nearly Maximum Likelihood estimates. In this work, the estimation of parameters was performed on Minitab and SPSS software package.

Step 3- Diagnostic Check of the model and residuals adequacy

It is a common practice found in ARIMA modelling to hesitantly fit more than one model form to the data by estimating the parameters for each model and then perform a diagnostic check to test each models' validity. According to the various statistical tests of fit the model which best fits is then selected for forecasting.

In particular, the following has to be performed:

a. A study of the residual series which is obtained after fitting the model to the data to see if any pattern remains accounted for. The PACF and ACF plots of the residual series help in detecting any unaccounted pattern.

b. A study of the sampling statistics of the current optimum solution to check if any further simplification of the model is possible. The residuals left over after fitting an ARIMA model should preferably be only random noise (white noise) with constant variance and zero mean.

The following statistical tests for lack of fit were used in the present work to check for the randomness of the residuals:

- PACF and ACF plots of residuals: The PACF plot of the residuals must show no significant points at any lag order. Similarly, the ACF of the residuals obtained after fitting a proper model to the data must show no significant autocorrelations at lag order. There is the demonstration of proper fitting in the absence of any significant spikes in the residual ACF and PACF plots. However, there are few spikes which are close to significance in practice. Even though the position of lag order also matters in deciding their importance and proper judgement used such spikes may not be a big concern.

- Ljung-Box Chi-Square Test: Another measure in checking for the randomness of residuals is the use of Ljung-Box Chi-Square test. The null hypothesis is white noise. The statistic measures the significance of the residual autocorrelations as a set and points out if they are collectively significant:

H_0 : The residual is random H_1 :
The residual is not random It is
computed as:

$$x_m^2 = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$$ (3.17) Where the sample size is n, r_k is the sample autocorrelation at lag k, and the m is the number of lags being tested. Each chi-square statistics is computed for all lags up to the indicated value and is not independent of the preceding chi-square values. If α is the significance level, then the null hypothesis is rejected if: $x_m^2 > x_{1-\alpha,f}^2$ the α quantile of the Chi-square distribution with f degrees of freedom. More tests need to be performed on the model itself in order to check its best fit and suitability in addition to the test in checking for residual randomness.

The most important criteria are the Akaike's Information Criteria (AIC) and Schwarz's Bayesian Criteria (SBC). The AIC and SBC are mostly used in comparing models fit which compete with each other to the same series. The statistically better fits are the models with the smaller AIC and SBC values.

- Akaike's Information Criteria (AIC)

  It is a statistical tool for model selection and is grounded in the concept of entropy. It can be non-statistically described as a measure of trade-off between the precision and complexity of the model. The absolute value of AIC is not useful; the relative comparison of AIC values of different competing models can be used to infer the best model. The model with lowest AIC value is the best fit. It is computed as:

$$AIC = -2\log likelihood + 2k \tag{3.18}$$

where k = the number for parameters in the model, $\log likelihood$ = maximum value of log likelihood function for the estimated model. Assuming the residuals to be normally and independently distributed, if the sum of square of the residual is denoted by R, the AIC criterion becomes:

$$AIC = n\left[\ln\frac{2\pi R}{n} + 1\right] + 2k, \quad R = \sum_{i=1}^{n} a_i^{-2} \tag{3.19}$$

With a minimum of parameters which are free, the AIC criterion tries to find out the best explained model. It imposes a punishment that is an increasing function of parameters which are estimated. This kind of penalty inhibits over fitting in estimation and thus leads to selection of a parsimonious model.

- Schwarz's Bayesian Criteria (SBC)

  It is also called Bayesian Information Criteria (BIC). SBC is also a statistical tool for model selection, which penalizes over fitting of estimation. The model with lower SBC is generally the best fit. It is computed as:

$$SBC = -2\log likelihood + k\ln(n) \tag{3.20}$$

For normally and independently distributed residuals,

$$SBC = n\left[\ln\frac{R}{n}\right] + k\ln(n) \tag{3.21}$$

The SBC criterion penalizes free parameters more heavily than the AIC criterion. It must be noted that both AIC and SBC tests may not generally point to a common model as the best fit. In such cases, proper judgment is required in choosing the best fit for the model. If, after performing the above checks of residual randomness and model adequacy, the model is found inadequate or if some significant autocorrelations are detected in the residual ACF plots, the identification stage should be revisited and a new model reformulated by

examining the ACF and PACF plots of the original series again and making a new interpretation. The knowledge of the leftover pattern in the residuals, as evidenced from the residual ACF and PACF series, may also be used in making a judgment to help identify a different form of tentative model. Thus, the three stages of the iterative process, that is identification, estimation and diagnostic checking may have to be repeated multiple times until a satisfactory model is generated. Another aspect to keep in mind while fitting models is the "principle of parsimony", which posits that models which are suitable for a given series is the very simplest model with least number of parameters which can be used to justify for the observed properties of the data. Thus, if two candidate models are finalized for the series under interest depending on the various tests outlined in this section and if they are comparable with respect to fitting adequacy and yielding white noise residuals, the model with minimum number of parameters must be preferred.

3.7 Forecasting Using ARIMA Models

Forecast can be generated using the models once a satisfactory and adequate model is fitted to the series of interest. There is the need to generate ARIMA model (3.4). $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$ (3.22)

The one-step ahead forecast for time $t + 1$ is given by, $y_{t+1} = \varphi_1 y_t + \varphi_2 y_{t-1} + \cdots + \varphi_p y_{t-p+1} + e_{t+1} - \theta_1 e_t - \theta_2 e_{t-1} - \cdots - \theta_q e_{t-q+1}$ (3.23)

Except e_{t+1}, the random shock at time $t+1$, all other parameters are known,

Thus setting $e_{t+1} = 0$, the one-step ahead forecasts generated is its true expected value. Correspondingly the future forecasts can be "bootstrapped" making use of these obtained forecasts y_t and setting the unrealized random shocks to 0 for each case. This therefore gives room for the 95% confidence interval for the forecasts to be calculated.

If future inputs are available, they are used in the forecasting model; if they are not available, the inputs can be forecasted using their corresponding prewhitening ARIMA filters prior to forecasting the outputs. As in the case of univariate ARIMA models, the goal is to fit a parsimonious model with minimum of parameters which satisfy the various goodness of fit criteria. The autocorrelation and partial autocorrelation plots are studied along with the Ljung Box Chi-square test results to test the hypothesis for randomness of the residual series e_t. AIC and SBC criteria are used to select the best fit out of a set of competing models.

3.8 Seasonal ARIMA Models

The pattern of a periodic series repeats every "s" time periods ($s > 1$), where "s" is also called the periodicity length. The ARIMA models for seasonal time series called SARIMA ("S" stands for seasonal), are built using the same three-stage iterative modelling procedure used for non-seasonal ARIMA models: identification,estimation, and diagnostic checking.

However, with seasonal data, attention must also be focused on the autocorrelation coefficients in the PACF and ACF plots occurring at seasonal lags $s,2s,3s,\ldots$ etc. If seasonal non-stationarity is present in the data, as evidenced from the fact that the autocorrelation coefficients at the seasonal lags of ACF plot will not die out rapidly,

proper order of seasonal differencing (denoted by "D") may be required to make the data seasonal stationary.

Secondly, the existence of seasonal autoregressive and moving average coefficients in the data needs to be determined on similar lines as was discussed for the non-seasonal ARIMA model identification, but with using the autocorrelation coefficients of PACF and ACF plots at the seasonal lags. The general notation for seasonal ARIMA model is ARMA (P, D, Q) where "P" is the order of seasonal autoregressive component, "Q" is the order of seasonal moving average coefficient and "D" is the order of seasonal differencing used. Generally, a time series often may contain both seasonal components and non-seasonal components.

Though the time series may be deseasonalized and a non-seasonal ARIMA model may be fitted to the remainder, experience suggests that Box-Jenkins methodology provides good forecasts of periodic data series. Thus, it may be advisable to leave the seasonal component in the data and fit a general class of ARIMA model which accounts for both seasonality and non-seasonality. Such a general ARIMA model can be represented by the form ARIMA (p, d, q). This is commonly referred to as a seasonal ARIMA multiplicative model and it is represented by:

$$\Phi(B^2)\Phi(B)(1-B^2)^D(1-B)^d y_t = \Theta(B^s)\Theta(B)e_t$$

where D represents the number of differencing. The forecasts from seasonal ARIMA multiplicative models are generated in the similar fashion as with nonseasonal ARIMA models.

3.9 Data Analysis Tool

The RStudio was utilized to examine the database of respondents and conduct statistical analysis of the trends. The research was examined through the respondents from the charts and figures of the trend of the time series analysis.

3.10 Sampling Techniques

This is the operation or methods of choosing a suitable sample size of a population in order to determine the features or parameters of the population wholly (Adinyira et al. (2013)). The sample must have sufficient size to merit statistical analysis. As a result of the unusually large population sizes used in research, the researcher cannot test every individual in the population just like in a census as a result of the expensiveness and waste of time associated with such method. To address this issue, researchers always depends on sampling techniques (Saunders et al. (2009))

CHAPTER 4

Data Analysis and Results

This chapter presents the analyses and results of the data collected with respect to the research objectives.

4.1 Descriptive Statistics on Malaria Cases in the Berekum Municipality

Table 4.1: Summary of the Malaria cases during the period of Jan 2011 to Dec 2015

	Min. Value	1st Quartile	Median	Mean	3rd Quartile	Max. Value	Totals
OPD	201	859	1245	1256	1736	2075	75361
In-Patients	23	54	80	83	106	106	4996
No. of Deaths	0	0	0	1	1	4	48

Source: Berekum Holy Family Hospital Survey, 2016

Table 4.1 demonstrates the summary of malaria cases from the Statistical Department at the Holy Family Hospital at Berekum Municipal, a total of 75361 out patient's malaria cases were recorded from January 2011 to December 2015. Between this period a total of 4996 patients were admitted in the hospital with 48 deaths recorded during that period. From the data, it was observed that, the hospital could record as high as 2075 out-patient malaria cases within a month and a minimum of 201 cases within a month. It was also observed that an average of 1256 out-patient malaria cases was recorded at the hospital every month. Out of the 75361 cases that were recorded during this period, about 4996 patients were admitted in the hospital. The data showed an average of 83 patients being admitted in the hospital every month. The records showed that, the hospital could admit as high as 106 patients in a month and a minimum of 11 patients were admitted in a month

during the period. A total of 48 deaths were recorded within this period with an average of 1 patient dying every month. The maximum number of deaths recorded within this period in a month was 4, whereas there were months without any death.

Table 4.2: Summary of OPD Malaria Cases

	Min. Value	1st Quartile	Median	Mean	3rd Quartile	Max. Value	Totals
<5 yrs	38	151	263	263	376	475	15780
> 5yrs	165	703	982	971	1321	1626	58237
Pregnant	0	6	16	22	34	85	1344

Source: Berekum Holy Family Hospital Survey, 2016

The Table 4.2 illustrates the summary of the out-patient malaria cases that were recorded at the statistical department of the Holy Family Hospital of Berekum Municipality from January 2011 to December 2015. The data was put into three categories; patients less than 5 years (< 5yrs), patients who are above 5 years (> 5yrs) and pregnant women. A total of about 15780 patients were less than 5 years old, 58237 patients were above 5 years old and 1344 pregnant women were recorded during this period. It was observed that an average of 263 patients were recorded for patients under 5 years old with a minimum of 38 patients in a month and a maximum of 475 patients were recorded in a month during this period. For patients over 5 years, the records showed that an average of 971 patients were recorded every month with a minimum and maximum number of patients recorded within a month during this period were 165 and 1626 respectively. An average of 22 pregnant women was recorded every month during this period and a maximum of 85 patients while no pregnant patients were recorded in some months during this period.

Table 4.3: Demonstrating the Summary of In-Patient Malaria Cases

	Min. Value	1st Quartile	Median	Mean	3rd Quartile	Max. Value	Totals
<5yrs	7	31	43	44	58	94	2657
>5yrs	4	17	29	33	45	94	2003
Pregnant	0	2	4	6	8	24	336

Source: Berekum Holy Family Hospital Survey, 2016

Table 4.3 shows a total of 4996 in-patients who were recorded at the hospital during the period under discussion with 2657 patients being under 5 years old, 2003 patients were over 5 years old while 336 patients were pregnant women. An average of 44 patients under 5 years were admitted at the hospital every month and 33 patients over 5 years were also admitted every month while 6 pregnant women were admitted every month during the period. This meant that, 27.3% of all out-patient malaria pregnant women were admitted every month, 16.7% of all OPD malaria patients under 5 years were admitted and 3.4% of patients over 5 years were admitted every month. On the average, 6.6% of all out-patient malaria patients were admitted.

Table 4.4: Summary of Death of Malaria Patients Recorded during the Period

	Min. Value	1st Quartile	Median	Mean	3rd Quartile	Max. Value	Totals
< 5yrs	0	0	1	1	1	4	32
> 5yrs	0	0	0	0	0	3	16

Source: Berekum Holy Family Hospital Survey, 2016

A total of 48 deaths were recorded at the hospital during this period. On the average no death was recorded for patients over 5 years every month at the hospital while 1 death for patients under 5 years per month was recorded. The records given did not include death of pregnant women. The maximum number of deaths recorded in a month during this period for patients under 5 years and over 5 years were 4 and 3 respectively.

4.2 Hypothesis Testing

The average malaria patients that visit the Holy Family Hospital of Berekum Municipality is normally distributed with a standard deviation of 528.07. A total of 80,357 patients were recorded during the period of January 2011 to December 2015, which gave an average of 1,339.28 patients per month. On the basis of these data, can one conclude that the average malaria cases is less than 1,500 patients per

month? (Take $\alpha = 0.01$) The parameter of interest in this study is the average number of malaria patients that visit the hospital every month. The standard error of estimate of the mean ($SE_{\bar{x}}$) was calculated using equation

(4.1);

$$SE_{\bar{x}} = \frac{SD_x}{\sqrt{n}} \tag{4.1}$$

where SD_x is the standard deviation of the population and n is the total number of data used to compute the sample mean.

The significance level chosen for our hypothesis test is 99%.

The null hypothesis: $H_0 : \mu = 1500$ Alternate hypothesis: $H_1 : \mu < 1500$

The test statistics is the quantity,

$$Z_{cal} = \frac{\bar{x} - \mu_o}{SE_{\bar{x}}}$$

Where $\bar{x} = 1339.28$ is sample mean, $\mu_o = 1500$ is population mean, $sd = 528.07$ is standard deviation and $n = 60$

$$Z_{cal} = \frac{1339.28 - 1500}{528.07/\sqrt{60}}$$

$$Z_{cal} = -2.36$$

The critical value was computed as $Z_{0.01} = -2.33$. Since $Z_{cal} < Z_{0.01}$, there is enough evidence to support the alternate hypothesis that the average number of malaria patients is less than 1,500, with a standard error of 68.17.

4.3 The Box Jenkins Modeling

The Box-Jenkins process involves identifying and plotting a time series data and transforming a non-stationary data into a stationary data. The dependence order of the model (ARIMA) is identified by examining the autocorrelation and the partial autocorrelation of the stationary sets. The parameter is estimated followed by diagnostic and residual analysis.

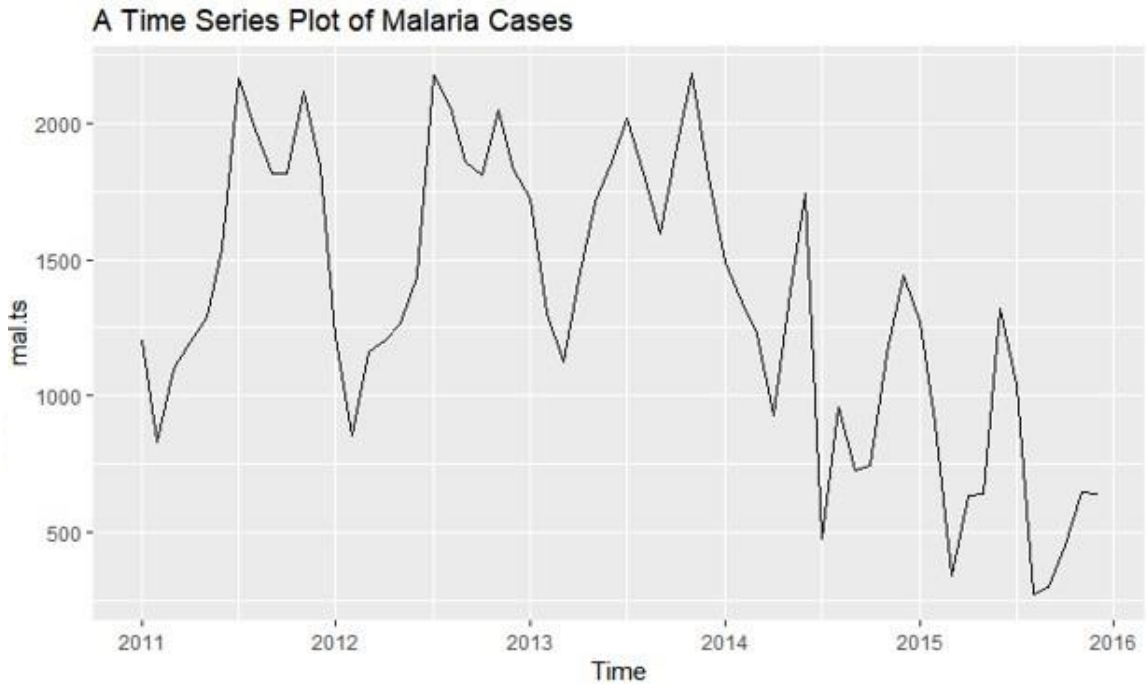


Figure 4.1: Time Series Plot of Malaria Cases

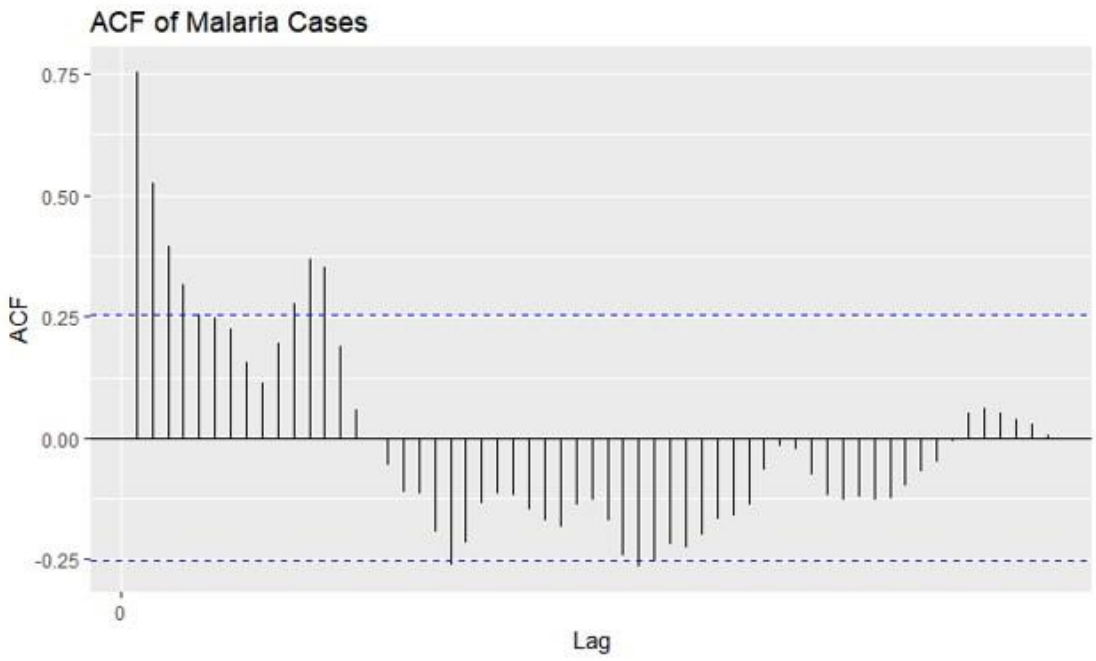


Figure 4.2: ACF of Original Malaria Cases

The ACF shows that the series is not stationary hence there is the need to difference the process to attain stationarity.

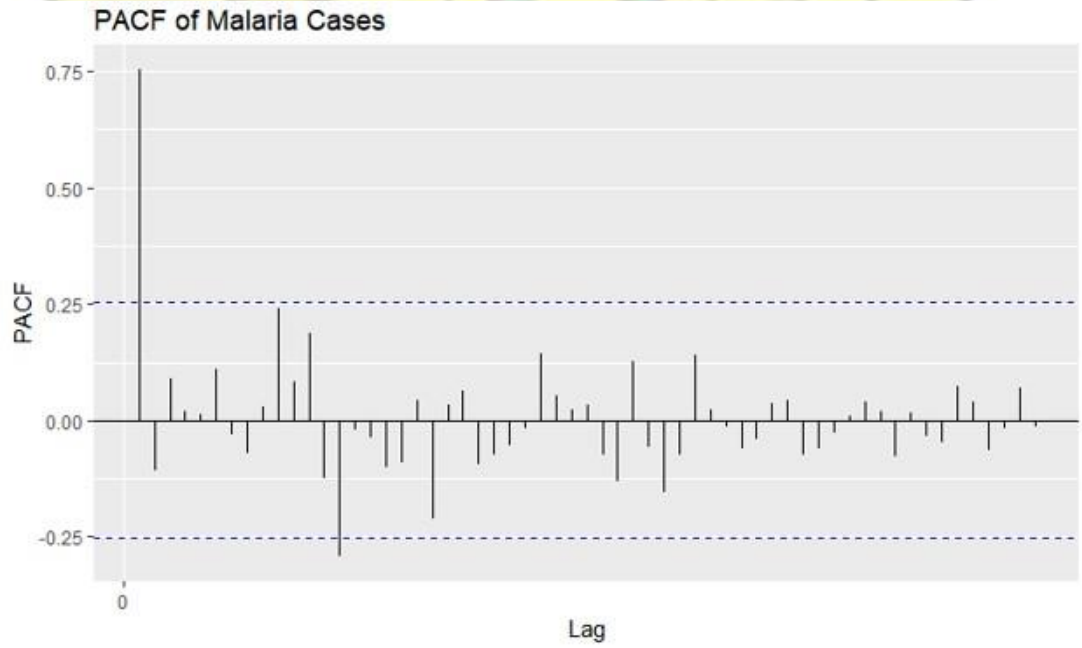


Figure 4.3: PACF of the Original Malaria Cases

The 95% boundary captures almost all the PACF values except lag 14 which falls outside the boundary due to the 5% chance. Hence there is no autoregressive process in this series.

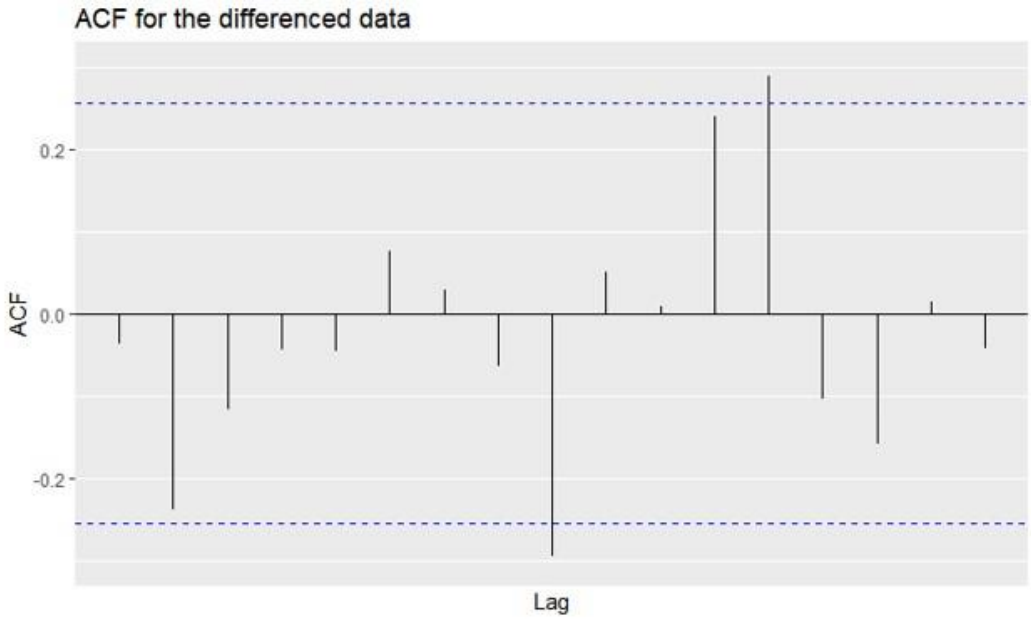


Figure 4.4: ACF for the Difference Data

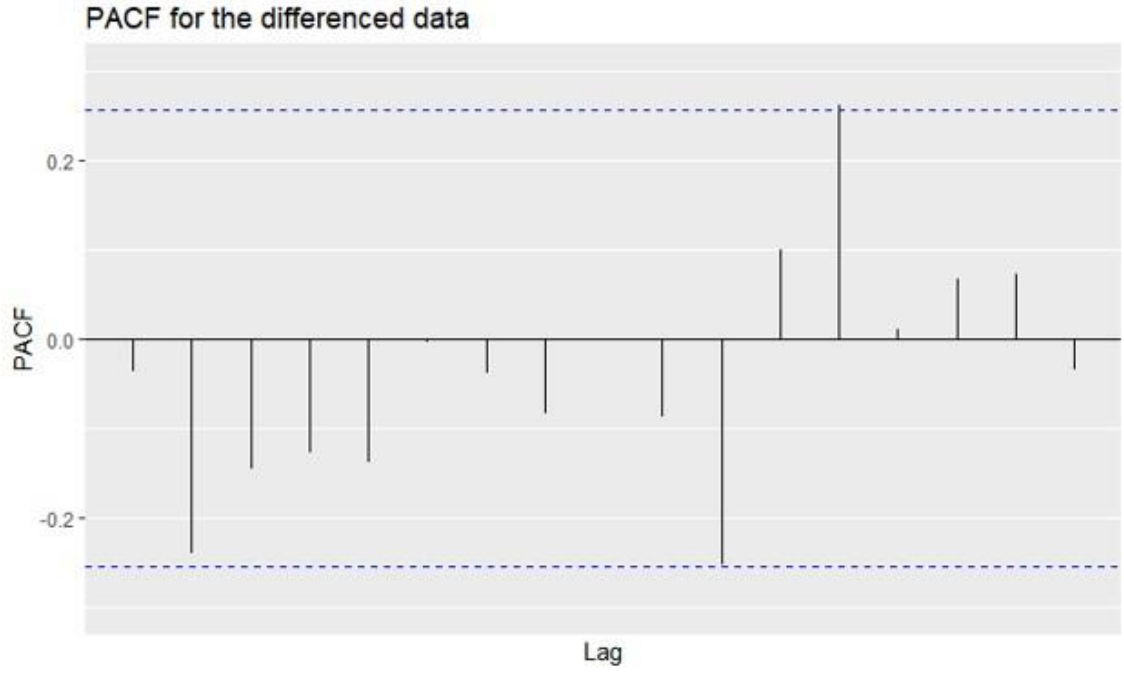


Figure 4.5: PACF of the Difference Data

4.4 Hypotheses testing for the malaria cases

(Stationarity Tests)

4.4.1 Stationarity Test for the original malaria cases

Null hypothesis (H_0): The process is not stationary.

Alternative hypothesis (H_1): The process is stationary

Table 4.5: Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value
-3.2242	3	0.09217

Since the p-value is greater than 5% we fail to reject the null hypothesis and conclude that the process is not stationary.

4.4.2 Stationarity Test for the differenced malaria cases

Null hypothesis (H_0): The process is not stationary.

Alternative hypothesis (H_1): The process is stationary

Table 4.6: Augmented Dickey - Fuller Test

Dickey-Fuller	Lag order	p-value
-5.0871	3	0.01

Since the p-value is less than 5% we reject the null hypothesis and conclude that the process is stationary.

Table 4.7: Model checking for Smaller AIC

Model	AIC	Log-likelihood	Standard Error
ARIMA(0,1,1)	865.16	-430.58	127772
ARIMA(0,1,2)	861.50	-427.75	115337
ARIMA(0,1,4)	863.50	-426.75	111285
ARIMA(0,1,7)	869.01	-426.50	109474
ARIMA(0,1,1)(1,0,0)[12]	858.87	-426.43	104569

ARIMA(0,1,2)(1,0,0)[12]	855.21	-423.60	96223
ARIMA(0,1,4)(1,0,0)[12]	858.93	-423.46	95755
ARIMA(0,1,2)(1,0,2)[12]	857.88	-422.94	87430
ARIMA(0,1,2)(1,0,1)[12]	857.18	-423.59	96116
ARIMA(0,1,2)(2,0,2)[12]	859.86	-422.93	86153

The table above gives some models of the *ARIMA(p,d,q) (P,D,Q)*$_S$ for the process. The interest here is to find the least AIC value, which will aid in fitting the model. From the table above, the least AIC value is 855.21 corresponding to *ARIMA(0,1,2) (1,0,0)*$_{12}$ process. Hence, this model will be used to for the forecasting.

4.5 Fitting the ARIMA Model

4.5.1 Test For Significance Of Model Estimates

Table 4.8: T-test table

Variable	Estimate	Standard Error	t.value	p.value
ma1	-0.3249	0.1272	-2.5550	0.0134
ma2	-0.3734	0.1272	-2.9351	0.0049
sar1	0.4146	0.1315	3.1519	0.0026
constant	-14.8988	19.2083	-0.7756	0.4413

The estimated parameters (ma1, ma2, and sar1) are all significant at the 5% level of significance. The model also gave a smaller AIC (855.21). The smaller the AIC the better for the model predicted.

4.5.2 Model Equation

The Model for *ARIMA*$(0,1,2)(1,0,0)_{12}$ is given by:

$$Y_t = \mu + Y_{t-1} + \phi_1^S Y_{t-12} - \phi_1^S Y_{t-13} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t \qquad (4.2)$$

Where ϕ_1^S is the parameter for the seasonal autoregressive process

$$e_t \sim N(0, \sigma_e^2)$$

μ is the mean of the process

$\Theta = (\theta_1, \theta_2)$ are the parameters of the white noise.

Fitted model

$$Y_t = -14.8988 + 0.3249e_{t-1} + 0.3734e_{t-2} \quad (4.3)$$

4.6 Forecast

Forecasted values for 2016 and 2017 are in Table 4.7 below:

Table 4.9: Forecasted/ Predicted Values for 2 years

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEPT	OCT	NOV	DEC
2016	617	521	289	412	414	697	581	260	272	335	418	413
2017	404	364	267	318	319	437	389	255	260	286	321	319

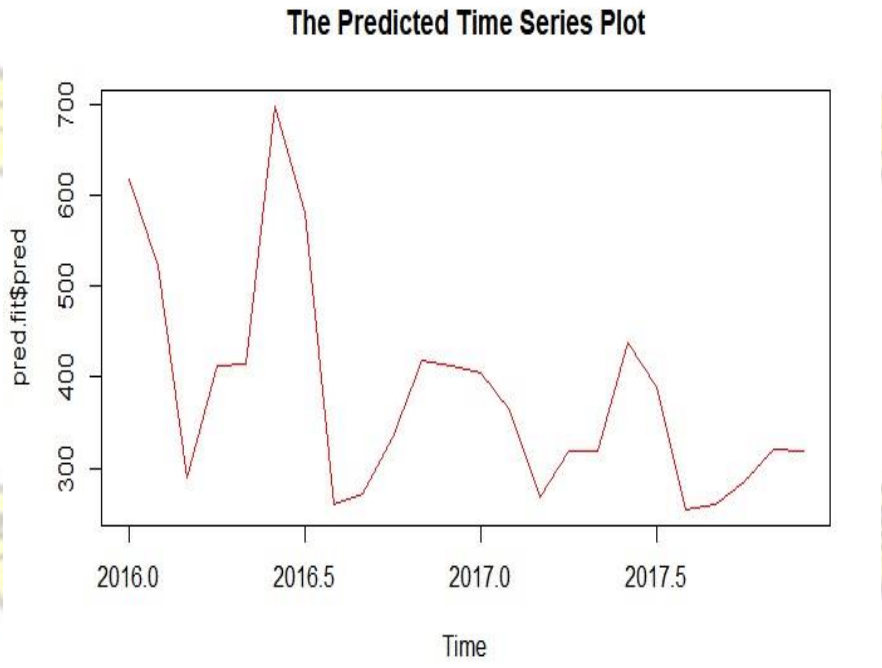


Figure 4.6: Predicted Time Series Plot for 2016 and 2017

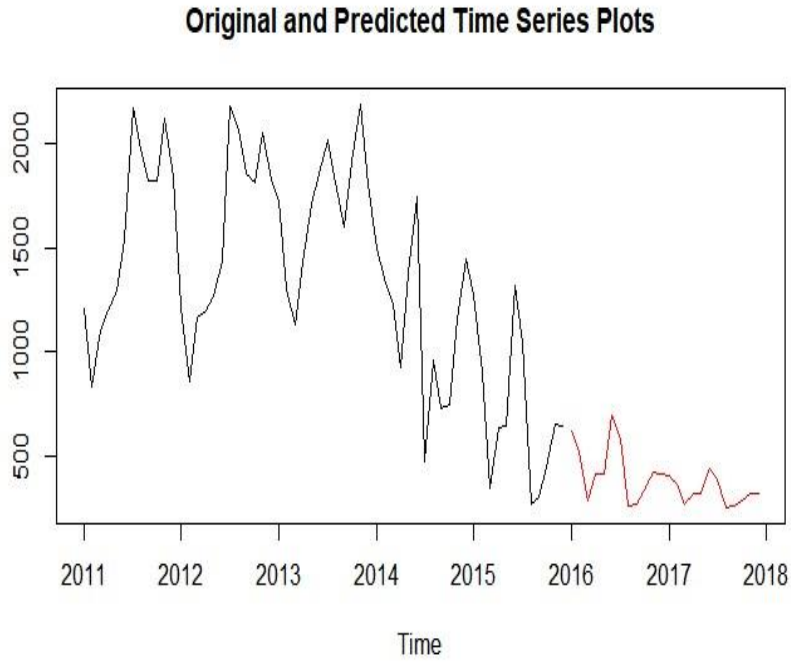
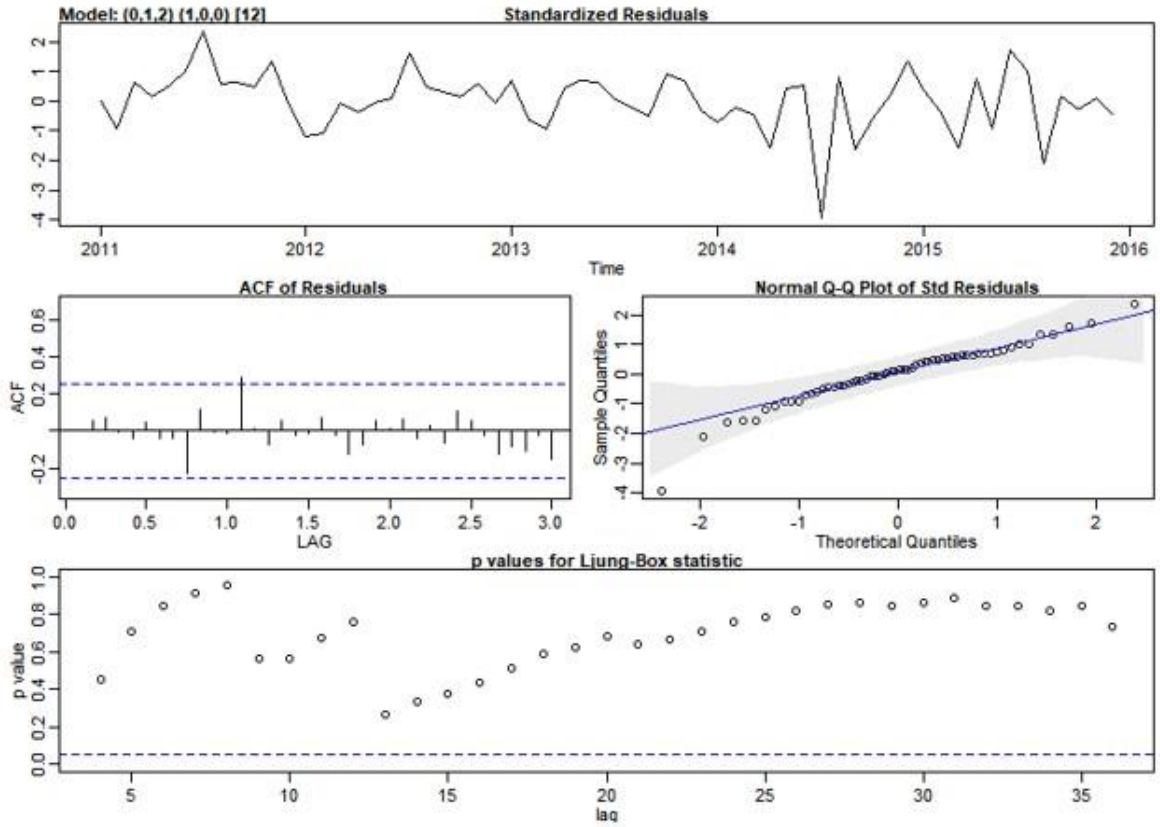


Figure 4.7: A Time Series plot for both Original and Predicted malaria cases.

Data collected during the study showed three trends from 2011 to 2017. From 2011-2013, a similar trend was seen where two peaks were recorded in a year (the middle and towards the end of the year) being the highest cases of malaria recorded. It further declined from 2014 to 2015 and continued for the predicted time series cases from 2016 to 2017.

4.7 Residual Analysis

The corresponding residual analysis is given below



Equal variance assumption (homoskedasticity)

The ACF of Residuals indicates that all the ACFs lies in the 95% confidence interval whereas the ACF that lies outside is due to chance. Hence the assumption of equal variance is valid.

Normality assumption

The Normal Q-Q plot of the standardized residuals indicates that the normality assumptions for the residual is valid. Almost all the data points lie on the Q-Q line in the plot above.

CHAPTER 5

Summary of Findings, Recommendations and Conclusion

5.1 Introduction

The research aims at identify critical success factors of variation management relating to project team characteristics. In an attempt to achieve the purpose of the study, two objectives were established.

This chapter consists of the summary of the issues that have been discussed in the preceding four chapters of the study. It also brings into the lime light the review of the study's recommendations based on the findings of the study. The chapter finally ends on concluding with the recommendation for future research studies.

5.2 Summary of Findings

The research study targets at probing the time series of Malaria Cases at the Holy Family Hospital in the Berekum Municipality and how it affects human life and economic development in general. The data were categorized into three areas. These are; patients less than 5years (< 5yrs), patients above 5 years (> 5yrs) and pregnant women. In a stab to answer the research objectives, there is the need to present the details of the salient points so as to acclaim appropriate measures to alleviate the menace associated with this problem. The summary of the findings is outline as follows:

5.2.1 Out-Patients DepartmentMalaria cases from January 2012 to December 2015

The number of Out-Patients Malaria Cases recorded at the Holy Family Hospital in the Berekum Municipality were 75,361. A total of 4,996 patients were admitted within this five-year period. It was observed that the Hospital recorded as high as 2,075 and as low as 201 out-patient Malaria cases within a month. An average of 1,256 out-patient malaria cases were recorded and an average of 83 patient being admitted in the hospital every month. The hospital admits a maximum of 106 patients and a

minimum of 11 patients in a month. The number of deaths were 48, having an average of 1 death monthly.

The number of out-patients less than 5 years old are about 15,780 patients, patients above 5 years are 58,237 and pregnant women are 1,344. Patients under 5 recorded 263 patients with a minimum of 38 patients and a maximum of 475 patients.

5.2.2 Demonstrating the In-patient Malaria Cases

The period saw a recording of 4,996 in-patients with 2,657 patients under 5 years old, patients who are over 5 years old were 2,003 with 336 patients being pregnant women. An average of 44 patients under 5 years old get admitted every month and 33 patients over 5 years were admitted with 6 pregnant women admitted every month.

5.2.3 Death of Malaria Patients during the Period

The number of death cases for the period stood at 48 patients. There was no record of death for patients over 5 years every month while a death for patients under 5 years per month was recorded.

5.3 Conclusion

The data on the malaria cases which were to be stationary was unstable. It was made stationary by differencing to remove the trend from the mean of the nonstationary malaria data acquired. It is however expedient that the Holy Family Hospital in the Berekum Municipality should expect a decline in the of cases of malaria.

5.4 Recommendations

The rudimentary intention of this research study is to cautiously explore a quantitative analysis of the impact of times series of Malaria cases by bridging the

information gap between technocrats point of view as well as the politicians acuity. In view of the findings of this research, the following are therefore recommended per the analysis. These are:

- There should be a modelling applicable to the modelling of malaria prevalence in different contexts. This is due to the fact that some context may be a geographical location with similar or different climatic factors; and
- There should be the need to relax the assumptions of constant prone population size and also inculcating the effect of interventions for Malaria control in the model.

REFERENCES

Abd, R. (2007). Rahman. *Yudariah Mohammad Yusof., & Mason, JH*.

Adinyira, E., Ahadzie, D., and Kwofie, T. (2013). Determining the unique features of mass housing projects (mhps). In *Proceedings of the 5th West Africa Built Environment Research (WABER) Conference*. Accra, Ghana.

Akpalu, W. and Codjoe, S. N. A. (2013). Economic analysis of climate variability impact on malaria prevalence: The case of ghana. *Sustainability*, 5(10):4362– 4378.

Ankamah, S., Nokoe, K. S., and Iddrisu, W. A. (2018). Modelling trends of climatic variability and malaria in ghana using vector autoregression. *Malaria Research and Treatment*, 2018.

Artavanis-Tsakonas, K., Tongren, J., and Riley, E. (2003). The war between the malaria parasite and the immune system: immunity, immunoregulation and immunopathology. *Clinical & Experimental Immunology*, 133(2):145–152.

Asenso-Okyere, K., Asante, F. A., Tarekegn, J., and Andam, K. S. (2009). *The linkages between agriculture and malaria: Issues for policy, research, and capacity strengthening*. Intl Food Policy Res Inst.

Cao, T., Pintér, E., Al-Rashed, S., Gerard, N., Hoult, J. R., and Brain, S. D. (2000). Neurokinin-1 receptor agonists are involved in mediating neutrophil accumulation in the inflamed, but not normal, cutaneous microvasculature: an in vivo study using neurokinin-1 receptor knockout mice. *The Journal of Immunology*, 164(10):5424–5429.

Carrington, A. (2001). Malaria: its human impact, challenges, and control strategies in nigeria. *Harvard Health Public Rev*, 2:54–60.

Dadzie, S., Boakye, D., Asoala, V., Koram, K., Kiszewski, A., and Appawu, M. (2013). A community-wide study of malaria reduction: evaluating efficacy and user-acceptance of a low-cost repellent in northern ghana. *The American journal of tropical medicine and hygiene*, 88(2):309–314.

Dan, E. D., Jude, O., and Idochi, O. (2014). Modelling and forecasting malaria mortality rate using sarima models (a case study of aboh mbaise general hospital, imo state nigeria). *Sci. J. Appl. Math. Stat*, 2:31–41.

Fillatre, L., Marakov, D., and Vaton, S. (2003). Forecasting seasonal traffic flows. *Computer Science Department, ENST Bretagne, Brest, Paris*.

Freckleton, R. (2009). The seven deadly sins of comparative analysis. *Journal of evolutionary biology*, 22(7):1367–1375.

Gilles, H. M. and Warrell, D. A. (2002). *Essential malariology*. Arnold.

Guinovart, C., Navia, M., Tanner, M., and Alonso, P. (2006). Malaria: burden of disease. *Current molecular medicine*, 6(2):137–140.

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., et al. (2005). A microrna polycistron as a potential human oncogene. *nature*, 435(7043):828.

Hoek, E., Carranza-Torres, C., and Corkum, B. (2002). Hoek-brown failure criterion-2002 edition. *Proceedings of NARMS-Tac*, 1:267–273.

Klein, E. Y., Smith, D. L., Laxminarayan, R., and Levin, S. (2012). Superinfection and the evolution of resistance to antimalarial drugs. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20121064.

Kristan, M., Abeku, T. A., Beard, J., Okia, M., Rapuoda, B., Sang, J., and Cox, J. (2008). Variations in entomological indices in relation to weather patterns and malaria incidence in east african highlands: implications for epidemic prevention and control. *Malaria journal*, 7(1):231.

Lin, H.-S., Yue, B.-D., and Ho, P. C. (2009). Determination of pterostilbene in rat plasma by a simple hplc-uv method and its application in pre-clinical pharmacokinetic study. *Biomedical Chromatography*, 23(12):1308–1315.

Manda, L. (2017). *Plasmodium Species Infections In Children Suspected Of Malaria At Yendi Municipal Hospital, Northern Region*. PhD thesis, University of Ghana.

McKenzie, F. E. and Bossert, W. H. (2005). An integrated model of plasmodium falciparum dynamics. *Journal of theoretical biology*, 232(3):411–426.

McKenzie, F. E., Smith, D. L., O'Meara, W. P., and Riley, E. M. (2008). Strain theory of malaria: the first 50 years. *Advances in parasitology*, 66:1–46.

Oduro, F. T., Harvim, P., Saviour, A. W., Borkor, R., and Francois, M. (2015). An epidemiological model of malaria transmission in ghana. *The International Journal of Science and Technoledge*, 3(1):171.

Okosun, K. O., Ouifki, R., and Marcus, N. (2011). Optimal control analysis of a malaria disease transmission model that includes treatment and vaccination with waning immunity. *Biosystems*, 106(2-3):136–145.

Organization, M. W. H. (2014). Global status report on alcohol and health, 2014.

Organization, W. H. (2015). *World malaria report 2014*. World Health

Organization.

Organization, W. H. et al. (2012a). Trends in maternal mortality: 1990 to 2010: Who, unicef, unfpa and the world bank estimates. *Trends in maternal mortality: 1990 to 2010: WHO, UNICEF, UNFPA and The World Bank estimates.*

Organization, W. H., Unicef, et al. (2012b). Trends in maternal mortality: 1990 to 2010: Who, unicef, unfpa and the world bank estimates.

Orkoh, E. and Annim, S. K. (2017). Source and use of insecticide treated net and malaria prevalence in ghana.

Ostovar, A., Haghdoost, A. A., Rahimiforoushani, A., Raeisi, A., and Majdzadeh, R. (2016). Time series analysis of meteorological factors influencing malaria in south eastern iran. *Journal of arthropod-borne diseases*, 10(2):222.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680.

Saunders, M., Lewis, P., and Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Takyi Appiah, S., Otoo, H., and Nabubie, I. (2015). Times series analysis of malaria cases in ejisu-juaben municipality. *Int. J. Sci. Technol. Res*, 4:220– 226.

Tio, T. L., Coene, P., Van Delden, O., and Tytgat, G. (1991). Colorectal carcinoma: preoperative tnm classification with endosonography. *Radiology*, 179(1):165–170.

Wikipedia contributors (2019). Malaria — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Malaria&oldid= 876921866. [Online; accessed 19-January-2019].

Worley, H. (2006). Best buys and challenges for malaria prevention in sub-saharan africa.

Worrall, E., Basu, S., and Hanson, K. (2005). Is malaria a disease of poverty? a review of the literature. *Tropical Medicine & International Health*, 10(10):1047– 1059.

Zinszer, K., Tamblyn, R., Bates, D. W., and Buckeridge, D. L. (2013). A qualitative study of health information technology in the canadian public health system. *BMC public health*, 13(1):509.