



Special Issue of First International Conference on Innovations in Engineering Sciences (ICIES 2020)

## Incidents of Malaria in India using ARIMA Models

K.Sukanya<sup>1</sup>, Dr.T.Suneetha<sup>2</sup>, T.Sukeerthi<sup>3</sup>, A.Vani<sup>4</sup>

<sup>1,3,4</sup>Department of Statistics, Academic Consultant, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.

<sup>2</sup>Department of Statistics, Assistant Professor, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.

sukanya.sudha27@gmail.com<sup>1</sup>, thiyyasuneetha@gmail.com<sup>2</sup>

### Abstract

*In India, the best embody of malaria occurred within the year 1950's with associate calculable 75 million cases and 0.8 million deaths per Annum (World Health Organization, country office for India). The model was used for the forecasting of the year wise incidence of Malaria whereas Auto regressive integrated moving average (ARIMA) models was used for forecasting for the years 2020 and 2022 in Republic of India (Bharat) our study provides that of the ARIMA model was designated as best suited model to predict the longer term incidents of malaria cases within the fourth approaching period in India.*

**Keywords:** Time Series, ARIMA models, Malaria

### 1.Introduction

The time series modeling parameters the relationship between the ascertained malaria from the indices past observations, while not mistreatment the other variables. Malaria is caused by the plasmodium. That is transmitted between humans by genus anopheles mosquitoes. These malaria victims are in the poorest and generally most remote parts of the world, increase the issue in findings support to cope with the disease. Curable and preventable with medication medical care, but a vaccine is not out there. In keeping with the World Health Organization in 2012, there were just about 207 million cases of malaria leading to 627,000 deaths (World Health Organization, 2004). The overwhelming majority is 90% of these cases happen in Africa (medical research council, 2001) most of the deaths occur in youngsters. However, the speed of deaths in youngsters has been reduced by 54% since 2000 (World Health Organization, 2014). Malaria is so a huge downside that player all phases that the society. Malaria remains a serious health challenge to humanity over the world.

### 1.1 Malaria in India

As the second hottest country within the world, with a population over and above one million people, India's public health system faces several challenges as well as implementation of surveillance programs to accurately estimate and management the national malaria burden. Traditionally, the best embody of malaria occurred in India occurred within the 1950's. With associate calculable 75 million cases and 0.8 Million deaths per Annum (World Health Organization, country workplace for India). The lunch of national malaria management program in 1953 resulted during a vital decline within the variety of reportable cases to <50,000 and no reportable morality by 1961. Despite it is close to elimination within the mid- 1960's malaria regarded to 6.45 million cases in 1976. Since then, confirmed cases have bit by bit cut to 1.6 million cases and 1,100 deaths in 2009. Recently it has been prompt that the malaria indecencies in between 9 to 50 times bigger than continual with a 13 fold below estimation of malaria connected morality. Such claims rail force the necessity for strong and

comprehensive epidemiological surveillance studies across the country.[1-4]

### 1.2. Data Collection

This study was carried by using secondary information of malaria cases and death across very different states of India. 23 years (1997- 2017) malaria cases information was obtained from sequent annual reports of national vector borne disease management program (NVBDCP), ministry of health and family welfare, government of India and the states was obtained from national commission on population, ministry of health and family welfare (5-9).

### 1.3. Objective

The aim of this project is to seek out about using ARIMA models on incidents of malaria in India.

- To value forecast accuracy additionally on compare among very different ARIMAfitted to a time series.
- To notice the simplest ARIMA model is used to forecast the incidents of malaria in India

## 2.Review of Literature

### 2.1. Introduction

Malaria could be a dangerous disease. It has generally transmitted through the bite of associate infected “Anopheles mosquito” infected mosquitoes carry the “Plasmodium parasite”. Once this mosquito bites you, the parasite is relaxed into your blood streams.

Once the parasites are entered into your body, they travel to measure wherever they matured. As they many days mature parasites entered the blood streams and begin to infect the used bloodcell.

In the United States, the centers for disease management prevent report 1, 700 cases of malaria annually.

Malaria is a mosquito- based infectious disease of humans associated different animals caused by protests ( a variety of microorganisms) of the genus plasmodium it begins with a bite from an insect feminine mosquito, once this bites you, the parasite is relaxed into your blood stream that in severe cases can progress two coma or death. Malaria is widespread in tropical sub-tropical regions in a broad bond around the equator , including much of sub Saharan, Africa America and Asia.

“The term Malaria originates from medieval Italian ; Mala aria - bad air“ The disease was formally referred to as Ague or marsh fever, due to its association with swamps and marshal Marshland Malaria was once common in most of the Europe and north America wherever it is no longer endemic, through imported cases do occur. Within 48, 72 hours, the parasites within the red blood cells multiply, inflicting the infected cells, ensuing in symptoms that ocean aim cycles that last 2-3 days at a time. Malaria generally found in tropical and sub-tropical climates were the parasites will live. The World Health Organization states that, in 2016, there were associate calculable 216 million cases of malaria in 91 countries. In the United states the centers for disease management prevention report 1,700 cases of malaria annually.

### 2.2. Causes of Malaria

Malaria will occur if a mosquito infected with the plasmodium parasite bites, you they are five kinds of malaria parasites infect humans

- Plasmodium Vivax
- Plasmodium ovule
- Plasmodium malaria
- Plasmodium falciparum
- Plasmodium Knowlesi

Plasmodium falciparum causes an additional severe type of the disease and those who contract this type of malaria havea higher risk of death. An infected mother will conjointly be transmitted through

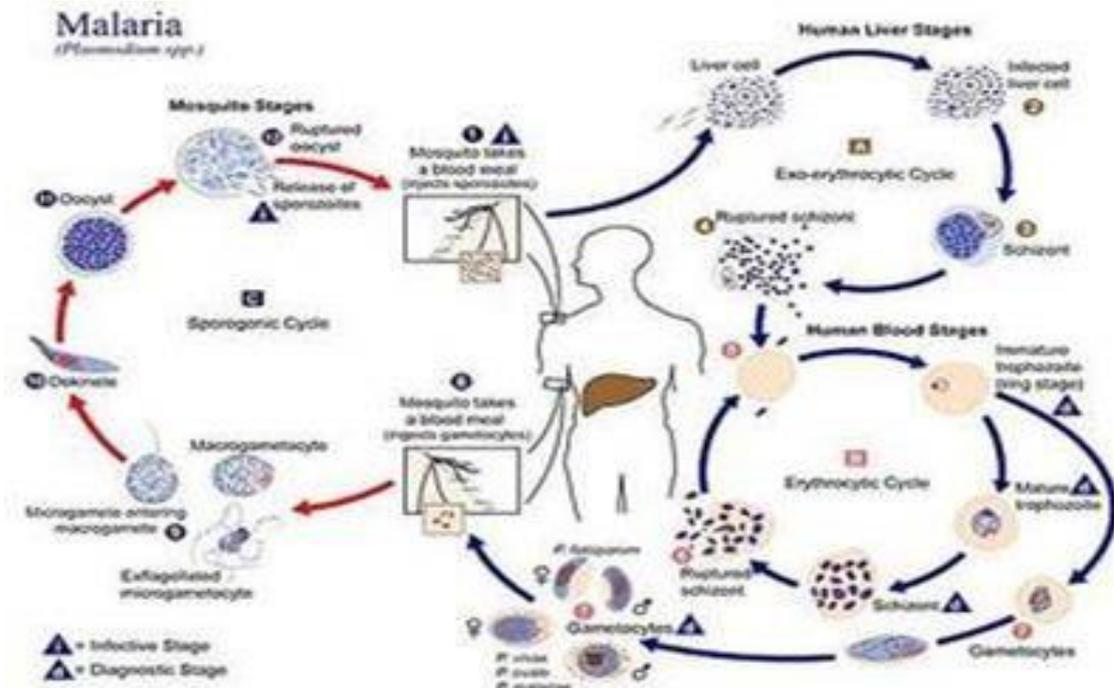
- An organ transplant
- A transfusion
- Use of shared needles or syringes

### 2.3.Symptoms of Malaria

Once bitten by an infect mosquito it could take 1- 4 weeks for the disease to show. In some rare cases, the parasite could keep dormant space for as long has a year and may recur over a period of symptoms of malaria . High Fever, Headache, Muscle Pain, Severe Chillis, serious Sweating, Fatigue, Dry Cough, Loose Motions.

### 2.4.Symptoms of Malaria

Life cycle of plasmodium is defined in both mosquito and humans begins as follows.



**Diagram 1: Life Cycle of Plasmodium**

**2.5. Symptoms of Malaria**

- 1945 - Insecticide properties of DDT known.
- 1953 - Calculated malaria cases in India-75 million calculated deaths.
- 1953 - Launching of national malaria management program.
- 1958 - 90% of malaria - launching of national educational malaria eradication program
- 1966 - Cases reduced to 0.1 million early 70 s - revitalization of malaria
- 1997 - Modified discovered of operations enforced.
- 1984 - International Bank power-assisted increased malaria management project.
- 1997 - Revivification of malaria in some states.
- 2005 - International fund power-assisted intense malaria management project.
- 2006 - ACT introduced in areas showing antimalarial resistant falciparum malaria
- 2008 - ACT introduced to high risk covering regarding.

**2.6. Treatment of Malaria**

- Completetreatmentideallyonce combination of identification.
- Use of artemisinin based mostly combination medical aid for treatment of falciparum.

- Constitution of country level task force for systematic drug resistance studies and observation
- Creation of network of researchers and establishments operating on observation of drug resistance of existingantimalarial
- Monitoring of resistance to chloroquine in this might be viewed as an emergency and steps might even be taken to prolong the use of chloroquine in the treatment of malaria.[7-10]

Chemoprophylaxis policy for the tourists should be developed. Specific recommendation for utterly different endemic states of traveller attraction might even be developed and hosted on the website of the national program, policy might be updated every 2 years and review of malaria scenario in the country. World malaria day was established in 2017 by the 60 session of the world health assembly. The day was established to offer “education and understanding of malaria” and unfold data of malaria hindrance and treatment in endemic areas.

**2.7 Malaria Control Strategies in India**

Directorate of national vector borne disease control programme (NVBDCP) is the central nodal agency for the hindrance and management of vector borne diseases together with malaria and alternative VBDs (dengue, lymphatic fulariasis, kalaazar, Japanese encephalitis and chikungunya) in India.

The national strategic set up (NSP) has been developed by NVBDCP. Ministry of health and family welfare, government of Republic of India with the support of WHO to produce a road map for creating India malaria free by 2027. Malaria control strategies consist of:

a. Early case detection and prompt treatment (EDPT): EDPT is that the main strategy of malaria management and radical treatment is important for all the cases of malaria to forestall transmission of malaria within the community.

b. Vector control: Mosquitoes within the community is controlled with the mixture of multiple activities like chemical and biological control. Personal protecting measures against mosquito bites:

- Use of mosquito repellent creams, liquids, coils, matsetc.
- Screening of the homes with wire mesh.
- Use of bed nets pre-treated with insect powder.
- Wearing clothes that cover most area of the body. Atmosphere management and community awareness concerning detection of mosquito breeding places and their elimination, Swatch Bharat mission promotes folks to stay section clean there by elimination of breeding places for mosquitoes.

## 2.8. Malaria Control Strategies in India

Each malaria day focuses on a specific theme. Current the past theme embraces the following

- 2019-“Zero malaria starts with me”
- 2018- - “Ready to beat malaria “
- 2017- “lets shut the gap”
- 2013-15 -“Invest in the future: defeat malaria”
- 2012 - “Achieving progress and impact”
- 2011 - “Counting malaria out”
- Malaria without boarders”.

## 3. Methodology

### 3.1 Introduction

In the analysis, the foremost methodology relies on the statistical models and foretelling ways severally ARIMA, basic foretelling ways in which are conducted to be ready to choose the only

methodology in step with the error results of those applications.

In Statistics and econometrics, and in especially in time series analysis, an autoregressive integrated moving average (ARIMA) model may well be a generalization of an autoregressive moving average (ARMA) model. Every of these models are fitted to time series knowledge either to higher understand the information or to predict future points among the series (forecasting). ARIMA models are applied in some cases wherever knowledge show proof of non-stationary, wherever an initial differencing step (corresponding to the “integrated” a neighborhood of the model) is applied one or tons of times to eliminate the non-stationary. The AR a part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged values. The NA part indicates that the regression error is essentially a linear combination of error terms whose values occurred contemporaneously and at various times among the past. The I (for “integrated”) indicate that the knowledge values are replaced with the excellence between their values and the previous values. The aim of each of this option is to make the model match the knowledge additionally as attainable. World Non-seasonal ARIMA models are typically denoted ARIM (p, d, q) where parameters p, d and q are non-negative integers, p is that the order (no of your time lags) of the autoregressive model, d is that the degree of differencing, and q is that the order of the moving average model. Seasonal ARIMA models sometimes denoted ARIMA (p, d, q) (P, D, Q) m, wherever m refers to the no's of periods in every season, and the upper case p, d, q refer to the autoregressive, differencing and moving average terms for seasonal a part of the ARIMA mode.

### 3.2. Autoregressive Integrated Moving Average Models (ARIMA)

ARIMA indicates that the Autoregressive Blended Moving Average is generally referred to as the Box-Jenkins Model. George Box and Gwilym Jenkins created the model for time series data (North Dakota State University). Univariate ARIMA can be a way to predict, referring only to its own sequence, when predicting (Morisson, n.d.). In order to be able to perform ARIMA time series analysis data, a minimum of 40 observation points should be included (Morisson, n.d.).

First, whether it is stationary or non-stationary, time series information that can be analyzed by ARIMA must be forced to be evaluated. Time series data should be stationary in order to be able to perform ARIMA, unless it is appropriate to implement different processes, which can be clarified in the following paragraph. Stationary means that the characteristics of a time series do not trust the time at which the series is produced (Otexts,2017).

In addition, the variation in time series data should be constant over time. The information shown in table one below, for example, is not stationary. Therefore, it shows a rise over time that visually demonstrates that time depends on data. Generally, the data is generally not stationary if there is a trend shifting the shift in data (Morisson,n.d.).

In fact, the precise opposite is shown in table two. There is no connection between the data and time and it varies haphazardly. Because of the figures because of the difficulties of the season, get extra dramatic and stationary information, stationary data mostly has stationary movement in most cases in a graph of seasonal information. However, once information from time series is not stationary, distinction technique is typically applied to stationary the current information. In fact, the technique of excellence is the act of subtracting an observation from the previous one to look at the excellence of recent and previous observations. For all findings, until subtracting is over.

Instead of time series details itself, the difference in time series data is used. Further applying the technique of excellence, if distinction information remains not stationary, a second distinction can be applied, and implying that in order to obtain stationary information, the distinction of the primary variations should be measured. Within the following equations, the main and hence, the second distinction technique shown is always shown:

$$D_t = x_t - x_{t-1}$$

$$D_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$$

Another theoretical input, very important for ARIMA is that the beginning of autocorrelations, that represents the degree of resemblance between a given time series and a lagged style of itself over its time intervals.

Lags are the number of periods (Morisson, n.d.) isolated from the observation. For instance, an autocorrelation at the primary lag calculates, but over time, the knowledge of the series is linked to any distinct one. For the time series analysis, autocorrelations are important to understand, especially for ARIMA, because of which the dependency of information on itself is seen and thus the dependency of the method changes over time with lags to itself.

Moreover, the components of it should be unwritten to be able to interpret ARIMA. Therefore, ARIMA has two components: Autoregressive models and Moving Average models.

### 3.3 Autoregressive Models (AR)

Any  $X_t$  observations can be clarified by adding the error variable, that is  $E_t$  (Morisson, n.d.), to different operations of its previous observation  $X_{t-1}$ . This therefore means that with  $X_{t-1}$  and all the other necessary constants and figures, which are derived from time series; it is possible to predict the  $X_t$  value. A way to achieve these statistics by the use of the application of opinions is described in a very completely different section of the methodology. The effective equation is to bring the analysis of the AR time series into a formula (the equation is taken from Jeff Morisson's onlinesource):

$$X_t = A_1 * X_{t-1} + E_t \text{ Where,}$$

$X_t$  = time series under examination

$A_1$  = the autoregressive constraint of order one

$X_{t-1}$  = the time series lagged one period  $E_t$  = the error phrase of the model

As mentioned constants and their figures like

$A_1$  and  $E_t$  would be obtained from Views. Moreover, the ways are explained in one in all the next sections. Returning to the equation higher than, it shows an AR technique with one lag that means it is merely analyzing the time series based on the previous series of observation. Thus, it is an AR1 technique. An AR2 technique with second lag would be as shown below:

$$X_t = A_1 * X_{t-1} + A_2 * X_{t-2} + E_t$$

(Morisson, n.d.)

With the equation above a forecast with 2 lags can be conducted, which means the  $X_t$  value is depending on previous 2 observations,  $X_{t-1}$ ,  $X_{t-2}$ . Next section will justify the Moving averages part which is second part of ARIMA models.

### 3.4. Moving Average (MA)

Second part of ARIMA technique is moving average models. The excellence between the autoregressive model and the moving average model is that the moving average model puts further concentrate on error constant of the preceding observations or also named as previous lags. For example, in Moving Averages model in its place of  $X_{t-1}$ ,  $X_{t-2}$ ,  $X_{t-3}$ , it evaluates the error constants like,  $E_{t-1}$ ,  $E_{t-2}$ , and  $E_{t-3}$ .

An example of a moving average equation is shown below:

$$X_t = -B_1 * E_{t-1} + E_t$$

$B_1$  is an MA of order one, as the equation on top indicates, and it is multiplied by the error term of lag

1. Therefore, this is the expected future value in line with the moving averages model  $X_t$ , usually depending on the term of error. With constant logic, it will be an equation with twolags:

$$X_t = -B_2 * E_{t-2} - B_1 * E_{t-1} + E_t$$

### 3.5 ARIMA

The thesis currently moves on to the mixed model, which is ARIMA (Box-Jenkins), as every model is known. In order to conduct a series of accurate forecasts, ARIMA primarily mixes every equation. A model can accept two AR terms in ARIMA, which can be seen as ARIMA of order (1, 0, 0), or a model can accept 2 MA terms, which can be shown as ARIMA of order (0, 0, 1) in ARIMA. Another example of an ARIMA model could be an ARIMA model order of (1, 1, 1) that could define the AR term of one order, the initial distinction of the analyzed time series data and the first-order MA term of the analyzed time series data and the initial order MA term. In addition, the equations are

generated with ARIMA orders and thus the excellence is reliable.

In the ARIMA model, the key problem is choosing the lags and judging increasing combination will be the easiest for the time series data we have. However, there is not always a regular solution to this problem and there is no correct model. After all, it continually depends on the degree of excellence, ARIMA's lags, and the abundance of combinations that come-at-able. Therefore, there is no correct model as stated, but the simplest fit from the models' choice options.

In short, the ARIMA model's expectations and wishes are firstly stationary, then application of differentiation if there is no stationary, secondly constant variance over time, last classification of autoregressive model lags and moving average models. The tests in view are applied for the competence of stationary and hence the number of lags, the lags are chosen in line with the results of those tests.

### 3.6 About SPSS

The SPSS (Social Science Statistical Package) can be a versatile and open software designed to perform a range of statistical procedures. During a sort of disciplines, SPSS code is commonly used and is out there among the University of Australian state from all personal computer equipment.

It should be remembered that SPSS is not the only statistical software—there are more than a few others that you may come across if you are pursuing a career that needs details to be found out. STATA and SAS (and there are many others) are included in a variety of other more popular statistical sets.

For this session, but on SPSS, the key goal is. Once you open SPSS on our laptop at the beginning, you should see something that looks familiar. Mechanically, SPSS assumes that you simply need to open an accessible file, and opens a dialogue box immediately to ask which file you would like to open. If we choose to open a file, it will make it easier to access the GUI and windows in the SPSS.

### 3.7 Modeling Methods

The available modelling methods are:

**Expert Modeller:** For each dependent sequence, the Expert Modeller eventually finds the best-

fitting model. If freelance (predictor) variables are listed, the Expert Modeller selects those who have a statistically significant relationship with the dependent series for incorporation into ARIMA models. Wherever possible, model variables are transformed by differentiation and/or a square root or natural log transformation. The Professional Modeller finds both exponential smoothing models and ARIMA models by design. You can, but restrict the expert modeller to only searching for ARIMA models.

**ARIMA:** To specify a custom ARIMA model, use this option. Because of the degree of distinction, autoregressive and moving average orders are expressly defined. For any or all of them, you can accept freelance (predictor) variables and definition transition functions. You may jointly specify automatic outlier detection or specify a defined set of outliers.

**Estimation and Forecast Periods:** The cycle of estimation determines the set of cases that will not determine the model. By default, all cases within the working dataset are included in the estimation period. To set the estimation length, select the time or case range supported in the desired case window. The estimate duration used by the procedure may vary by dependent variable, depending on the information provided, and thus differ from the value displayed. The true estimation period is the period left for a given dependent variable when all contiguous missing values of the variable occurring at the beginning or end of the specified estimation period are excluded.

**Forecast Period:** The forecast period begins at the primary case after the estimation period, and by default goes through to the last case inside the active dataset. You will set the tip of the forecast period from the alternatives tab.

### Specifying Options for the Expert Modeller

The Expert Modeller provides choices for qualified the set of candidate models, specifying the usage of outliers, and as event variables.

### Model Selection and Event Specification

The Model tab permits you to specify the categories of models thought-about by the Expert Modeller and to specify event variables.

**Model Type:** The subsequent options are available:

- **All models.** The Expert Modeler considers every ARIMA and exponential smoothing models.
- **Exponential smoothing models only.** The Expert Modeler only considers exponential smoothing models.
- **ARIMA models only.** The Expert Modeler only considers ARIMA models.

### 3.8 Model Specification for ARIMA Models

The Model tab permits you to specify the structure of a custom ARIMA model.

**ARIMA Orders:** In the corresponding cells of the structure grid, enter values for the different ARIMA components of your model. All values should have non-negative integer values. The value reflects the primary order for autoregressive and moving average components. All positive lower orders inside the model are enclosed. As an example, if you state two, orders two and one are included in the model. For the active dataset, cells inside the seasonal column are only allowed if a periodicity has been outlined.

**Autoregressive (p):** The number of autoregressive commands within the model. Autoregressive orders state that current values are typically predicted by previous values from the sequence. As an example, an autoregressive order of two specifies that this value will not be predicted by the value of the sequence of two times in the past.

**Difference (d):** The order of differentiation applied to the series prior to model estimation is defined. When trends are present, differentiation is necessary (series with trends are usually non-stationary and ARIMA modelling assumes stationary) and is used to get rid of their effect. The order of differentiation corresponds to the degree of series trend — linear trend accounts of first-order differentiation accounts, quadratic trend accounts of second-order differentiation accounts, etc.

**Moving Average (q):** Within the model, the number of moving average orders. However, moving average orders specify deviations from the mean series for previous values that are not supposed to forecast current values. For example, one and two moving- average orders state that deviations from the mean of the series from each of

the last two times are thought-about until the series' present values are expected.

#### **Statistics and Forecast Tables:-**

The Statistics tab offers options for showing tables of the outcomes of modelling.

**Display fit measures, Ljung-Box statistic, and number of outliers by model:** Select (check) this function to display a table containing the selected fit measurements, the value of the Ljung-Box, and thus the number of outliers for each calculable model.

**Fit Measures:** For each calculable model, you will be ready to select one or several of the next for inclusion within the table containing fit measures:

- Stationary-square
- R-square
- Root mean sq.error
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC

**Statistics for Comparing Models:** Tables containing statistics calculated across all estimated models are shown by this cluster of choice controls. A separate table is created by any option. You will be ready to select one of the following choices, or a lot of them:

Goodness of Fit:

#### **Table of outline statistics and percentiles for stationary R-square, R-square, root**

mean square error, mean absolute percentage error, mean absolute error, maximum absolute percentage error, maximum absolute error, and normalized Bayesian data Criterion.

**Residual Autocorrelation Function (ACF):** Table of outline statistics and percentiles for Autocorrelations of the residuals across all calculable models.

**Residual Partial Autocorrelation Function (PACF):** Table of outline statistics and percentiles for partial autocorrelations of the residuals across all calculable models.

**Statistics for Individual Models:** This cluster of choice controls show of tables containing elaborated data for every estimated model. Every choice generates a separate table. You will be ready to choose one or many of the subsequent options:

**Parameter Estimates:** For each approximate model, a table of parameter estimates is shown. For Exponential Smoothing and ARIMA models, separate tables are shown. If there are outliers, parameter estimates for them are also shown in a separate table.

**Residual Autocorrelation Function (ACF):** For each approximate model, it shows a table of residual autocorrelations by lag. Boldness intervals for the autocorrelations are included in the table.

**Residual Partial Autocorrelation Function (PACF):** For each approximate model, a table of residual partial autocorrelations by lag is shown. The table contains the intervals of arrogance for the partial autocorrelations.

**Display Forecasts:** Displays a table of model forecasts and confidence intervals for every estimated model. The forecast period is about from the alternatives tab.

**Plots:** The Plots tab provides decisions for displaying plots of the modelling results. Plots for examination Models, this cluster of decisions controls show of plots containing statistics calculated across all denumerable models. Every selection generates a separate plot. You will be able to opt for one or lot of the next options:

- Stationary-squares
- R-square
- Root mean sq.error
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC

• Residual autocorrelation function (ACF)

Plots for Individual Models series, opt for (check) this feature to urge plots of the expected values for each calculable model. You will select one or a lot of the next for inclusion at intervals the plot:

- **Observed values:** The determined values of the dependent series.
- **Forecasts:** The model expected values for the forecast period.
- **Fit values:** The model predicted values for the

estimation period. Solutions require knowledge of different scientific fields, such as analytical methods, differential equations, and stochastic. Very often, the faculty and students interact directly with experimentalists to see the results of their research.

- **Confidence intervals for forecasts:** The boldness intervals for the forecast period.
- **Confidence intervals for fit values:** The arrogance intervals for the estimation period.

**Residual Autocorrelation Function (ACF):** Displays a plot of residual autocorrelations for every calculable model.

**Residual Partial Autocorrelation Function (PACF):** Displays a plot of residual partial autocorrelations for every estimated model. . The table includes the arrogance intervals for the partial autocorrelations.

**Display forecasts:** Shows a table of model forecasts and confidence intervals for every estimated model. The forecast period is concerning from the alternatives tab. Plots tab provides selection for displaying plots of the modelling results.

### 3.9 ARIMA Model in SPSS

This suggests estimating an ARIMA (0, 1, 1) model. To estimate an ARIMA model in SPSS, follow the menus

Analyze → Forecasting → Create Traditional Models

This opens the panel of the Time Series Modeller.

The malaria cases variable, called malaria cases, is first recognized in the variable list on the left side of the window. "Opt for it and use the top arrow to transfer it to the" Based Variables: "window. Next, there is a" Process "button in the middle of the screen. It would possibly reveal the default "Expert Modeller". Click the button and instead opt for "ARIMA." Another button labelled "Criteria ..." opts for the button right next to it and another window will open, called Time Series Modeller: ARIMA Criteria. The middle of this box has what looks like a small table with rows labelled "Autoregressive (p)", "Difference (d)" and "Changing Average (q)." You will see zeros by

default under the heading "No Seasonal." You would like to change the auto-regressive (p) entries from zeros to at least one to approximate an ARIMA (1, 0, 0) prototype. Just click and enter the number '1' for each one in the box. When you have completed, click the continue button within the lower right-hand corner.

You will return to the Time Series Modeller panel once you have done with the Time Series Modeller: ARIMA requirements panel and press proceed. You can see that the model form is set to ARIMA (1, 0, 0) under the button where you pick the strategy to be ARIMA. Next, near the top of the Time Series Modeler dialogue box, press the Statistics tab. Make sure that the check box next to the "Show fit steps, Ljung-Box statistics, and number of outliers by standard" label is checked at the top left. Under the heading "Fit Steps," check the boxes next to "Stationary R sq." and "R square." Because we only estimate one model, under the heading "Statistics for Comparing Models" you can uncheck the box next to "Goodness of fit," shows what this window might look like when you are finished. Statistics window in the SPSS Modeller Time Series row. You should click OK to estimate the model once you have done doing this. Executing this process will produce results that include two tables and one plot. Model explanation and model results from estimating the ARIMA (1, 0, 0) model for this instance in SPSS..

## 4. Empirical Work.

### 4.1 The SPSS Procedure

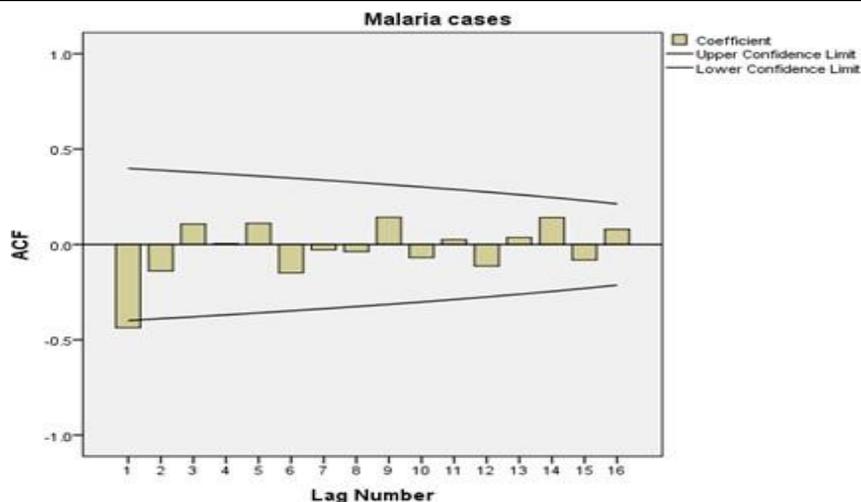
To create an ARIMA model of a single time series, you initially got to produce an autocorrelation function (ACF) and a partial autocorrelation function (PACF) for the time series variable and also the process for doing thus is represented carefully within the SAGE Research Methods Datasets example for time series ACFs and PACFs. The model ACF and PACF functions of original information seem to suggest that the data follow ARIMA processes. The failure of the correlations in Figure 1 to converge to zero indicates that the time series is non-stationary and may be differenced.

**Table 1: ACF for Incidence of Malaria in India**

Autocorrelations					
Series: Malaria cases					
Lag	Autocorrelation	Std. Error <sup>a</sup>	Box-Ljung Statistic		
			Value	df	Sig. <sup>b</sup>
1	-.437	.199	4.794	1	.029
2	-.138	.195	5.299	2	.071
3	.107	.190	5.617	3	.132
4	.005	.185	5.618	4	.230
5	.111	.179	5.999	5	.306
6	-.149	.174	6.731	6	.346
7	-.028	.169	6.759	7	.454
8	-.038	.163	6.812	8	.557
9	.143	.157	7.638	9	.571
10	-.069	.151	7.848	10	.644
11	.025	.144	7.877	11	.724
12	-.113	.138	8.552	12	.741
13	.036	.131	8.629	13	.800
14	.141	.123	9.942	14	.766
15	-.081	.115	10.441	15	.791
16	.079	.107	10.995	16	.810

a. The underlying process assumed is independence (white noise).

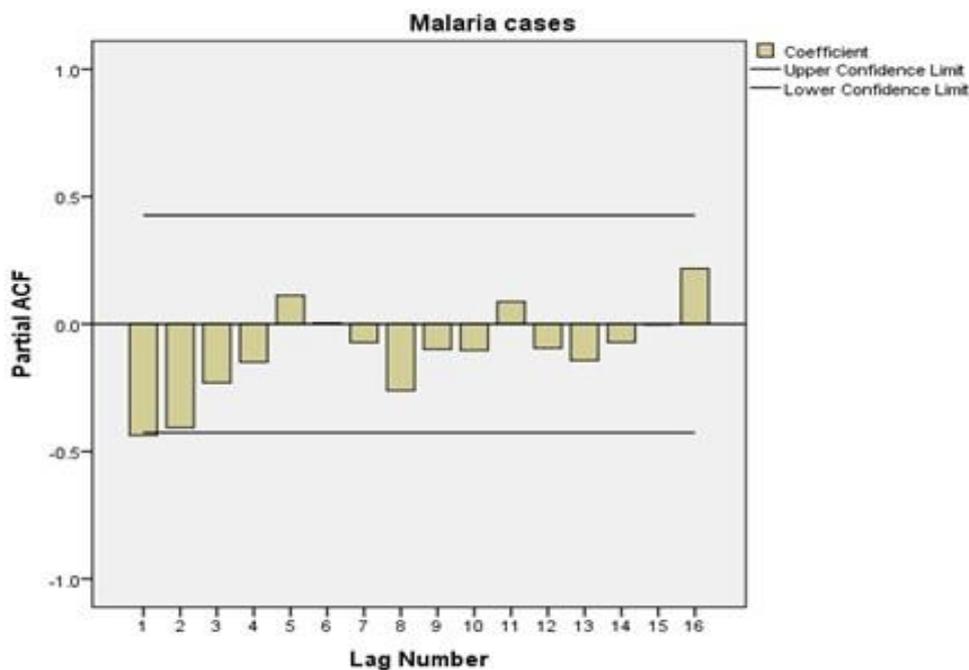
b. Based on the asymptotic chi-square approximation.



**Chart 1: ACF for malaria cases in India**

**Table 2: PACF for Malaria Cases in India.**

<b>Partial Autocorrelations</b>		
Series: Malaria cases		
Lag	Partial Autocorrelation	Std. Error
1	-.437	.213
2	-.407	.213
3	-.230	.213
4	-.149	.213
5	.112	.213
6	.004	.213
7	-.072	.213
8	-.261	.213
9	-.099	.213
10	-.103	.213
11	.087	.213
12	-.095	.213
13	-.143	.213
14	-.071	.213
15	-.003	.213
16	.218	.213



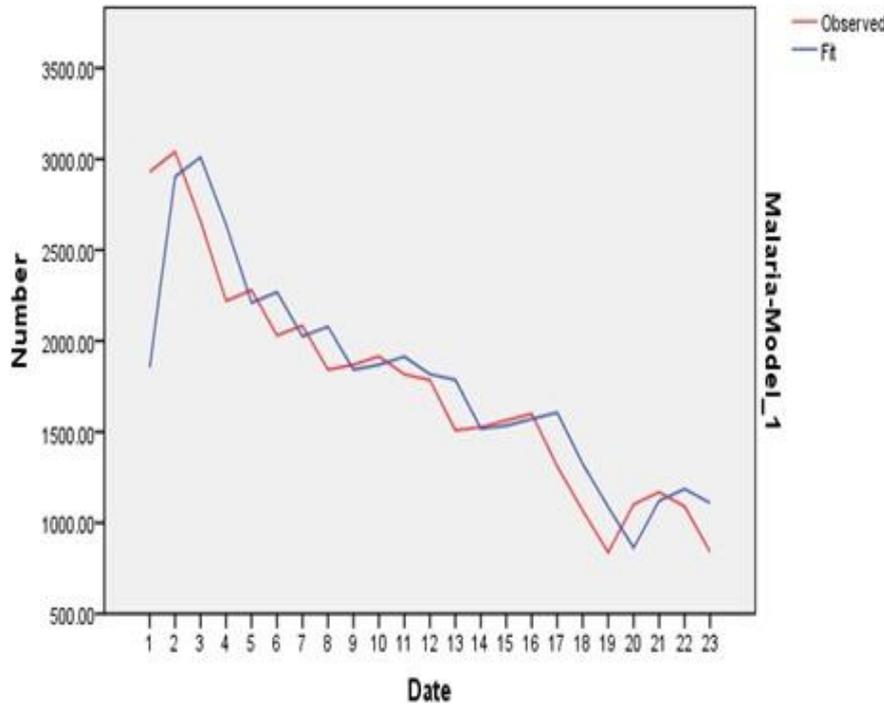
**Chart 2: PACF for malaria cases in India**

Shows a moderately large negative spike at the initial lag followed by correlations that bounce around between being positive and negative and

each one of are either not statistically vital or just barely cross the verge of statistical significance

**Table 3: Model Description**

			Model Type
Model ID	Malaria Cases	Mode l_1	ARIM A (1,0,0)



**Chart 3: ARIMA model parameters for malaria**

The first table identifies the variable utilized throughout this analysis and that the model calculable was an ARIMA (1, 0, 0) model. The second table reports the stationary R-squared of 0.753 and therefore the R-squared of 0.753. It additionally reports the worth of the Ljung–Box Q statistic (16.262), its associated degrees of freedom (17), and its associated level of statistical significance (0.505). This, collective with the Ljung-Box Q statistic, suggests that the ARIMA (1, 0, 0) model suitably modeled the dynamics for this time series

**4.1 Model Validation:**

Several models were created and among many

models, the simplest appropriate designated supported 3 measures, notably normalized baysian information criteria (BIC) mean absolute percentage error (MAPE) and stationary R-squared. Whereas, worth (lower price) is that the MAPE price where performed, the consecutive worth of stationary square suggests an even bigger proportion of variance of the variable explained by the right model.. After careful examination, it's apparent that the mean absolute error percentage is least i.e. MAPE = 9.150, in the model ARIMA (1, 1, 1). The model ARIMA (1, 1, 1) was the foremost appropriate prophetic model in our study.

**Table 4: MAPE, Normalized Bic and Stationary R-Squared Values for ARIMA Models**

Models	MAPE	Normalized BIC	Stationary R-squared
ARIMA(1,0,0)	12.207	11.762	0.753
ARIMA(0,1,0)	10.353	10.572	0.003
ARIMA(0,0,1)	21.345	12.465	0.501
ARIMA(1,1,0)	10.436	10.756	0.005
ARIMA(1,0,1)	12.134	11.962	0.749
ARIMA(0,1,1)	10.523	10.752	1.010
ARIMA(1,1,1)	9.150	10.817	0.128

**Table5: Predicted ARIMA (1, 1, 1) Model for Malaria Cases in India**

Year	Malaria cases (in thousands)	Predicted values
1995	2930.000	-----
1996	3040.000	2840.962
1997	2660.000	2907.286
1998	2220.000	2610.485
1999	2280.000	2264.952
2000	2030.000	2261.522
2001	2085.484	2055.091
2002	1842.019	2050.319
2003	1869.403	1854.876
2004	1915.363	1832.414
2005	1816.569	1825.737
2006	1785.129	1730.547
2007	1508.927	1677.506
2008	1526.210	1472.658
2009	1563.574	1446.519
2010	1599.986	1435.097
2011	1310.656	1425.377
2012	1067.824	1218.066
2013	836.916	1305.384
2014	1102.205	857.474
2015	1169.261	978.0167
2016	1090.724	984.890
2017	840.838	906.098

## Conclusion

The time series modelling of accumulated malaria cases in Republic of India. Our analysis of the information has shown that the accumulated malaria cases that exhibits a ACF , PACF and ARIMA(1, 0, 0), ARIMA(0, 1, 0), ARIMA(0,0, 1), ARIMA(1, 1, 0), ARIMA(1, 0, 1), ARIMA(0, 1, 1) and ARIMA(1, 1, 1) among all these models ARIMA(1,1,1) was the most suitable predictive model in our study. Which showed the lowest MAPE 9.150ARIMA (1,1,1) , models are useful in modelling the temporal dependence structure of a time series as they explicitly assume temporal dependences between notably for seasonal diseases, ARIMA(1, 1, 1) models have shown to be adequate tools for use in epidemiology surveillance. Our study provides ARIMA (1,1, was selected as best suited model to predict the future incidents of malaria cases within the fourth coming period in India. More advanced predictive models may be developed taking into account precipitation, extrinsic incubation period etc. for more prediction that is accurate. The model may be applied at smaller periods or geographical levels such as district level to predict the malaria fever incidents.

## References

- [1].Malaria fact sheet No 94 March 2018.
- [2].World malaria day 2019.
- [3].Time series forecasting system sas institute retrieved 19 march 2017.
- [4].Forecasting: principles and practice may2015.
- [5].Box Jenkins model. Sap retrieved 20014.
- [6].Malaria site, Web portal:  
www.malariasite.com/malaria/treatment.
- [7]. IBM SPSS statistics 25.0.0.2 –detailed system requirements.
- [8]. Sarma, K.V.S (2010), Statistics made simple, Do it yourself on PC, Prentice Hall of India.
- [9].Cryer, J. D. 1986. Time series analysis.Boston, Mass.: Duxbury Press.
- [10]. Kohn, R., and C. Ansley. 1986. Estimation, prediction, and interpolation for ARIMA models with missing.