**Chapter One**

**Introduction**

* 1. **Background of the study**

Recent years Bangladesh has achieved a remarkable progress in controlling communicable diseases, but country has still been facing a tremendous pressure in respect of public health problems especially controlling the emerging or re-emerging diseases.

The disease is transmitted through the bite of the Aedes aegypti and Ae. albopictus mosquitoes. The epidemic in 2000was likely due to introduction of a dengue virus strain from a nearby endemic country, probably Thailand. The first epidemic of dengue was reported in the capital city, Dhaka in the year 2000[10,11]. Since then the disease has shown an annual occurrence in all major cities of the country. During January 2000–December 2021, Bangladesh recorded a total of 22 245 cases and 233 deaths (1.04%). Of these, Dhaka accounted for 20 115 cases and 181 deaths (0.9%).

* 1. **Historical background of dengue**

Dengue is an emerging mosquito-borne flavi-virus infection which is now endemic in more than 100 countries with approximately 4 billion people at risk of contracting the disease and 96 million human cases per year (WHO, 2018). Bangladesh is one of the 10 countries of the World Health Organization (WHO) Southeast Asia Region where 52% of world’s total population at risk of dengue lives (WHO, 2011). Dengue fever was first recorded in the 1960s in Bangladesh (then East Pakistan) and was known as “Dacca fever” (Russell et al., 1966), but it was not until the 21st Century that it became epidemic and finally endemic with establishment of the “domesticated” mosquito vector Aedes aegypti and urban cycles.[[1](#_ENREF_1)]

Since January 2021, Bangladesh has reported over 15,000 dengue cases and at least 57 deaths. In the most recent updates from the Directorate General of Health Services (DGHS), on September 3, the country had reported 330 new cases in the previous 24-hour period. This is the highest single day number of dengue cases reported in the current outbreak. According to the DGHS, Dhaka accounts for 88% of the total cases since January.

* 1. **Transmission of dengue virus**

Dengue is a mosquito-borne viral infection causing a severe flu-like illness and, sometimes causing a potentially lethal complication called severe dengue. The incidence of dengue has increased 30-fold over the last 50 years. Up to 50-100 million infections are now estimated to occur annually in over 100 endemic countries, putting almost half of the world’s population at risk.

Dengue is a vector-borne disease transmitted by the bite of an infected mosquito. There are 4 serotypes of the virus that causes dengue. These are known as DEN-1, DEN-2, DEN-3, DEN-4.Severe dengue is a potentially lethal complication which can develop from dengue infections. It is estimated that there are over 50-100 million cases of dengue worldwide each year and 3 billion people living in dengue endemic countries.

Dengue is mainly transmitted by a mosquito (Aedes aegypti) and is distributed across all tropical countries. Ae. aegypti and other species such as Ae. albopictus are highly adaptive and their combined distribution can spread dengue higher up north across Europe or North America during summer. (Note: Travellers already infected with the virus also spread the disease when they get bitten by the local Aedes mosquito population).Dengue outbreaks can occur anytime, as long as the mosquitoes are still active. However, in general, high humidity and temperature are conditions that favour mosquito survival, increasing the likelihood of transmission.

**Dengue fever:**

Dengue causes flu-like symptoms and lasts for 2-7 days. Dengue fever usually occurs after an incubation period of 4-10 days after the bite of the infected mosquito.

High Fever (40°C/ 104°F) is usually accompanied by at least two of the following symptoms:

• Headaches

• Pain behind eyes

• Nausea, vomiting

• Swollen glands

• Joint, bone or muscle pains

• Rash Severe dengue

When developing into severe dengue, the critical phase takes place around 3-7 days after the first sign of illness. Temperature will decrease; this does **NOT** mean the person is necessarily recovering. On the other hand, special attention needs to be given to these warning signs as it could lead to severe dengue:

• Severe abdominal pain

• Persistent vomiting

• Bleeding gums

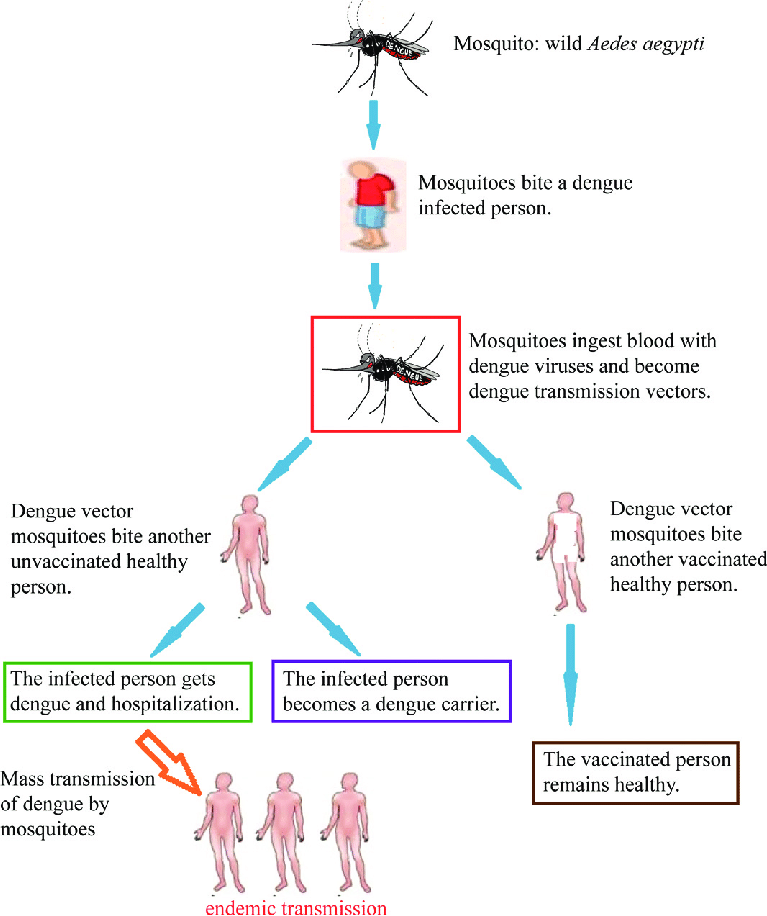
• Vomiting blood

• Rapid breathing

Fatigue/ restlessness When severe dengue is suspected, the person should be rushed to the emergency room or to the closest health care provider as it causes:

* Plasma leaking that may lead to shock and/or fluid accumulation with/without respiratory distress;
* Severe bleeding;
* Severe organ impairment.

Dengue viruses are transmitted to humans through the bites of infective female Aedes mosquitoes. Most commonly, the mosquitoes involved are Aedes aegypti and Aedes albopictus, two species which can also transmit other mosquito-borne viruses, including zika and chikungunya. Other infection routes are reported from mother to child as well as blood transmission.



* 1. Fig: Transmission of dengue virus
  2. **Problems and motivations**

When we dealing with time series data for forecasting purpose, a number of special problem arise that often data are non-stationary. But for our analysis purpose we assume that stationary. Most of the epidemiological time series data, such as deaths and Cases are often exhibit the phenomenon of non –stationary data. So we are interested to see at what lag the data become stationary, because without stationary data no valid forecast is made.

* 1. **Objectives of the study**
* To monthly reported dengue cases from January 2000 to December 2021 were analyzed to estimate monthly growth
* Controlling the mosquitoes earlier might be more efficient in limiting the dengue outbreaks in Bangladesh
* To forecast the future epidemic outbreak of dengue in Bangladesh
  1. **Source of data**

The data were collected from the onset of the epidemic in Bangladesh from the Directorate General of Health Service (DGHS) and Institute of Epidemiology, Disease Control and Research (IEDCR). The original data format was inappropriate for Time Series Analysis; in this regard, the researchers had to process the data in an appropriate format for further analysis using R programming.

* 1. **Limitation of the study**

The data we have collect is a secondary data. It is quiet impossible to collect primary data of whole Bangladesh that is time consuming and not economical efficient. On the other hand our collected data dengue Cases and Deaths have some zero or missing value which is very difficult to handle.

The limitation of this study is considering only the dengue incidences found in Bangladesh. It is believed that the prevalence of dengue may be influenced by other climatic and environmental factors like rainfall, humidity, and so on. Therefore, there is a lot of scope for future studies to develop and validate the model for the better findings.

**Chapter: Two**

**Literature Review**

**2.1 Introduction**

The main purpose of this chapter is to review the available studies related to the present one. Review of literature provides an opportunity for reviewing the stock of knowledge and information for the proposed research which gives guideline in designing the future research and validating the findings. The important studies which have been conducted in the past are discussed below. In reviewing literatures, focus is given not only on the studies dealing with exported commodities but also on methodologies adopted by different past studies that are related to the methodology of this study.

**2.2 Historic epidemic pattern**

Historically, it has not been determined when DENV first appeared in human populations, mainly because the disease is often asymptomatic and is therefore not diagnosed. The earliest record of dengue comes from a Chinese medical encyclopedia dating back to 992 BC. Moreover, before the end of the 18th century, intermittent epidemics of a specific disease with a strong similarity to dengue occurred in Asia and the Americas; therefore, there is a hypothesis that between the 19th and 20th centuries, the virus probably spread throughout the tropics and subtropics.

Dengue is an emerging mosquito-borne flavivirus infection which is now endemic in more than 100 countries with approximately 4 billion people at risk of contracting the disease and 96 million human cases per year (WHO, 2018). Bangladesh is one of the 10 countries of the World Health Organization (WHO) Southeast Asia Region where 52% of world’s total population at risk of dengue lives (WHO, 2011). Dengue fever was first recorded in the 1960s in Bangladesh (then East Pakistan) and was known as “Dacca fever” ,but it was not until the 21st Century that it became epidemic and finally endemic with establishment of the “domesticated” mosquito vector Aedes aegypti and urban cycles. Dengue is an important cause of morbidity and mortality in Bangladesh.

Managing dengue fever outbreaks in tropical countries where temperatures remain favourable for mosquito breeding and viral replication around the year is a critical challenge. Further, controlling large dengue outbreaks is expensive and involves use of insecticides to control the mosquito population and significant resources to manage sick patients at hospitals. Current vector-control in Bangladesh is a mandate of the local government – the City Corporation/Municipality is responsible for controlling mosquito populations. The programme is largely based on spraying insecticidal chemicals which aim to kill adult mosquitoes, especially during the peak dengue period (August–September) when usually a large number of dengue cases are detected in hospitals. Detecting the critical control points is central to managing dengue. One of these points of intervention involves suppressing the mosquito vector activity at the optimal time in annual population fluctuations, in order to achieve the lowest biting population when environmental conditions for emergence and transmission are most favorable.

Dengue virus is transmitted between humans and mosquitoes: Ae. aegypti and Ae. albopictus are the principal vectors of the disease, and both mosquito species are widely distributed in Bangladesh. In 2019 the country experienced one of the largest dengue fever outbreaks in the history of the country. More than 100,354 people are hospitalized and 266 people (0.1% of hospitalized cases) dying between 1st January 2019 and 31st December 2019. Case data are regularly updated by the Directorate General of Health Services (DGHS, 2019) and monthly updates are publicly shared by DGHS’s surveillance and research wing, Institute of Epidemiology, Disease Control and Research (IEDCR) (IEDCR, 2019). Approximately 90% of the cases are reported in the capital city, Dhaka, consistent with spatial models predicting 94% cases to be localized in Dhaka. An observational study reported 90% of observed cases between June and November during 2010–2014, whilst a model predicted more than 92% cases between August and September in 2000–2021

**2.3 Dengue overview in Bangladesh**

Dengue is endemic in Bangladesh however a surge of cases started in June 2022. Currently, all eight divisions in the country are reporting cases and deaths. This is the second-largest outbreak since 2000, with the largest having occurred in 2019. The current dengue outbreak is unusual in its scale and seasonality.

Between 1 January and 20 November 2022, a total of 52 807 dengue cases including 230 related deaths (case fatality rate = 0.44%) were reported by the Ministry of Health & Family Welfare (MOHFW) (Figure 1). According to information available for 40% of reported cases (n=20 982) the median age is 25 years (range: 0 – 89) with males accounting for 60% of the cases. This is the second highest annual number of cases since 2000, the highest having occurred in 2019, when 101 354 cases including 164 deaths were reported (Figure 2 & 3).

The most affected division is Dhaka, accounting for 70.6% of cases and 60.4% of deaths. Dhaka city, the largest city in Bangladesh, located in Dhaka division, has reported 64.5% (n= 34 071) of the total number of cases. Other affected divisions include Chattogram division (13.2% of cases and 24.8% of deaths) and Khulna division (5.5% of cases and 4.8% of deaths) (**WHO 2022**)

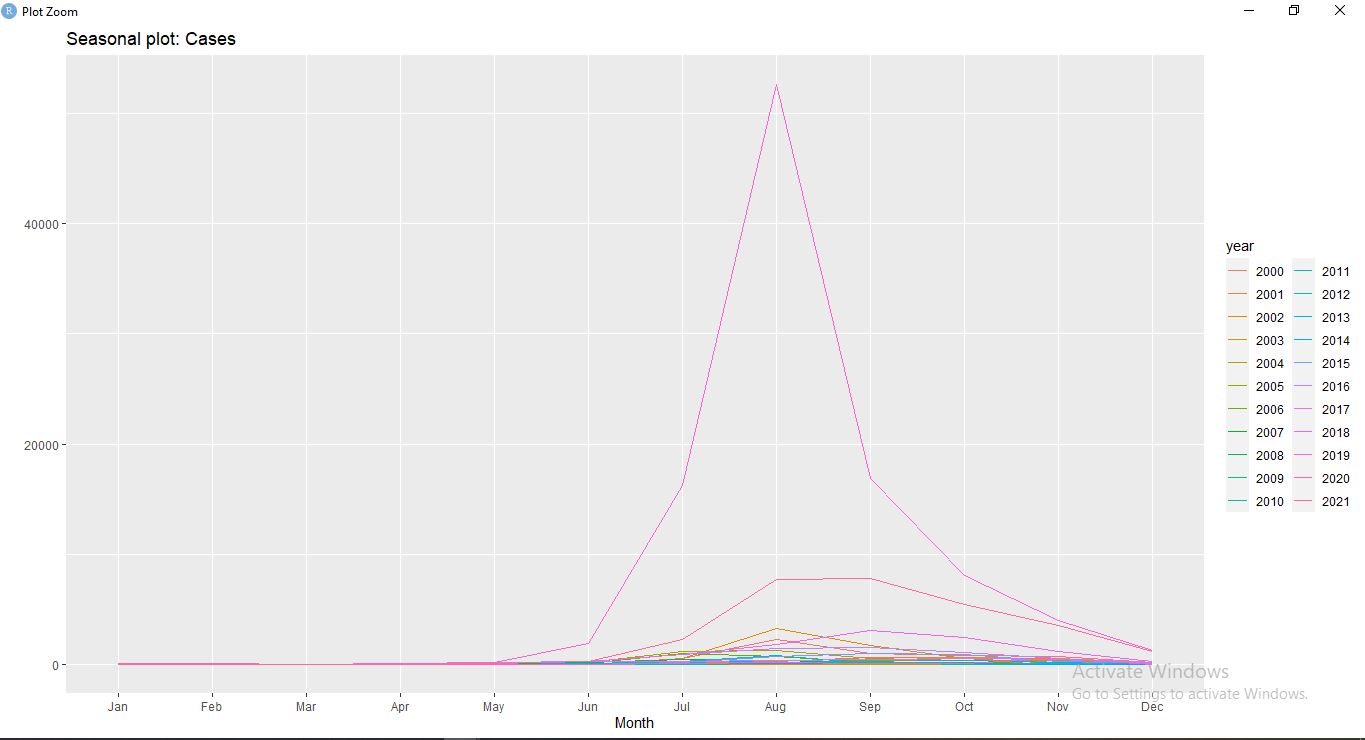


Fig. : Monthly Dengue Cases in Bangladesh 2000-2021

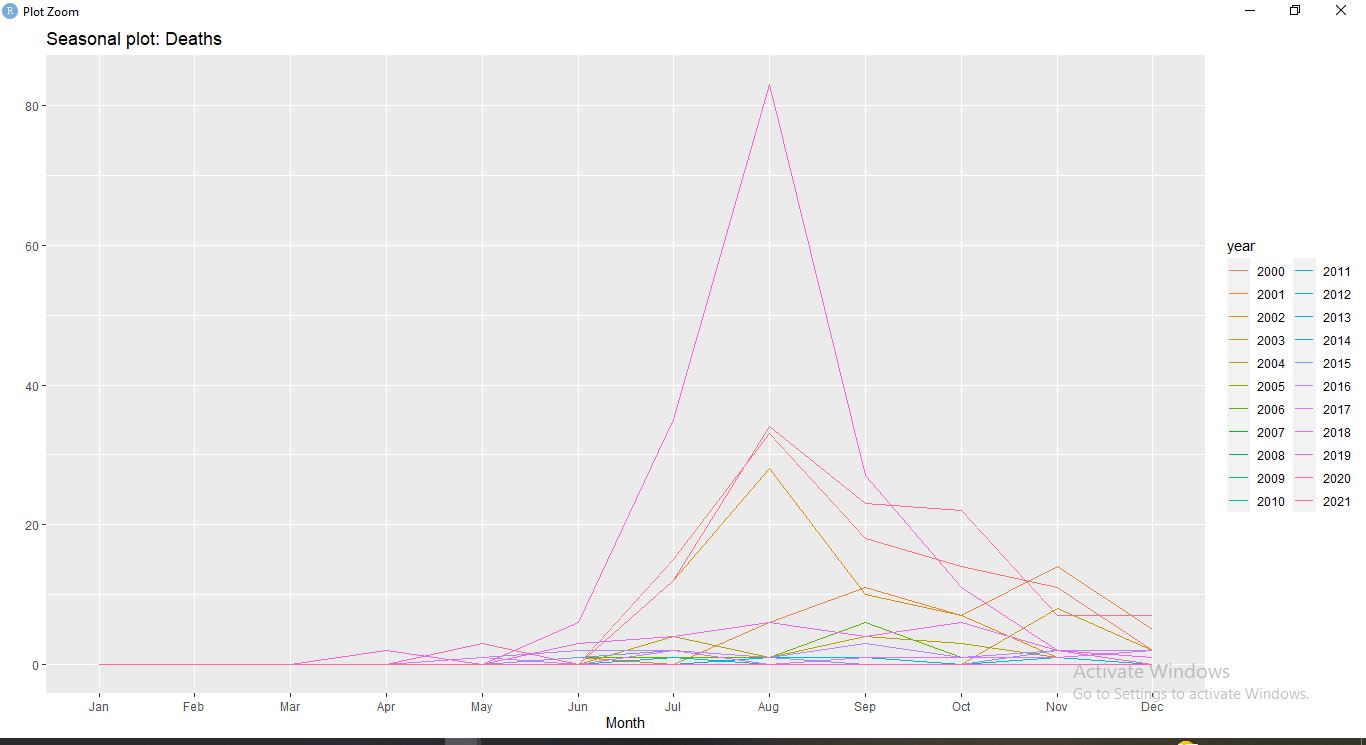


Fig. : Monthly Dengue Deaths in Bangladesh 2000-2021

**2.4 Earlier work on trend Analysis and stationary**

It is one of the most often used time-series models in a variety of sectors of data analysis because it accounts for changing trends, periodic variations, and random disturbances in the data.[[2](#_ENREF_2)] In the statistical analysis of time series, a trend-stationary process is a stochastic process from which an underlying trend (function solely of time) can be removed, leaving a stationary process. The trend does not have to be linear.

Conversely, if the process requires differencing to be made stationary, then it is called difference stationary and possesses one or more unit roots. Those two concepts may sometimes be confused, but while they share many properties, they are different in many aspects. It is possible for a time series to be non-stationary, yet have no unit root and be trend-stationary. In both unit root and trend-stationary processes, the mean can be growing or decreasing over time; however, in the presence of a shock, trend-stationary processes are mean-reverting

**2.5 Earlier work on ARIMA modeling and forecasting**

In the study[[3](#_ENREF_3)] of dengue infected from 2008-2020 shows that AIC value as the indicator, ARIMA (2,1,2) shows the lowest value of AIC, which means ARIMA (2,1,2) is better than the ARIMA (2,1,2) (0,0,1) Therefore, we use the ARIMA model for subsequent analysis. The tentative order of the ARIMA model was identified by inspecting the ACF and PACF plots.

In another research[[4](#_ENREF_4)] shows that Seasonal Autoregressive Integrated Moving Average (SARIMA) models have been developed on the monthly data collected from January 2000 to October 2007 and validated using the data from September 2006 to October 2007. The results showed that the predicted values were consistent with the upturns and downturns of the observed series. The SARIMA (1,0,0)(1,1,1)12 model has been found as the most suitable model with least Normalized Bayesian Information Criteria (BIC)

From the previous studies, it is clear that ARIMA can be used to forecast. Hence, this study aims on selecting the best ARIMA model and to forecast dengue cases in Selangor. The data used for this analysis are secondary data provided by the Vector Borne Disease Control Section of the Ministry of Health Selangor. The study focuses on the collection of data in Selangor. The data showed epidemic weekly dengue outbreak data from January 2018 to November 2020[[5](#_ENREF_5)]. To select the best ARIMA model, their goodness of fit had been compared using the Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE) measurement errors by looking at their smallest error values.

**Chapter: Three**

**Methodology**

**3.1 Introduction**

One of the important types of data used in empirical analysis is time series data. The empirical work based on time series data assumes that the underlying time series is stationary. The time series analysis based on the stationary time series data. In this chapter we will try to give some brief discuss on stationary time series and some basic definitions which is related to this topic such definitions may be data types. ACF, PACF etc. there are different methods of testing stationary of a data set, and Dickey Fuller Test is one of them. In this chapter I discuss methodology of the study.

**3.2 Concept of time series**

A time series is a series of data points listed (or graphed) in time order. Most commonly a time series is a sequence taken at successive equally spaced points in time. A time series is a set of observations Y, (1,2,......n) each one being recorded at a specified time. In other words, arrangement of statistical data in accordance with occurrence of time is known as time series.

According to Ya-lun Chou, "A Time Series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables".

Mathematically, a time series is defined by the functional relationship Y, f(), where Y, is the value of the phenomenon (variable) under consideration at timer.

Thus, the values of a phenomenon or variable at times t1,t2........tn are Y1 Y₂,..... ,Yn respectively, then the series:

t : t1, t2........tn .

Yt : Y1, Y₂,..... ,Yn .

Constitute time series.

**3.3 Example of time series**

1. Many time series arise in economics, such as share prices on successive days, average incomes in successive months, company profits in successive years etc

2. Many types of time series occur in the physical sciences, particularly in methodology marine science and geophysics Examples are rainfall in successive days and air temperature measured in successive hours or months etc.

3. The population of Bangladesh measured at every ten years is an example of demographic time series

4. Time series in commerce such as sales figure in successive weeks or months, advertising expenditure in successive time period.

**3.4 Type of data**

There are three types of data may be available for empirical analysis

1. **Time series data:** A time series is a set of observation on the values that a variable takes at different times. Such as data may be collected at regular time intervals such as weekly, monthly, quarterly, annually, quinquennially, decennially etc.
2. **Cross sectional data:** Cross section data are data on one or more variables collected at the same point in time such as the census of population conducted by the Bangladesh Bureau Statistics (BBS) every 10 years.
3. **Polled data:** Combination of time series data and cross-sectional data are called pooled data.

**3.5 Assumption of a time series**

The main assumptions of a time series are given below:

1. The time gap between various values must be as for as possible equal
2. It must consist of a homogeneous set of values
3. Data must be available for a long period

**3.6 Uses of time series**

The time series analysis is of great importance not only to businessman or an economist but also to people working in various disciplines in natural, social and physical sciences. Some of its use is enumerated below

1. It enables us to determine the type and nature of the variations in the data
2. It helps in planning and forecasting.
3. It is very essential in business and economics with the help of time series we can make plans for the future.
4. It helps to compare the actual current performance of accomplishments with the expected ones and analyze the causes of such variation, if any.
5. It enables us to predict or estimate or forecast the behavior of the phenomenon in future which is very essential for business planning**.**

**3.7 Main component of time series**

The important components of time series are:

1. Trend or secular trend variation
2. Periodic changes or short term movement
3. Seasonal variation
4. Cyclical variation
5. Irregular or random variation

**3.8** **Different type of model used in time series**

The following are the models commonly used for the decomposition of a time series into its components.

* Decomposition by additive hypothesis.
* Decomposition by multiplicative hypothesis.
* Mixed models.

**3.9 Estimation and elimination of trend and seasonal analysis**

The first step in the analysis of any time series is to plat the data. If there are any apparent discontinuities in series, such as a sudden change of level, then break it into homogeneous segments. If there are outlying observations, they should be studied carefully to check whether there is any justification for discarding them. The classical decomposition model

Yt=Tt + St + Ut: t =1,2,3,........,n

Where,

Tt is a slowly changing function known as a trend component.

St is a function with known period referred to as a seasonal component and

Ut is a random noise component**.**

If the seasonal and noise fluctuations appear to increase with the level of the process, then a preliminary transformation of the data is often used to make the transformed data more compatible with the model. Our aim is to estimate and extract the deterministic components T t and St in the hope that the residual or noise component Ut will turn out to be a stationary time series. We can then use the theory of such processes to find a satisfactory probabilistic model for the process Ut to analyze its properties and to use it in conjunction with Tt and St for purpose of prediction and simulation of {Yt} Another approach, developed by Box and Jenkins is to apply differencing operators repeatedly to the series {Yt} until the differenced observations resemble a realization of some Stationary time series {Wt}, we can then use the theory of stationary processes for the modeling, analysis and prediction of {Wt}, and hence of the original process

**3.10 Different time series process**

A timeobservations made sequentially through time Examples occur in a variety of variety of fields, ranging from economies to engineering. Given a set of series data Y1,Y2,Y3, .......,Yn each associated random variable it consist of a mean, say met and zero mean random part. The different time series process is as follows:

**3.10.1 Stochastic process**

A stochastic process may be defined as a collection of random variables {Yt, t∈T}, where T denotes the set of time points. We denote the random variable at time t by Yt if T is discrete and by Yt, if T is continuous. Thus a stochastic process is a collection of random variables which are ordered in time.

In time series analysis, it is often impossible to make more than one observation at time t. So that we may only have one observation on the random variable at time t.

**3.10.2 Stationary process**

A time series {Yt, t=0,±1,± 2,± 3,........} is said to be stationary if it has similar Statistical properties to the "time shifted" series {Yt+h, t= 0,±1,± 2,± 3,........} for each integer h**.**

Simply, a time series {Yt, t=0,±1,± 2,± 3,........}) is said to be stationary time series if it is independent of time t.

**3.10.3 Non stationary process**

A time series {Yt, t=0,±1,± 2,± 3,........}) is said to be non-stationary time series if it is dependent of time t.

**3.10.4 White noise process**

If {Yt} is a sequence of uncorrelated random variables, each with zero mean and variance **σ2,** then clearly {Yt} is stationary with the same covariance function as the IID noise as

γyh =

Which does not depends on t. Such a sequence is referred to as white noise. This is indicated by the notation

{Yt}~ WN(**σ2,0**)

**3.11 Test of stationary**

There are several test of stationary .some important test of them are given below:

1. Graphical analysis(time series plot)
2. Sample autocorrelation function
3. Unit root test

**Graphical analysis (time series plot)**

The single most important thing to do when first exploring the data is visualize the data through graphs. The basic feature of the data including patters and usual observations are mostly easily graphs. Sometimes graphs are also suggest possible explanations for some of the variation in the data. The first step in the analysis of any time series as to plot the data.

**3.12 Approaches to precise forecasting**

There are many approaches to precise forecasting which is given below:

**Moving average (MA)**

The best-known forecasting methods is the moving averages or simply takes a certain number of past periods and add them together, then divide by the number of periods. Simple Moving Averages (MA) is effective and efficient approach provided the time series is stationary in both mean and variance. The following formula is used in finding the moving average of order n, MA (n) for a period t+1.

MA t+1=

Where n is the number of observations used in the calculation

The forecast for time period t+1 is the forecast for all future time periods. However, this forecast is revised only when new data becomes available.

**Single exponential smoothing**

It calculates the smoothed series as a damping coefficient times the actual series plus 1 minus the damping coefficient times the lagged value of the smoothed series. The extrapolated smoothed series is a constant, equal to the last value of the smoothed series during the period when actual data on the underlying series are available. While the simple Moving Average method is a special case of the ES, the ES is more parsimonious in its data usage.

Ft+1= σ Dt + (1- σ)Dt

Where,

Dt is the actual value

Ft is the forecasted value

σ is the weighting factor, which ranges from 0 to 1

T is the current time period.

**Double exponential smoothing**

An exponential smoothing over an already smoothed time series is called double-exponential smoothing. It applies the process described above three to account for linear trend. The extrapolated series has a constant growth rate, equal to the growth of the smoothed series at the end of the data period.

**ARIMA modeling**

The ARIMA model (an acronym for Auto-Regressive Integrated Moving Average), essentially creates a linear equation which describes and forecasts your time series data. This equation is generated through three separate parts which can be described as:

* AR — auto-regression: equation terms created based on past data points
* I — integration or differencing: accounting for overall “trend” in the data
* MA — moving average: equation terms of error or noise based on past data points

The ARIMA model is almost always represented as ARIMA(p, d, q) where each of the letters corresponds to one of the three parts described above. These three letters represent parameters that you will have to provide, and are described as follows:

* p determines the number of autoregressive (AR) terms
* d determines the order of differencing
* q determines the number of moving average (MA) terms

**SARIMA model**

A seasonal autoregressive integrated moving average (SARIMA) model is one step different from an ARIMA model based on the concept of seasonal trends. In many time series data, frequent seasonal effects come into play. The seasonal model may be dividing into two types based on its complexity: an additive model (simple seasonal model) and a product seasonal model. The mathematical expression of the simple seasonal models is:

Xt = St +Tt +It

Where, St, Tt, and It denote seasonal information, trend information, and random fluctuation information in the data, respectively.

Non - seasonal ARIMA models are generally denoted ARIMA (p,d,q ) where parameters p, d and q are non - negative integers , p is the order of the autoregressive model , d is the degree of differencing and q is the order of the moving - average model Seasonal ARIMA models are usually denoted ARIMA ( p,d,q) ( P , D , Q ) m , where m refers to the number of periods in each season , and the uppercase P.D.Q refer to the autoregressive , differencing , and moving average terms for the seasonal part of the ARIMA model .

**3.13 Difference operator for non stationary time series**

If time series values y1, y2, y3,..............,yn indicate that these values are non -stationary we can transform the non-stationary series values by taking the first difference of the non - stationary time series values . That is

The differences of time series values y1, y2, y3,..............,yn are

zt=yt-yt-1 where t= 2,3,4,..........n

Although taking first differences sometimes will transform non stationary time series value into stationary time series values, we sometimes need to use other forms of differencing to produce stationary time series values. We can produce stationary time series value by taking the second differences (the first differences of the first differences of the original time series values).

Therefore,

The second differences of the time series values y1, y2, y3,..............,yn are

zt=(yt-yt-1)-(yt-1-yt-2)

= yt-yt-1-yt-2 for t= 3,4,5......n

**3.14 Auto correlation function (ACF)**

Seasonal patterns of time series can be examined via correlograms . The correlogram ( auto correlogram ) displays graphically and numerically the autocorrelation function ( ACF ) , that is , serial correlation coefficients ( and their standard errors ) for consecutive lags in a specified range of lags ( c.g , 1 through 30 ) . Ranges of two standard errors for each lag are usually marked in correlograms but typically the size of auto correlation is of more interest than its reliability because we are usually interested only in very strong autocorrelations. While examining correlograms one should keep in mind that autocorrelations for consecutive lags are formally dependent. Consider the following example. If the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one, etc. This implies that the pattern of serial dependencies can change considerably after removing the first order auto correlation ( ie . , after differencing the series with a lag of 1 )

The autocorrelation function (ACF) reveals how the correlation between any two values of the signal changes as their separation changes. It is a time domain measure of the stochastic process memory, and does not reveal any information about the frequency content of the process. Generally, for an error signal, et, the ACF is defined as,

*P*k =

**3.15 Partial auto correlation function (PACF)**

Another useful method to examine serial dependencies is to examine the partial autocorrelation function (PACF) - an extension of autocorrelation, where the dependence on the intermediate elements (those within the lag) is removed. In other words the partial autocorrelation is similar to autocorrelation, except that when calculating it, the (auto) correlations with all the elements within the lng are partial led out. If a lag of 1 is specified (i.e. there are no intermediate elements within the lag) then the partial autocorrelation is equivalent to auto correlation. In a sense, the partial autocorrelation provides a “cleaner " picture of serial dependencies for individual lags (not confounded by other serial dependencies)

In time series analysis, the partial autocorrelation function (PACF) gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags.

This function plays an important role in data analysis aimed at identifying the extent of the lag in an autoregressive (AR) model. The use of this function was introduced as part of the Box–Jenkins approach to time series modeling, whereby plotting the partial auto correlative functions one could determine the appropriate lags p in an AR (p) model or in an extended ARIMA (p,d,q) model.

**3.16 Mean absolute deviation (MAD)**

Expresses accuracy in the same units as the data, which helps conceptualize the amount of error Outliers have less of an effect on MAD than on MSD. The equation is:

Where yt equals the actual value, equals the fitted value, ŷt and n equals the number of observations.

**3.17 Mean absolute percentage error (MAPE)**

Expresses accuracy as a percentage of the error. Because this number is a percentage, it can be easier to understand than the other statistics. For example, if the MAPE is 5, on average, the forecast is off by 5 %. The equation is:

×100(yt ≠0)

Where yt equals the actual value, equals the fitted value, ŷt and n equals the number of observations.

**Akaike's Information Criterion (AIC)**

The AIC estimates the accuracy of each model and it helps to choose the appropriate model and calculated utilizing the following equation:

AIC=−2() + (); where, AIC = Akaike Information Criterion, LL = log likelihood.

With the lowest value of AIC is preferred.

In comparing two or more model with the lowest value of AIC is preferred. One advantage of AIC is that it is useful for not only in sample but also out of sample forecasting performance of a regression model. Also it is useful for both nested and non nested models. It has been used to determine the lag length in an AR(p) model.

**Bayesian Information Criterion (BIC)**

The BIC is analogous to the formula of AIC, however, having a distinct penalty for the number of parameters and it is calculated as:

BIC=−2LL + kln (T); where, LL = log likelihood, k = number of total parameters, T = number of observations.

**3.18 Conclusion**

In this chapter, we discuss about some basic topics of time series that will help us to complete our further work. Thus time series estimation needs to consider many things which are given above. Our aim is to estimate model and forecast future values

**Chapter: Four**

**Analytical Result**

**4.1 Identification of data patterns**

At first we summarize the data and plot them

**Fig:** Dengue Death and cases in Bangladesh from 2000-2021

From the graph we can observe that dengue increasing Bangladesh year to year with irregular fluctuation.

**4.2 Summary of data:**

**Table:** Mean, maximum, minimum and standard deviation of (Cases and Death)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Total | Minimum | Maximum | Mean | Standard Deviation | Median |
| Cases | 181622 | 375 | 100949 | 8256 | 21645 | 1975 |
| Death | 566 | 0 | 166 | 26 | 43 | 7 |

**Mean:**

A mean is the simple mathematical average of a set of two or more numbers. The mean for a given set of numbers can be computed in more than one way, including the arithmetic mean method, which uses the sum of the numbers in the series, and the geometric mean method, which is the average of a set of products. However, all of the primary methods of computing a simple average produce the same approximate result most of the time.

Formulation: 𝑀𝑒𝑎𝑛=

On verge yearly dengue confirm Cases of Bangladesh is 8256 and Deaths is 26.

**Standard Deviation:**

The standard deviation is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The standard deviation is calculated as the square root of variance by determining each data point's deviation relative to the mean.

Formulation:

𝑆𝑥=

Here,

n = The number of data points

𝑥𝑖 = Each of the values of the data

𝑥̅ = The mean of 𝑥𝑖

The actual amount of dengue Cases and Death in Bangladesh on average differs from the mean is about 21645 and 43 respectively.

**R-Squared:**

R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model.

𝑅2=1− =

Where,

ŷ is estimated value of dependent variable.

𝑦̅ denotes the average value of y.

**4.3 Checking and obtaining Stationary of Death and Cases**

By using plot we can determine stationary

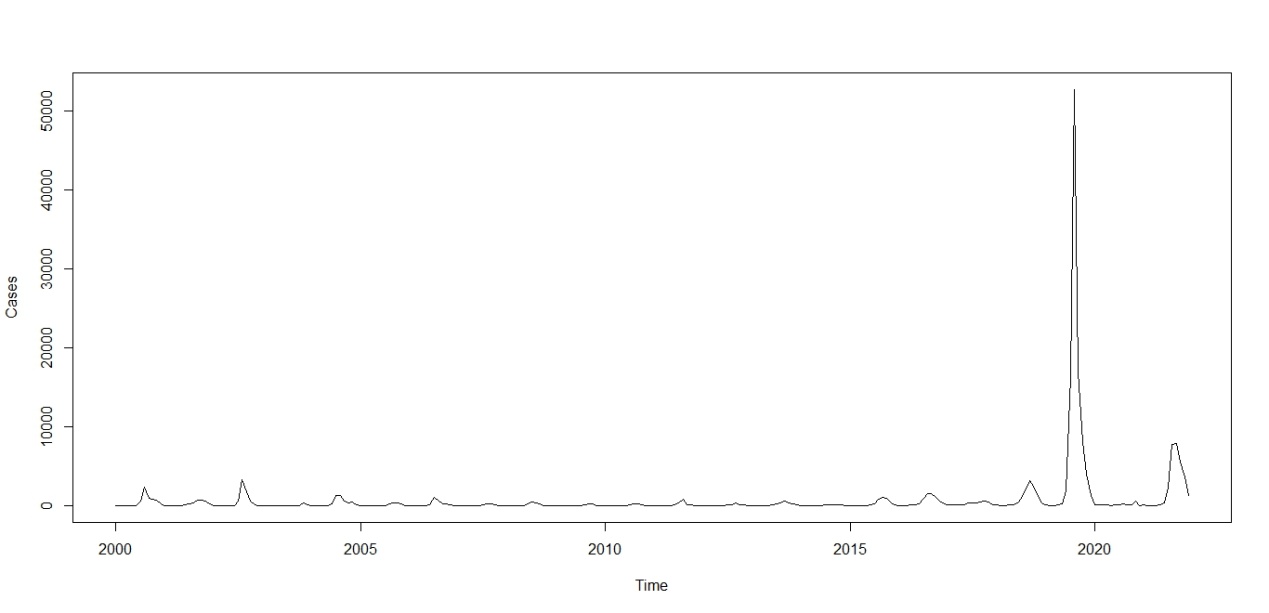


Fig: Time series plot of dengue‐infected people in Bangladesh from 2000-2021

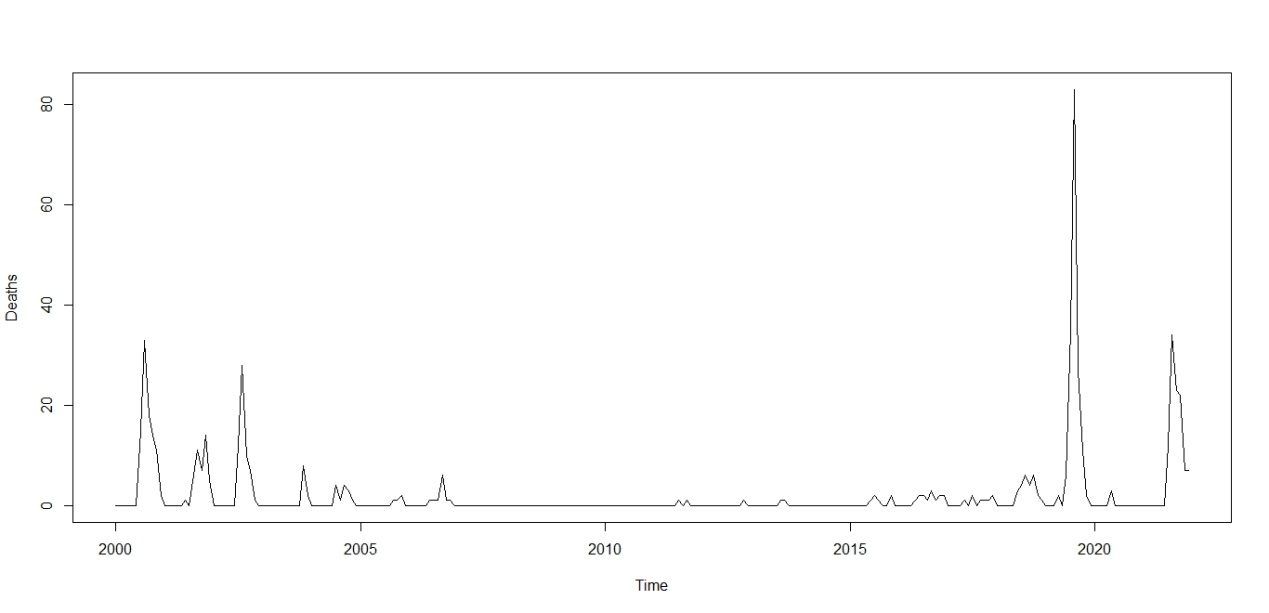


Fig: Time series plot of dengue‐Deaths people in Bangladesh from 2000-2021

From this plot it is clear that the data must take account of the obvious increasing trend. Hence the original series is not stationary.

A mathematical transformation is a convenient method for accounting for the increasing variation. So we use log transformation to stabilize the variation whose time plot is shows in figure

**4.4 Model selection of Death and Cases**

The plot of ACF and PACF can give a primary guess about the parameter p and q for the ARIMA model.

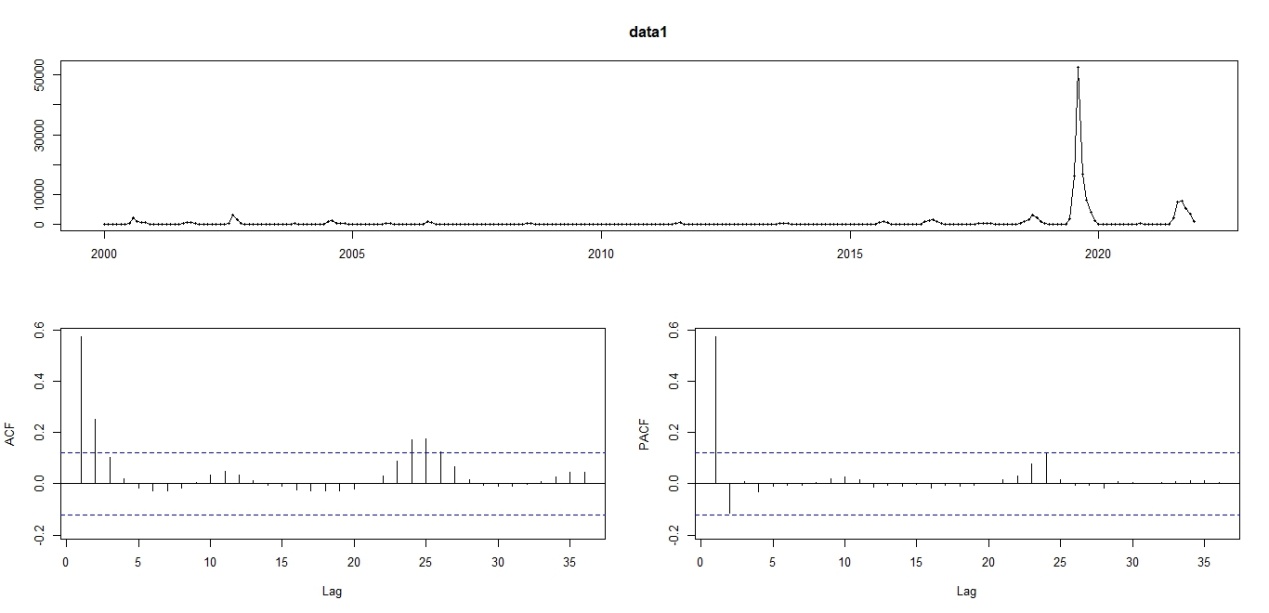


Fig: ACF and PACF for Cases of dengue in Bangladesh

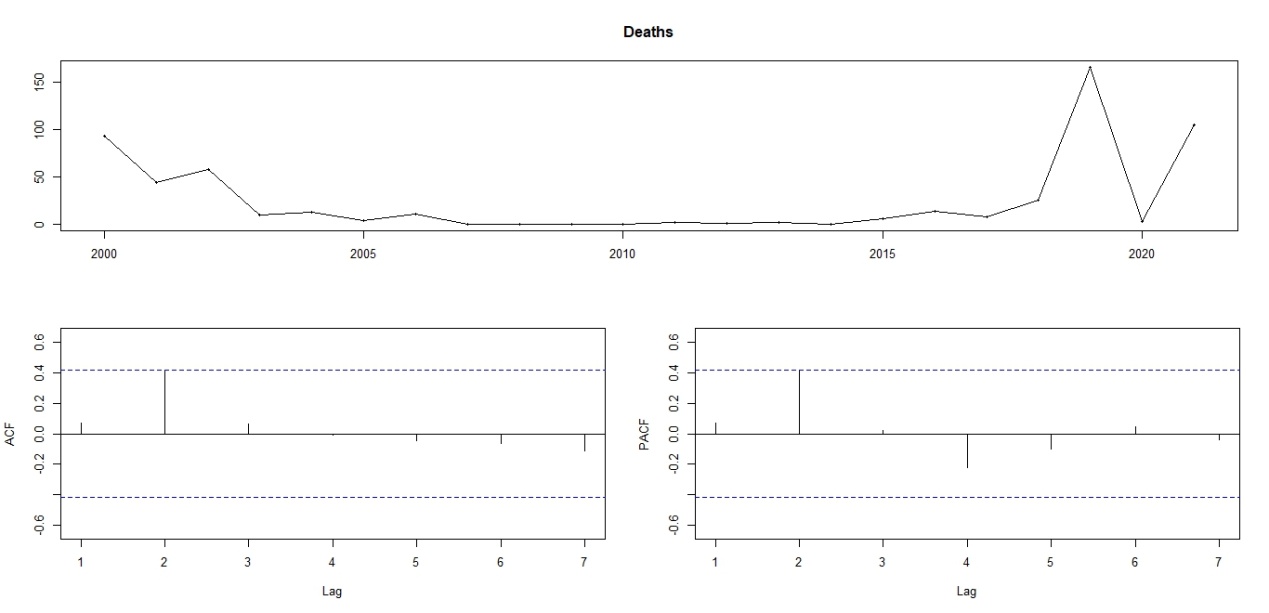


Fig: ACF and PACF for Deaths of dengue in Bangladesh

Now we consider the different types of tentative models as much as possible from which we select the best model using the model selection criterion. Since the characteristics of a good ARIMA model is parsimonious ignoring the higher order of p and q, the tentative models on the basis of model selection criterion are as follows:

Tab: Different ARIMA model comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **SSE** | **AIC** | **BIC** | **Adjusted R2** |
| ARIMA(0,1,1) | 0.1364 | 226.02 | 227.19 | 0.225 |
| ARIMA(1,1,1) | 0.2793 | 224.69 | 224.78 | 0.300 |
| ARIMA(0,1,2) | 0.0321 | 220.49 | 222.27 | 0.305 |
| ARIMA(2,1,2) | 0.0267 | 219.57 | 220.34 | 0.471 |

ARIMA: Autoregressive Integrated Moving Average SSE: Sum of Squares Error

AIC: Akaike Information Criterion BIC: Bayesian Information Criterion

From the Above table we see that for the model ARIMA(2,1,2) ; SSE,AIC and BIC value are smaller than other model and adjusted R2 is high .So the modelARIMA (2,1,2) is the best model and we use this model for our forecasting purpose.

**Chapter: Five**

**Forecasting**

**5.1 Forecasting with ARIMA model of Dengue Cases**

**5.2 Forecasting with ARIMA model of Dengue Death**

**5.3 Forecasting with SARIMA model of Dengue Cases**

**5.4 Forecasting with SARIMA model of Dengue Death**

**5.5 Discussion**

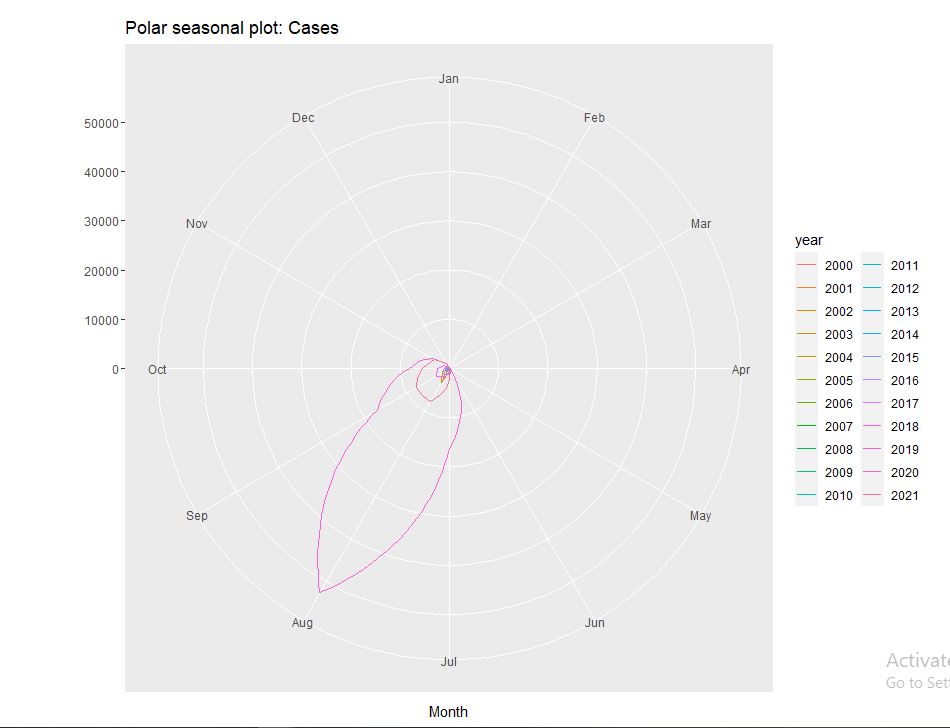
The goal of this study was to determine the trend of dengue outbreaks and make a short‐term forecast. Though the prevalence of dengue was not taken as a serious consideration before the outbreak was found in 2019. However, the outbreak creates potential risk to people in Bangladesh. For forecasting, statistical approach such as Box–Jenkins technique is normally one of the preferable approaches instead of system dynamics (SD) model as this model shows the seasonality‐periodic fluctuations for studies. However these models have some limitations in terms of feedback process. For this reason, this study aims to forecast dengue incidence using ARIMA model to forecast algorithm based on the assumption that previous values of a time series can also be used to forecast values also other researchers used this model for forecasting purpose. Previous studies [[6](#_ENREF_6)]also highlighted that the ARIMA models performed better than other models to forecast the dengue incidences in different countries which is consistent with our study and reported that dengue fever outbreaks are more likely to occur during hot, dry weather with high daily temperatures. Here, AIC, and BIC are employed to select the model and a previous study also used these criteria to select the most suitable ARIMA model for forecast purpose.

**Chapter: Six**

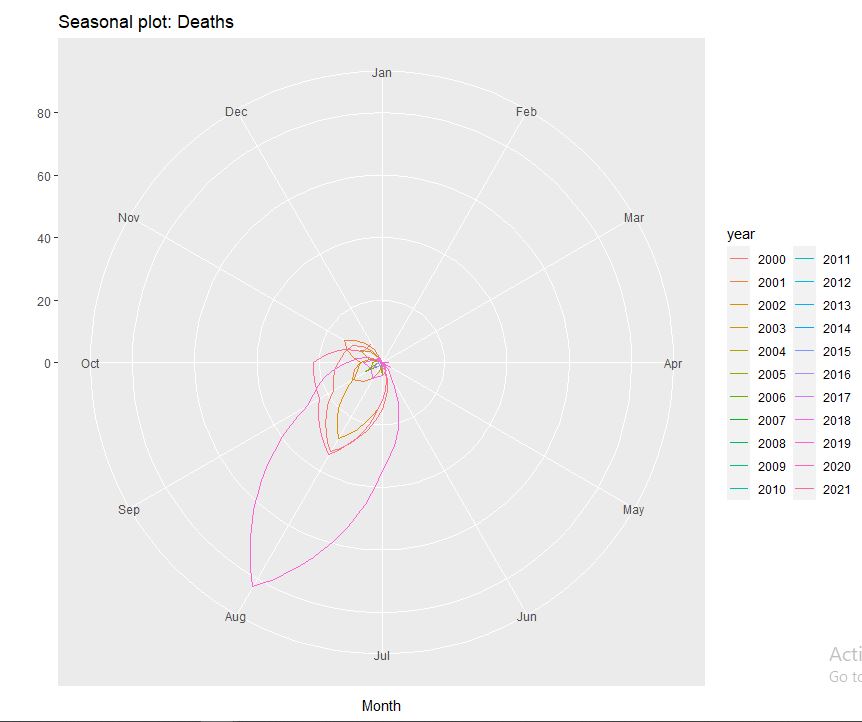
**Result and Conclusion**

**6.1 Result**

The variable “dengue” is the sum of all dengue infected patients all over Bangladesh. The authors plot the dengue‐infected patients by months all over the study period. The results depict that the dengue incidence was more frequent in August, September, and October compared to all other months. This indicates that the seasonality may be present in the data set. Thus, the authors considered seasonal ARIMA model and make a comparison with the ARIMA model From Figure , it is seen that an epidemic of dengue attacks was held in 2019. Dengue‐infected numbers have increased every year varying slightly, but in 2019 it was an extreme value. First, we verify the stationarity condition of the data series. This is due to the fact that the ARIMA model (which is a linear regression model) works best when the perdition is uncorrelated and independent to each other.



**Fig:** Polar seasonal plot of dengue cases in Bangladesh 2000-2021

****

**Fig:** Polar seasonal plot of dengue death in Bangladesh 2000-2021

We can see that there is a seasonality in dengue Cases and death in Bangladesh from 2000-2021, which is August, September, and October compared to all other months. It also clear that an epidemic of dengue attacks held on August month which is extremely in 2019.

**6.2 Conclusion**

We develop a model that may predict dengue incidences in Bangladesh. The results of the model selection criteria considered in this study depict that ARIMA (2,1,2) is the most suitable model for fitting dengue data and forecasting the future dengue scenario in Bangladesh. The findings depict that the incidence of dengue in Bangladesh is expected to rise in the future. Therefore, in addition to existing policies, government officials, nongovernmental organizations, and policymakers must take nationwide activities, and public awareness should be raised through community education campaigns to tackle the upcoming challenges due to dengue prevalence. This forecasting result helps us to get an indication of the future number of dengue infected people, enable us to assist public health policymakers to forecast dengue outbreaks and prepare preventive measures, and take suitable policy and strategies to control dengue outbreak in future prospect in Bangladesh.

**6.3 Recommendation for future study**

To improve the study design, researchers should be attentive to the overestimated self report data Also we should look out for the best statistical modeling for this kind of small data set. Although it seems as though this current study coincides with findings from past research, some limitations and assumptions did exist by the researcher, and suggestions for future research should be discussed. The sample size that was selected may be smaller than whist is preferred, however the researcher assumed it to be representative of the entire population, a larger random sample size should be suggested for future research. A few limitations of the study include the possible lack of random sampling and self reported surveys In research random sampling is important in order for everyone in the population to have an equal chance of being included in the study . While the researcher of this study tried to have a random sample this was complicated to perform with limited resources Also, because the survey was conducted in a relatively small geographic area, some caution should be taken in generalizing the findings This information, although not done so by the researcher, would be interesting to lest in order to see if this clinical test also had a significant effect on their academic performance.

**1. Haider, N., et al., *Dengue outbreaks in Bangladesh: Historic epidemic patterns suggest earlier mosquito control intervention in the transmission season could reduce the monthly growth factor and extent of epidemics.* Current research in parasitology & vector-borne diseases, 2021. 1: p. 100063. DOI: https://doi.org/10.1016/j.crpvbd.2021.100063.**

**2. Rahman, M.S., A.H. Chowdhury, and M. Amrin, *Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh.* PLOS Global Public Health, 2022. 2(5): p. e0000495. DOI: org/10.1371/journal.pgph.0000495.**

**3. Naher, S., et al., *Forecasting the incidence of dengue in Bangladesh—Application of time series model.* Health Science Reports, 2022. 5(4): p. e666. DOI: https://doi.org/10.1002/hsr2.666.**

**4. Choudhury, Z.M., S. Banu, and A.M. Islam, *Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis.* 2008. DOI: https://apps.who.int/iris/handle/10665/170465.**

**5. Nayak, M.S.D.P. and K. Narayan, *Forecasting dengue fever incidence using ARIMA analysis.* International Journal of Collaborative Research on Internal Medicine & Public Health, 2019. 11(6): p. 924-932. DOI: https://**[**www.iomcworld.org/articles/forecasting-dengue-fever-incidence-using-arima-analysis-44475.html**](http://www.iomcworld.org/articles/forecasting-dengue-fever-incidence-using-arima-analysis-44475.html)**.**

**6. Hossain, M. and F. Abdulla, *Jute production in Bangladesh: a time series analysis.* Journal of Mathematics and Statistics, 2015. 11(3): p. 93-98. DOI: 10.3844/jmssp.2015.93.98.**