

FORECASTING OF DENGUE CASES IN CHENNAI USING STATISTICAL APPROACH

Submitted in partial fulfillment of the requirements for the award of
Master of Science degree in Mathematics

By

B.SHREEVIJAYALAKSHMI
(42630007)



**DEPARTMENT OF MATHEMATICS
SCHOOL OF SCIENCE AND HUMANITIES**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A++" by NAAC

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI-600119**

APRIL - 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A++" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF MATHEMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **B.SHREEVIJAYALAKSHMI (42630007)** who carried out the project entitled **"Forecasting Of Dengue Cases In Chennai Using Statistical Approach"** under our supervision from **November 2023 to April 2024.**

Internal Guide
DR.M.MALLIKA

Head of the Department
DR.M.NIRMALA

Submitted for Viva voce Examination held on April 20-2024

Internal Examiner

External Examiner

DECLARATION

I **B.SHREEVIJAYALAKSHMI (42630007)** hereby declare that the Project Report entitled **“FORECASTING OF DENGUE CASES IN CHENNAI USING STATISTICAL APPROACH ”** done by me under the guidance of **Dr.M.MALLIKA** at **Sathyabama Institute of Science and Technology** is submitted in partial fulfillment of the requirements for the award of Master of Science degree in Mathematics.

DATE : 20-04-2024

PLACE : CHENNAI



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to The Board of Management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKRAVARTHI**, Dean, School of Science and Humanities and **Dr. M. NIRMALA** , Head of the department, Dept. of Mathematics for providing me necessary support during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. M .MALLIKA**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Mathematics who were helpful in many ways for the completion of the project.

ABSTRACT:

Dengue (break-bone fever) is a viral infection that spreads from mosquitoes to people. It is more common in tropical and subtropical climates. Most people who get dengue won't have symptoms. But for those that do, the most common symptoms are high fever, headache, body aches, nausea and rash. The study compares the Moving Average model and Exponential Smoothing model in terms of forecasting the monthly Dengue cases in Chennai from January 2000 to December 2019. The models were used to analyze data and evaluated based on the MAPE values. By comparing the error measures of the models Exponential smoothing model is found to be best model that fits the data. Therefore the Exponential smoothing model is used to predict the dengue cases in Chennai from January 2020 to December 2020.

KEYWORDS:

Dengue cases prediction, Time series analysis, Moving Average, Exponential Smoothing, Mean Absolute Percentage Error.

CONTENTS

CHAPTER NO	TITLE	PAGE NO
1	INTRODUCTION	
	1.1.TIME SERIES	1
	1.2.TIME SERIES ANALYSIS	1
	1.3.COMPONENTS OF TIME SERIES	2
	1.4.FORECASTING	2
	1.5.DENGUE	3
	1.5.1.Symptoms	3
	1.5.2.Treatment	3
2	LITERATURE SURVEY	4
3	AIM AND SCOPE	5
4	METHODOLOGY	
	4.1.MOVING AVERAGE	6
	4.2.EXPONENTIAL SMOOTHING	7
	4.3.ERROR MEASURE	8
	4.4.MEAN ABSOLUTE PERCENTAGE ERROR(MAPE)	8
5	RESULT AND DISCUSSION	9
6	SUMMARY AND CONCLUSION	14
	REFERENCES	15

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
5.1	Comparison of MAPE value for the 2 models.	12
5.2	Forecast value of dengue cases for the year 2020.	13

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
5.1	Average dengue cases of every month from the year 1999 to the year 2019.	9
5.2	Average dengue cases using 2 Month Moving average and its seasonally adjusted forecast value.	10
5.3	Average dengue cases using Exponential Smoothing ($\alpha = 0.1$) and its seasonally adjusted forecast value.	11

LIST OF ABBREVIATIONS

DHF	- Dengue Hemorrhagic Fever
GIS	- Geographical Information System
MAPE	- Mean Absolute Percentage Error
NVBDCP	- National Vector Borne Disease Control Programme

CHAPTER - 1

INTRODUCTION

1.1 TIME SERIES

Time Series is defined as the set of data collected according to time of assurance. It consists of data arranged in chronological order. When quantitative data are arranged in the order of their occurrence then the resulting series is known as Time Series.

EXAMPLE:

- Network data
- Economic Indicators
- Patient health evolution metrics

1.2. TIME SERIES ANALYSIS

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time.

Time series are analyzed to discover past patterns of variability that can be used to forecast future values.

Time series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data

Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Industry forecasts

1.3.COMPONENTS OF TIME SERIES

The two main components of time series are Trend and Seasonality. To do the forecast, seasonality and trend should be removed to get accurate result.

Trend: The trend shows the general tendency of the data to increase or decrease during a long period of time. A trend is a smooth, general, long-term, average tendency. It is not always necessary that the increase or decrease is in the same direction throughout the given period of time.

Example: Gradual increase in the Earth's average temperature with time due to climate change.

Seasonality: Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year.

Seasonality can occur on different time spans such as daily, weekly, monthly, or yearly.

Example: an increase in water consumption in summer due to warmer weather.

1.4. FORECASTING

The process of predicting future conditions that will influence in the performance of an organization is known as forecasting. It is also a process of analysing time series data using statistics and modelling to make predictions and inform strategic decision making.

1.5. DENGUE

- Dengue is a viral disease
- It is transmitted by the infective bite of *Aedes aegypti* and *Aedes albopictus* mosquitoes
- Man develops disease after 5-6 days of being bitten by an infective mosquito

1.5.1. SYMPTOMS

Mild symptoms of dengue can be confused with other illnesses that cause fever, aches and pains, or a rash.

The most common symptom of dengue is fever with any of the following:

- Nausea, vomiting
- Rash
- Aches and pains (eye pain, typically behind the eyes, muscle, joint, or bone pain)

Symptoms of dengue typically last 2–7 days. Most people will recover after about a week.

1.5.2. TREATMENT

No drug or vaccine is available for the treatment of Dengue/ DHF

In dengue shock syndrome, the following treatment is recommended:

- Replacement of plasma losses
- Correction of electrolyte and metabolic disturbances
- Blood transfusion

With early detection and proper case management and symptomatic treatment, mortality can be reduced substantially.

CHAPTER - 2

LITERATURE SURVEY

Karuppasamy, Ramanathan et al., in their paper discussed the Pathogenesis and Epidemiology using Statistical Analysis of State wise Dengue Data in India. Also, a natural therapy, in the form of papaya leaf juice has been identified and discussed in the paper. Mutations found in the virus have been discussed, [6]. Kumar, M., & Mala, S. studied the Application of Geographical Information System (GIS) and Statistical Modelling for Dengue Fever Surveillance in Delhi India. The work is aimed at using GIS techniques such as Point Density method and Empirical Bayesian Kriging method to produce optimal spatial distribution and prediction of Dengue cases in Delhi using historical Dengue Fever case counts and their geographic locations [7]. Olmoguez, I. L. G., Catindig, M. A. C et al., studied to predict the probability of dengue incidence in Iligan City per barangay based on the relationship of climatic factors and dengue cases using different predictive models with data from 2008 to 2017 and have used the methods Poisson Regression, Multiple Linear Regression and Random Forest. [9]. Vaddadi Srinivas, Vaddadi Radha Srinivas discussed the Journal of Evolution of Medical and Dental Sciences and aimed to control the aforementioned causes to prevent disease spread and reduce epidemic flare up. [13]. V. G. Ramachandran and et al., studied the Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity. This retrospective study was planned to investigate whether climatic factors could be used to predict the occurrence of dengue in East Delhi. [10]

In this research article two models, 2 monthly moving average and exponential smoothing ($\alpha = 0.1$) are compared to forecast the monthly dengue cases in Chennai city for the year 2020.

CHAPTER - 3

AIM AND SCOPE

The objective of this research project is to obtain moving average and exponential smoothing model for forecasting the dengue cases of Chennai city for the year 2020

CHAPTER - 4

METHODOLOGY

4.1. MOVING AVERAGE

A moving average is a series of averages, calculated from historic data. Moving averages can be calculated for any number of time periods, for example a three-month moving average, a seven-day moving average, or a four-quarter moving average.

Moving averages method is used in statistics to analyze data points, which are calculated by averaging several subsets of a larger dataset. A moving average is a measure of how well a piece of work is doing over a given period of time.

Moving Average can be calculated for the Time period 2, 3, 4....n.

FORMULA

The forecast value using Simple Moving Average is given by,

$$F = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}}{n}$$

Where,

- F_t = Forecast for the period 't'.
- n = Number of terms to be averaged.
- $A_{t-1}, A_{t-2}, A_{t-3} \dots$ are the actual occurrences in the past period.
two period ago, three period ago and so on respectively .

4.2. EXPONENTIAL SMOOTHING

Exponential Smoothing is a popular method used for forecasting Time series data. It is a Statistical technique that involves smoothing a time series by giving more weight to recent observations while gradually decreasing the weights of older observations.

In Exponential Smoothing, the forecast for the next period is weighted average of past observations, where the weights decrease exponentially as the observations get older. The method works on the principle that more recent data points are more relevant in predicting future trends than older ones

FORMULA

Formally, the Exponential Smoothing equation is

$$F_{t+1} = \alpha y_t + (1-\alpha) F_t$$

F_{t+1} = Forecast for the next Period.

α = Smoothing constant.

y_t = Observed value of series in period t.

F_t = Old forecast for period.

The forecast F_{t+1} is based on weighting the most recent observation y_t with a weight α and weighting the most recent forecast F_t with a weight of $1-\alpha$.

The value of smoothing constant α must be between 0 and 1.

α cannot be equal to 0 or 1

4.3. ERROR MEASURE

Error Measure is a metric that defines the actual quantity and how it deviates from the forecasted quantity. Error Measure focuses on the magnitude of the error rather than whether it is positive or negative.

4.4. MEAN ABSOLUTE PERCENTAGE ERROR(MAPE)

The Mean Absolute Percentage Error (MAPE) is also called the Mean Absolute Percentage Deviation (MAPD). It measures accuracy of a forecast system. It represents the average of the absolute percentage errors of each entry in the dataset to calculate how accurate the forecasted quantities were in comparison with the actual quantities.

MAPE is calculated using following formula:

$$MAPE = \frac{1}{n} \sum \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Where,

- n = sample size
- A_t = Actual value
- F_t = Forecast value

CHAPTER - 5

RESULTS AND DISCUSSION

The monthly data set for the dengue cases in Chennai city from January 1999 to December 2019 was considered for this project work. This data is collected from the Vector Control Department of Chennai City Corporation, the Municipal body of the State government of Tamil Nadu. The average of every month for all the years is calculated and these values are plotted and are shown below.

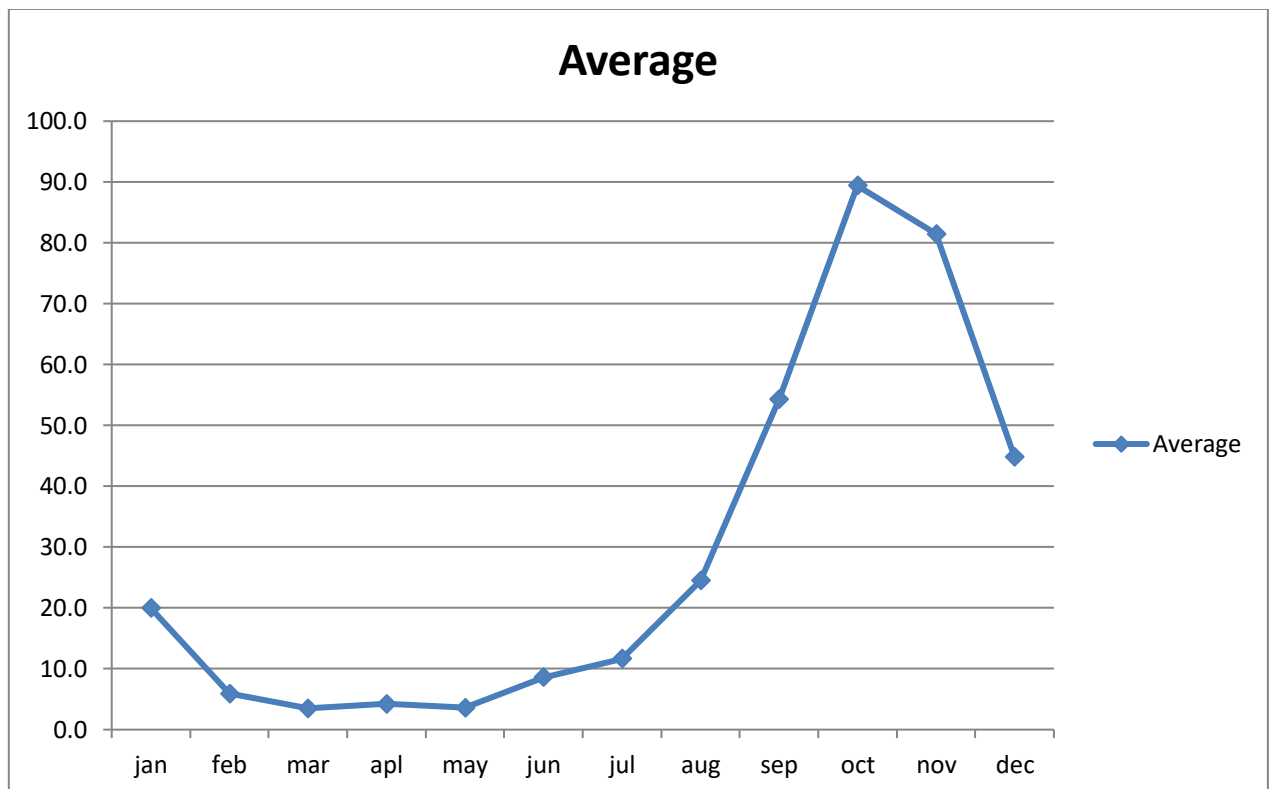


Fig.5.1: Average dengue cases of every month from the year 1999 to the year 2019.

The graph shown above depicts that there is seasonality for the given data. Hence seasonally adjusted forecast values should be evaluated. For doing this process, firstly a trend line of the form $Y = AX + B$ where A =Slope and B =Intercept is fitted for finding the total number of dengue cases for the year 2019. Secondly, seasonal factor is calculated for every month for the year 2019 by dividing monthly total by the overall total. Overall total is the overall number of dengue cases from January 1999 to December 2018.

Finally the seasonally adjusted forecast values for every month for the year 2019 is calculated by multiplying the seasonal factor of every month by the total.

In this research work, 2 month moving average, 3 month moving average.....10 month moving average are applied to forecast values of the given data. Among them 2 month moving average gave the least MAPE value. So with these 2 month moving average forecast values, seasonally adjusted forecast values are found for every month for the year 2019 using the procedure discussed above.

The graph below shows the comparison of every month for the year 2019 using simply 2 month moving average values against seasonally adjusted forecast values using 2 month moving average values.

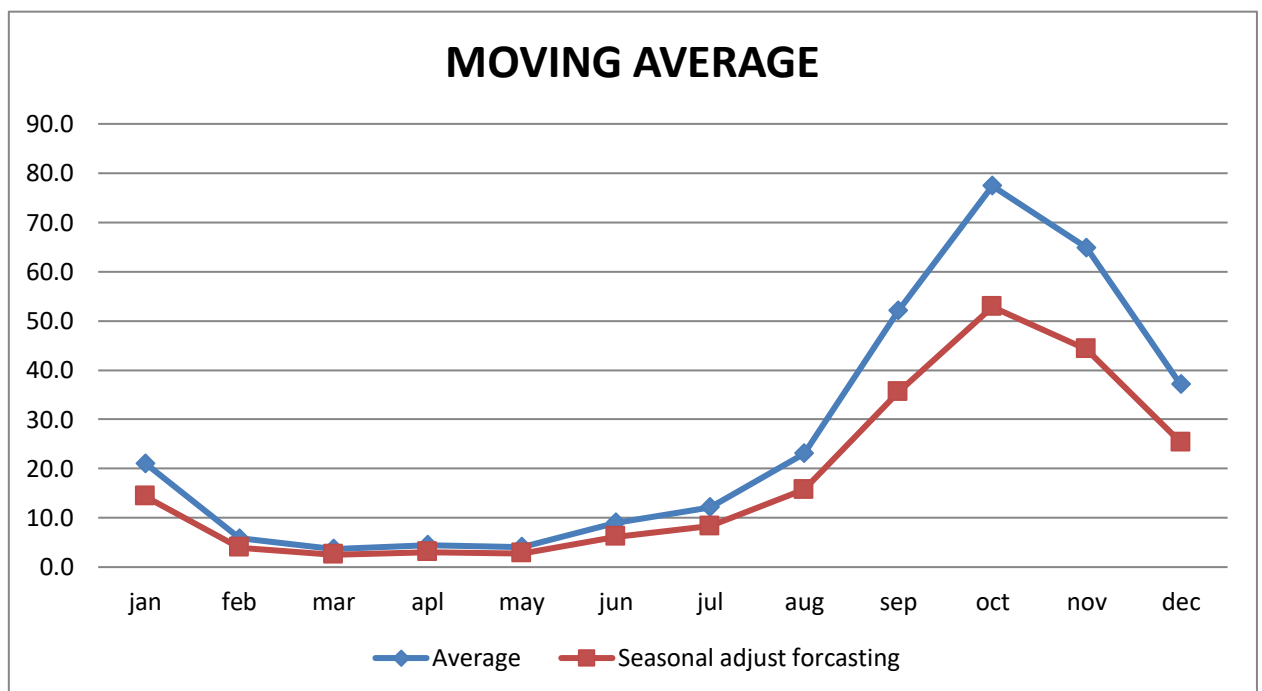


Fig.5.2: Average dengue cases using 2 month moving average and its seasonally adjusted forecast value.

Also for the given data, exponential smoothing for $\alpha = 0.1, 0.2 \dots 0.9$ are applied to forecast values of the given data. Among them $\alpha = 0.1$ gave the least MAPE value. Therefore, using the forecast values by exponential smoothing for $\alpha = 0.1$, seasonally adjusted forecast values are found for every month for the year 2019 using the same procedure done for moving average.

The graph below shows the comparison of every month for the year 2019 using simply exponential smoothing values for $\alpha = 0.1$ against seasonally adjusted forecast values using exponential smoothing values for $\alpha = 0.1$.

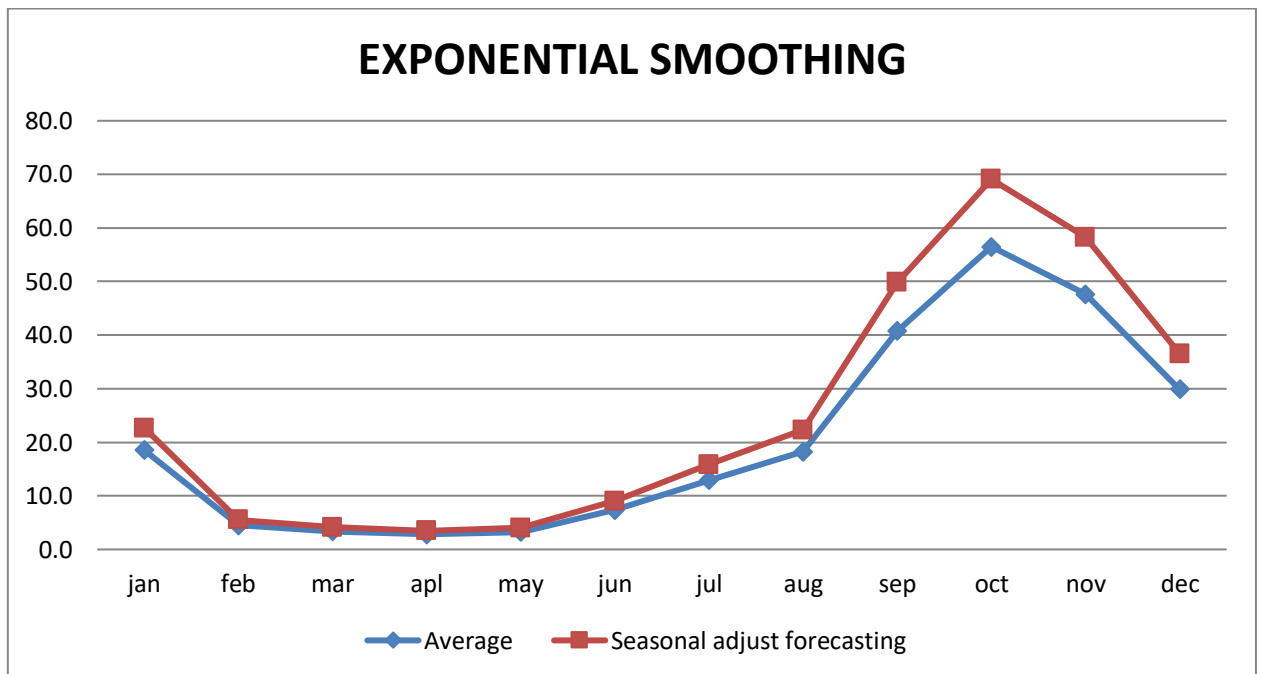


Fig.5.3: Average dengue cases using Exponential Smoothing ($\alpha = 0.1$) and its seasonally adjusted forecast value.

The seasonally adjusted forecast values using 2 month moving average and exponential smoothing for $\alpha = 0.1$ are compared and their MAPE value are obtained and are shown in the table below.

Table. 5.1: Comparison of MAPE value for the 2 models.

ERROR MEASURE	2 MONTHLY MOVING AVERAGE	EXPONENTIAL SMOOTHING($\alpha=0.1$)
MAPE	68.01	54.9

From the above table we find the MAPE value corresponding to the exponential smoothing for $\alpha = 0.1$ gave the least MAPE value. Hence forecast for the Dengue cases for every month for the year 2020 are done using exponential smoothing model for $\alpha = 0.1$. The values thus calculated are shown in the table below.

Table.5.2: Forecast value of dengue cases for the year 2020.

MONTH	CASES
JANUARY	23
FEBRUARY	6
MARCH	4
APRIL	4
MAY	4
JUNE	9
JULY	16
AUGUST	23
SEPTEMBER	51
OCTOBER	70
NOVEMBER	59
DECEMBER	37

CHAPTER - 6

SUMMARY AND CONCLUSION

Dengue is one of the most dreadful disease the world faces. This research work compares the Moving Average model and Exponential Smoothing model in terms of forecasting the monthly Dengue cases in Chennai from January 2000 to December 2019. The models were used to analyze data and evaluated based on the MAPE values. By comparing the error measures of the models, exponential smoothing model is found to be best model that fits the data. Therefore, Exponential smoothing model for $\alpha = 0.1$ is used to predict the dengue cases in Chennai from January 2020 to December 2020. By doing the forecast one can plan for the steps to be taken to reduce the number of cases in the future as the saying goes "Prevention is better than cure".

REFERENCES:

- [1] Bhatnagar S, Lal V, Gupta S.D, Gupta O.P, "Forecasting incidence of dengue in Rajasthan using time series analyses", Indian Journal Public Health, vol 56,2012, 281-5.
- [2] Box and Jenkins –"Time Series Analysis- Forecasting and Control" (5th edition), John Wiley and sons ,2016.
- [3] Choudhury M.A, Banu S, Islam M.A, "Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis",Dengue Bulletin, vol 32,2008, 29-37.
- [4] Chris Chatfield – "The Analysis of Time Series: An introduction", (6th edition), Taylor and Francis e-Library ,2009.
- [5] <https://tnhealth.tn.gov.in/tngovin/dph/dphpm.php#>.
- [6] Karuppasamy, Ramanathan & Shreenidhi, R. & Vidya V.S &Veerappapillai, Shanthi, "Statistical Analysis of Statewise Dengue Data in India",Research Journal of Pharmaceutical, Biological and Chemical Sciences, vol 6,2015, 1557-1567.
- [7] Kumar. M., & Mala.S, "Application of GIS and Statistical Modelling for Dengue Fever Surveillance in Delhi India", Fourth International Conference on Advances in Civil, 2016.

- [8] Lal V, Gupta S, Gupta O, Bhatnagar S, “ Forecasting incidence of dengue in Rajasthan, using time series analyses”,Indian Journal Public Health,Vol 56, 2012, 281-285.
- [9] Olmoguez I.L.G., Catindig, M.A.C., Amongos, M.F.L., & Lazan, F.G, “Developing a dengue forecasting model: a case study in Iligan City”, International Journal of Advanced Computer Science and Applications, vol 10 Issue 9,2019.
- [10] Ramachandran V.G and e. al, "Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity: a 19-year retrospective analysis in East Delhi," Epidemiology and health, vol 38, 2016.
- [11] Silawan T, Singhasivanon P, Kaewkungwal J, Nimmanitya S, Suwonkerd W, “Temporal patterns and forecast of dengue infection in northeastern Thailand”, Southeast Asian Journal Trop Med Public Health,Vol 39 ,2008,39.
- [12] Siva Durga Prasad Nayak. M and Narayan K.M, “Forecasting Dengue Fever Incidence Using ARIMA Analysis”, International Journal of Collaborative Research on Internal Medicine & Public Health ,Vol 11 No. 3 ,2019.
- [13] Vaddadi Srinivas, Vaddadi Radha Srinivas. “Dengue Fever: A Review Article”, Journal of Evolution of Medical and Dental Sciences, 2015; Vol. 4, Issue 29 Page: 5048-5058.