Project Report

On

DENGUE AND LEPTOSPIROSIS IN KERALA: ANALYSIS AND PREDICTION

Submitted in partial fulfilment of the requirements for the degree of

BACHELOR OF SCIENCE

in

MATHEMATICS

by

NAVYA UNNI- AB21BMAT043

Under the Supervision of

DR. SUSAN MATHEW PANAKKAL

DEPARTMENT OF MATHEMATICS

ST TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI, 682011

APRIL 2024

ST TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **DENGUE AND LEPTOSPIROSIS IN KERALA: ANALYSIS AND PREDICTION** is a bonafide record of the work done by **NAVYA UNNI** under my guidance as partial fulfilment of the award of the degree of Bachelor of Science in Mathematics and Statistics at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date: 18-03-2024

Place: Ernakulam

DR. SUSAN MATHEW PANAKKAL

Assistant Professor,

Department of Mathematics,

St. Teresa's College (Autonomous),

Ernakulam.

DR. URSALA PAUL

Associate Professor and HOD,

Department of Mathematics

St. Teresa's College (Autonomous),

Ernakulam.

External Examiners

1.

2.

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of SUSAN MATHEW PANAKKAL, Assistant Professor, Department of Mathematics and Statistics, St. Teresa's College (Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Ernakulam

Date: 18-03-2024

NAVYA UNNI-AB21BMAT043

ACKNOWLEDGEMENT

I extend my heartfelt appreciation to Mrs. DR. SUSAN MATHEW PANAKKAL, Assistant Professor at St. Teresa’s College, for her invaluable coordination, guiding us through the project's stages, and to my group members for their engaging defence and insightful contributions. Gratitude goes to Dr. Ursala Paul, Dr. Susan Mathew Panakkal, and divine guidance for steering me through challenges, culminating in the completion of this project.

Ernakulam

Date: 18-03-2024

NAVYA UNNI-AB21BMAT043

CONTENTS

Chapter 1

INTRODUCTION

1.1 DENGUE

Dengue, a virus transmitted by mosquitoes, is the predominant arthropod-borne viral disease globally. Referred to as breakbone fever due to intense muscle spasms, it is also known as dandy fever or seven-day fever, reflecting the typical symptom duration. Despite most cases being asymptomatic, severe outcomes, including death, can occur. Aedes mosquitoes, prevalent in tropical regions, transmit the virus. Dengue's incidence has surged in recent decades, with some areas now endemic. In some instances, individuals previously infected with one dengue virus subspecies may experience severe capillary issues and bleeding if infected with another, leading to dengue haemorrhagic fever. This overview explores dengue's pathophysiology, symptomatic presentation, management, and emphasizes interprofessional collaboration for effective care coordination and communication, crucial for prompt diagnosis and improved patient outcomes.

Home treatment with pain medication is typically sufficient for most dengue fever cases, emphasizing the importance of preventing mosquito bites as the primary preventive measure. Dengue lacks a specific cure, with pain symptom management being the primary focus. Acetaminophen is commonly used for pain control, while avoiding non-steroidal anti-inflammatory drugs like ibuprofen and aspirin due to their potential to increase bleeding risk. Dengvaxia, a vaccine, is available for individuals who have experienced dengue and reside in areas where the disease is prevalent. Severe dengue cases often require hospitalization. [1]

1.1.1 DENGUE CONDITION IN KERALA

Dengue cases, including fatalities, were first reported in Kerala in 1997. However, antibodies and viruses associated with dengue had been identified in the state since

1979. Over the years, Kerala's dengue-related deaths increased, with the state becoming hyper-endemic, harbouring all four serotypes. Thiruvananthapuram district reported the highest cases, especially in urban areas. Despite global vector control efforts, dengue remains a significant cause of mortality. While no vaccine or antiviral treatment exists, improved clinical management has reduced severe dengue mortality. The WHO aims to cut dengue mortality by 50%, emphasizing early case detection and effective management. This study focused on identifying mortality risk factors in dengue patients at the Medical College Hospital, Thiruvananthapuram, from 2005 to 2008. [2]

1.1.2 ABOUT CLIMATIC FACTORS IN KERALA

Kerala has witnessed a decrease in annual and monsoon rainfall alongside an increase in temperature over the past decades. According to a study by the India Meteorological Department, the mean annual maximum temperature in Kerala rose by 0.8 degrees Celsius, the minimum temperature by 0.2 degrees Celsius, and the average temperature by 0.5 degrees Celsius between 1961 and 2003. The state experiences rainfall in four spells: winter (January-February), pre-monsoon (March-May), monsoon (June-September), and post-monsoon (October-December). Rainfall is scarce in the pre-monsoon period, whereas the southwest monsoon, occurring from June to September, contributes to heavy rainfall. The northeast monsoon, starting in October and lasting until December, is marked by moderate rainfall. Of the total annual rainfall (945 mm) in the state, 48% is received during the southwest monsoon, 32% during the northeast monsoon, and the remaining percentage during other seasons.

Kerala faces a high number of dengue cases, possibly due to the abundance of breeding grounds, a higher percentage of infected mosquitoes, and favourable temperature ranges (23.5–30 °C) with short incubation periods throughout the year (9–14 days), including the rainy season. These temperature ranges are conducive to mosquito development and virus transmission. In areas with lower temperatures within the dengue distribution range (~17–18 °C), disease transmission is limited due to its impact on the extrinsic incubation period (EIP). Conversely, high temperatures (~35 °C, depending on the vector species) can decrease disease risk by limiting mosquito survival. Consequently, future climate change might further impact the burden of dengue and other vector-borne diseases.

The connection between the spread and prevalence of dengue in Kerala and climatic factors is transmission of Aedes mosquitoes responsible for the disease. Temperature plays a crucial role, with warm and humid conditions in Kerala providing an ideal breeding ground for Aedes mosquitoes. The consistent high temperatures contribute to year-round mosquito activity, facilitating continuous dengue transmission. Additionally, heavy rainfall during the monsoon season, combined with elevated humidity, creates favourable conditions for mosquito breeding and, consequently, an increased risk of dengue transmission. Seasonal variations, influenced by climatic patterns, contribute to the surge in dengue cases during the post-monsoon and pre-monsoon periods. Understanding these variations is essential for implementing targeted preventive measures during high-risk periods. Moreover, the impact of climate change on dengue transmission emphasizes the need for adaptive strategies by public health authorities to mitigate the evolving climatic conditions affecting disease dynamics. [3]

1.2 LEPTOSPIROSIS

Leptospirosis, caused by various species of the Leptospira genus, is a bacterial infection prevalent in tropical and subtropical regions. The bacteria, found in the urine of infected animals, especially rodents, can be transmitted to humans through contact with contaminated water, soil, or food. Common carriers include not only rodents but also animals like dogs, cattle, pigs, and wildlife. The disease manifests with a range of symptoms, from mild to severe, including fever, chills, muscle aches, headache, and jaundice. In severe cases, organ failure can occur, posing a risk of fatality if not promptly treated. Diagnosis relies on laboratory tests like serological tests and polymerase chain reaction (PCR). Seeking medical attention is crucial, particularly if symptoms are present, and there has been potential exposure to contaminated environments.

Treatment involves antibiotics such as doxycycline or penicillin, with early intervention essential to prevent complications and reduce the severity of the illness. Preventive measures encompass avoiding contact with potentially contaminated elements, wearing protective clothing in high-risk areas, and maintaining good hygiene. Vaccines are available for specific high-risk groups like farmers and military personnel.

Leptospirosis has a global distribution but is more common in tropical and subtropical regions. Outbreaks may occur following heavy rainfall or flooding, as these events heighten the risk of exposure to contaminated water. Timely medical attention, combined with preventive measures, is crucial for managing and mitigating the impact of leptospirosis. [4]

1.2.1 LEPTOSPIROSIS CONDITION IN KERALA

Leptospirosis has long been a major threat to the State of Kerala with more than 1,000 cases is being reported annually. Nationally, it causes the highest number of deaths among all communicable diseases in the state of Kerala. At least 100 deaths were reported yearly in Kerala before 2010. In 2006, there were 1,821 cases of rat fever of which 104 (5.7%) died and in 2007 there were 1,359 cases with 229 (16.9%) deaths. The number of leptospirosis cases in 2008, 2009 and 2010 were 1305, 1237 and 1016 with mortality rate of 136 (10.4%), 107 (8.6%) and 85 (8.4%), respectively [8]. In 2011 and 2012, the number of confirmed cases was 944 and 736 with death rate of 70 (7.4%) and 18 (2.4%). It has been also reported that in 2013 and 2014, confirmed cases was 814 and 717 with 34 (4.2%) and 19 (2.6%) deaths respectively [9]. Notably, the incidence and mortality of leptospirosis in Kerala for the following years showed a declining trend as compared to the previous years. In 2015, 43 people died of rat fever and in the subsequent years the death toll was found to be 35 and 80 in 2016 and 2017, respectively. From the month of January till July 2018 (before the flood), 28 deaths were reported due to leptospirosis in Kerala. [5]

1.2.2 ABOUT CLIMATIC FACTORS IN KERALA

Leptospirosis, caused by the spirochete bacteria Leptospira, is a zoonotic disease prevalent in the southwestern Indian state of Kerala. The region grapples with a high incidence of leptospirosis, and its prevalence and transmission are significantly influenced by climatic factors. Understanding how these factors contribute to the disease's incidence is imperative due to Kerala's unique geographical and climatic features, creating an environment conducive to leptospirosis spread.

From June to September, Kerala witness's intense monsoon rains, providing an ideal breeding ground for Leptospira bacteria. Flooding during rainfall results in stagnant water bodies and waterlogged areas, creating a medium for bacteria multiplication. Leptospira, thriving in warm, moist conditions, are commonly found in the urine of infected animals, posing an increased risk of transmission to human. Kerala's warm and humid climate year-round supports the survival of Leptospira in the environment. High temperatures and humidity create favourable conditions for the bacteria to persist in water, soil, and vegetation. This extended survival increases the likelihood of human exposure to the bacteria.

Kerala is blessed with an extensive network of rivers, backwaters, and other water bodies. While these water resources are vital for agriculture and daily life, they also contribute to the transmission of leptospirosis. Flooding during the monsoon season not only contaminates water sources but also forces humans and animals to come into closer contact with the infected water.

Kerala's diverse ecology, encompassing dense forests and wildlife habitats, complicates leptospirosis transmission. Wild animals act as reservoir hosts for Leptospira, adding to the risk. Human contact with contaminated water sources in forested areas, coupled with deforestation and habitat encroachment, further amplifies the interaction between wildlife and human populations. [6]

1.3 OBJECTIVES

- To determine the relation of Dengue and Leptospirosis with various Climatic factors
- To predict the model of Dengue and leptospirosis with Climatic variables using Multiple Regression and Ridge Regression
- Model Seasonal Time Series data of Dengue with SARIMA

1.4 SIGNIFICANCE OF THE STUDY

This study on climatic factors influencing dengue in Kerala is crucial due to the complex link between climate and the Aedes mosquito, the disease vector. With dengue's susceptibility to climate variations and the backdrop of climate change, understanding these influences is vital. The research explores how climate acts as a catalyst for dengue development in Kerala, given the region's favourable conditions for Aedes mosquito breeding. It also addresses the consequences of climate change on vector-borne diseases, emphasizing the need for adaptive strategies. The study's practical significance lies in informing decision-makers, aiding in the formulation of preventive measures, resource allocation, and adaptive strategies to counter the evolving impact of climate change on vector-borne diseases. Additionally, it lays the foundation for proactive disease control efforts by identifying specific climatic conditions conducive to dengue transmission. This forward-looking perspective is essential in the face of a changing climate, providing a framework for sustainable and adaptive disease control strategies.

Comprehensive research on leptospirosis is crucial for global public health, given its impact on both humans and animals worldwide. A thorough understanding of the disease's prevalence and impact is essential for the development of effective prevention and control strategies. The zoonotic nature of leptospirosis highlights the interconnectedness of human, animal, and environmental health, emphasizing the need for interdisciplinary collaboration. Environmental factors, including rainfall and temperature, play a significant role in the transmission of leptospirosis, presenting challenges for disease management. The enhancement of diagnostic tools and techniques is vital for the early detection and treatment of the disease. In summary, a holistic approach to studying leptospirosis is key to addressing its complexities and advancing public health responses.

Chapter 2
LITERATURE REVIEW

- Md. Nazmul Karim, Saif Ullah Munshi, Nazneen Anwar & Md. Shah Alam "Climatic factors influencing Dengue cases in Dhaka city: a model Dengue prediction". This study was carried out to examine whether the Climatic factors data can be used to predict yearly dengue cases of Dhaka City, Bangladesh. Their results showed that the climate had a major effect on the occurrence of dengue infection in Dhaka City. Though the prediction model had some limitations in predicting the monthly number of dengue cases it could forecast possible outbreak two months in advance with considerable accuracy. [7]
- Srinivasa Rao Mutheneni, Andrew P Morse, Cyril Caminade and Suryanarayana Murty Upadhyayula, "Dengue Burden in India: recent trends and importance of climatic parameters" Dengue is driven by complex interactions among host, vector and virus that are influenced by climatic factors. This study is focused on the extrinsic incubation period (EIP) and its variability in different Climatic zones in India. These results suggest that temperature is important in virus development in climatic regions and may be useful in understanding spatio-temporal variations in dengue risk. [8]
- Juliana Lucia Duarte, Fredi Alexander Diaz-Quijano, Antônio Carlos Batista, Leandro Luiz Giatti "Climatic variables associated with dengue incidence in a city of the Western Brazilian Amazon region." This study aimed to examine the impact of climate variability on the incidence of dengue fever in the city of Rio Branco, Brazil. The relation between the monthly incidence of dengue fever and climate variables was evaluated, using generalized autoregressive moving average models with negative binomial distribution. [9]
- Micheline SZS Coelho, Eduardo Massad "The impact of climate on Leptospirosis in São Paulo, Brazil" International journal of biometeorology 56, 233-241, 2012. In this

work, they correlate the daily number of human leptospirosis cases with several climatic factors. They used a negative binomial model that considers hospital daily admissions due to leptospirosis as the dependent variable, and the climatic variables of daily precipitation pattern, and maximum and minimum temperature as independent variables. [10]

- Shivshakti D Pawar, Maruti Kore, A Athalye, PS Thombre, "Seasonality of leptospirosis and its association with rainfall and humidity in Ratnagiri, Maharashtra". The aim of the study was (1) to study the seasonal pattern of leptospirosis cases and with rainfall and relative humidity and (2) to forecast the leptospirosis cases' occurrence based on the model. Seasonal pattern of cases of leptospirosis was observed along with the correlation of rainfall. This forecasted model could be used by health administrators effectively to arrange the adequate resources on time to manage the outbreaks. [11]

Chapter 3

METHODOLOGY

3.1 DATA COLLECTION

Data for Climatic factors is collected from the Indian Meteorological Department. Data is taken from the year 2012 to 2021. This study is about the relationship between the Dengue and the various climatic factors like rainfall, Maximum Temperature, Minimum Temperature, humidity. The data is arranged in an order in which the year is divided into four seasons such as Winter (Jan-Feb), Pre-monsoon (March-May), Southwest-Monsoon (June-Sept), Post-Monsoon (Oct-Dec). Here Dengue is the dependent variable and various climatic factors are the independent variables.

3.2 ANOVA TEST

ANOVA is to test for differences among the means of the population by examining the amount of variation within each sample, relative to the amount of variation between the samples. It assesses whether the variability within groups is like the variability between groups. If the variance between groups is significantly greater than the variance within groups, it suggests that there are differences in the means. ANOVA provides an overall p-value, and if it is below a chosen significance level, you can conclude that at least one group differs from the others. [12]

3.3 MULTIPLE REGRESSION

Multiple regression is a statistical technique that analyses the relationship between multiple independent variables and a dependent variable. It extends simple linear regression, allowing for the examination of the combined impact of several predictors on an outcome. The model aims to find the best-fitting linear equation by estimating coefficients for each predictor. It is commonly used in various fields, such as economics, social sciences, and business, to understand and predict complex relationships.

$$Y=a+b_1X_1+b_2X_2+..........+b_nX_n \qquad (1)$$

Here Y is the dependent variable, and $X_1,\ldots X_n$ are the n independent variables.

In calculating a, $b_1,\ldots,b_n$, regression analysis ensures maximal prediction of the dependent variable from the set of independent variables. This is usually done by least squares estimation. This can be used to analyse multivariate time series data when one of the variables is dependent on a set of other variables. We can model the dependent variable Y on the set of independent variables. [13]

3.4 RIDGE REGRESSION

Ridge Regression is a linear regression technique that adds a regularization term to the cost function. It is designed to prevent overfitting by penalizing large coefficients. The regularization term is a multiple of the squared sum of coefficients, controlled by a hyperparameter alpha. Ridge Regression is particularly useful when there's multicollinearity among the predictors.

Assumptions of ridge regression are the same as those of linear regression; linearity, constant variables, and independence. However, as ridge regression does not provide confidence limits, which can be proved effective for overfitting. This technique makes a good fit model by adding a penalty and shrinking the beta coefficients. One of the main challenges of ridge regression is choosing the right value of alpha, the parameter that controls the amount of regularization. If alpha is too small, then the model will be complex and overfit. If the alpha is too large then the model will be too simple and underfit the data. [14]

3.5 COMPARISON OF ACCURACY OF THE MODELS FOR REGRESSION

3.5.1 Mean Absolute Error

Mean Absolute Error (MAE) is a metric used to evaluate the accuracy of a prediction model. It measures the average absolute difference between the predicted values and the actual values. The Formula for MAE is the sum of the absolute differences. [15]

$$\frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

n: number of observations

y_i: the actual value of the i^{th} observation

$\hat{y}_i$: the predicted value of the i^{th} observation

3.5.2 R^2 Score

The R^2 score or Coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variable. It ranges from 0 to 1, where 1 indicates a perfect fit.

$$R^2 = 1 - \frac{SSres}{SStot}$$

where SS_{res} is the sum of squared residuals and SS_{tot} is the total sum of squares. It is a measure of how well the model explains the variability in the dependent variable. [16]

3.6 SARIMA MODEL

Time Series Analysis deals with data collected over time, where observations are ordered chronologically. Time series analysis focuses on understanding the patterns and trends within the data over time. It assumes that there is some form of pattern or trend in the data that can be modelled and forecasted. It also assumes that observations are correlated with each other due to their temporal nature. It includes techniques like ARIMA (AutoRegressive Integrated Moving Average), Exponential Smoothing, and Seasonal Decomposition, which are specifically designed to capture patterns and trends in time series data. Feature engineering may involve lagging variables, differencing, and other transformations specific to time series data to capture seasonality, trends, and other patterns.

Seasonal AutoRegressive Integrated Moving Average (SARIMA) Model:

SARIMA is an extension of the ARIMA (AutoRegressive Integrated Moving Average) model, which is a popular time series forecasting technique. SARIMA incorporates seasonality into the ARIMA framework, making it suitable for data with clear seasonal patterns.

Components of SARIMA:

1. Seasonal Component (S): This represents the repeating pattern in the data over fixed intervals (e.g., days, months, quarters). The seasonal component captures the seasonality in the data.
2. Autoregressive Component (AR): This component models the relationship between an observation and several lagged observations (its own past values). It accounts for the dependency of the current observation on its previous values.
3. Integrated Component (I): This component accounts for the differencing of the time series data to make it stationary. Stationarity implies that the statistical properties of the time series, such as mean and variance, remain constant over time. [17]

The Seasonal AutoRegressive Integrated Moving Average (SARIMA) model is represented by the following equation.

Let y_t represent the time series data at time t.

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps})(1-B)^d(1-B^s)D_{yt} =$$
$$(1+\theta_1 B+\theta_2 B^2+\ldots+\theta_q B^q)(1+\Theta_1 B^s+\Theta_2 B^{2s}+\ldots+\Theta_Q B^{Qs})\varepsilon_t$$

B is the backshift operator.

$\phi_1,\phi_2,\ldots,\phi_p$ are the non-seasonal autoregressive parameters.

$\Phi_1,\Phi_2,\ldots,\Phi_P$ are the seasonal autoregressive parameters.

p and P are the orders of the non-seasonal and seasonal autoregressive parts, respectively.

d and D are the orders of non-seasonal and seasonal differencing, respectively.

$\theta_1,\theta_2,\ldots,\theta_q$ are the non-seasonal moving average parameters.

$\Theta_1, \Theta_2,\ldots,\Theta_Q$ are the seasonal moving average parameters.

q and Q are the orders of the non-seasonal and seasonal moving average parts, respectively. s is the seasonal period. ε_t is white noise. [18]

Chapter 4

RELATION OF DENGUE AND LEPTOSPIROSIS WITH CLIMATIC VARIABLES

4.1 TEMPORAL DISTRIBUTION OF DENGUE CASES

Over the 10-year period (2012-2021), total of 62,582 dengue cases were reported. The number of dengue cases reported varied year by year. Over the period, the highest dengue cases were reported in 2017 and the lowest in 2014. In each year, dengue cases showed a specific pattern of occurrence.

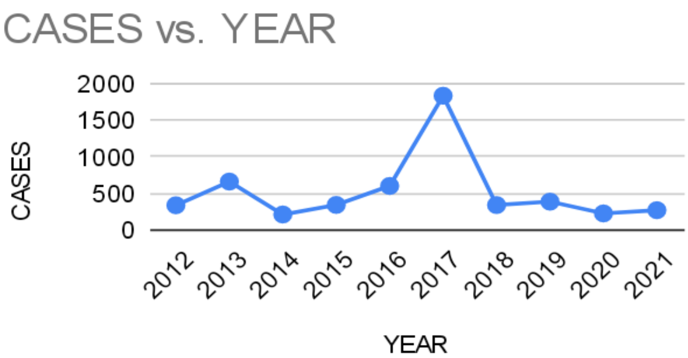


Figure 4.1 Dengue through years

Most of the cases were reported during Southwest Monsoon period (i.e., from June-September) in each year compared to other periods. It is clear from the table that the highest dengue cases were reported in the year 2017 and the lowest in 2014. During the ten years, it is in winter period that there is a little decrease in the occurrence of dengue cases. Highest number of cases were reported in the months of June and July and a significant number of cases were reported in August and gradually declined at the time of beginning and at the end of the year.

Year	Winter (Jan-Feb)	Pre-monsoon (March-May)	Southwest Monsoon (June-Sept)	Post Monsoon (Oct-Dec)
2012	178.5	121.33	543	387.67
2013	257	557.33	1290	197.33
2014	43.75	88	338.5	251.67
2015	102.5	139	643.5	306
2016	279.5	353.67	1102.5	396
2017	358	1433.33	3884.5	479.67
2018	131	195.33	701.75	145
2019	65.5	89	620.5	590.33
2020	322	163.33	342.25	73.33
2021	104	136	491.75	222.67

Figure 4.2 Seasonal Variation in number of reported Dengue cases in each year

4.2 CLIMATIC INFLUENCE

The dengue outbreak in Kerala coincided mainly with Southwest Monsoon period with heavy rainfall started from June and lasted up to September, followed by relatively less rainfall during Post-Monsoon period from October to December. The period from January to February is considered as the winter period. The period from March to May is considered as the Pre-Monsoon period.

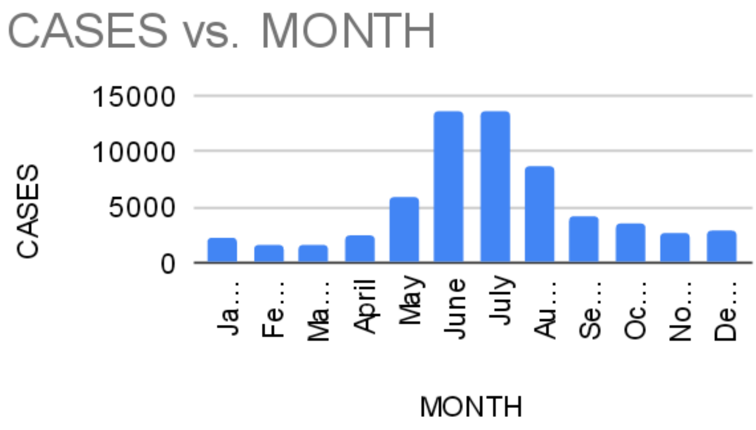


Figure 4.3 Dengue Cases through Months

The number of dengue cases was plotted against the climatic factors (Rainfall, Temperature, Relative Humidity, Mean vapour pressure, Mean wind speed) to assess their contribution across the years. As the rainfall increased, the cases of dengue also started rising from June. With declining rainfall from October to December, dengue cases also showed gradual decline. Over the years, the average relative humidity showed an increasing trend during the dengue outbreak period.

4.3 TEMPORAL DISTRIBUTION OF LEPTOSPIROSIS CASES

Over the ten years (2012-2021) period, a total of 12,915 cases of Leptospirosis were reported. The number of reported leptospirosis cases varied by year. Over the study period, the highest cases of dengue were reported in 2018 and the lowest in 2012.

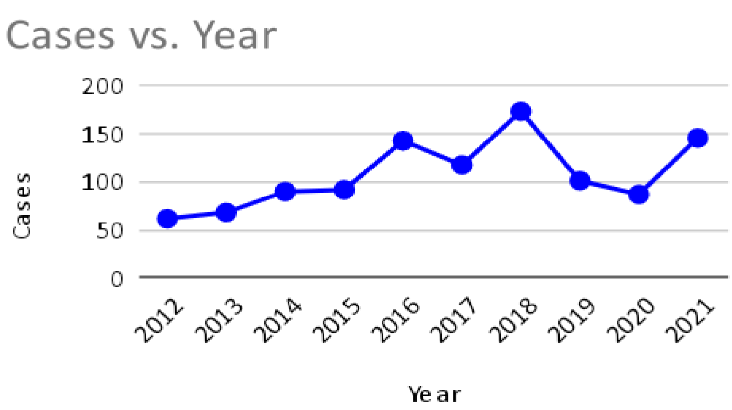


Figure 4.4 Leptospirosis cases through Years

In each year, leptospirosis cases showed a specific pattern of occurrence. Most of the cases were reported during the Southwest-Monsoon period (i.e., from June-September) in each year compared to other periods. It is clear from the table that the highest leptospirosis cases were reported in the year 2018 and the lowest in 2012. During the ten years, it is in winter period that there is a decrease in the occurrence of leptospirosis cases. Significant number of cases were reported from May to August and reached a peak in September and gradually declined at the end of the year.

Year	Winter (Jan-Feb)	Pre-Monsoon (March-May)	Southwest-Monsoon (June-Sept)	Post-Monsoon (Oct-Dec)
2012	43.5	37.33	74.75	79.33
2013	25.5	73.33	87.5	64.33
2014	24	53.67	130	115.33
2015	48.5	39	117	138.67
2016	88	100.67	216.75	121.67
2017	101	113.33	113.75	137
2018	44	46	348.75	152.67
2019	64	58	114.75	150
2020	58.5	44.33	113.5	111.67
2021	93	46.33	161	258.67

Figure 4.5 Seasonal Variation in number of reported Leptospirosis cases in each year

4.4 CLIMATIC INFLUENCE

The outbreak of leptospirosis cases in Kerala coincided mainly with the Southwest Monsoon period with heavy rainfall started from June and lasted up to September, followed by relatively less rainfall during post-monsoon period from October to

December. The period from January to February is considered as the winter period. The period from March to May is considered as the Pre-Monsoon period.

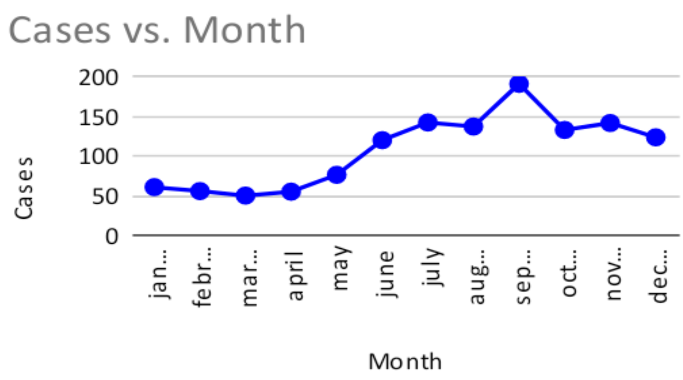


Figure 4.6 Leptospirosis Cases through Months

The number of leptospirosis cases was plotted against the climatic factors (Rainfall, Temperature, Relative Humidity, Mean vapour pressure, Mean wind speed) to assess their contribution across the years. As the rainfall increased, the cases of leptospirosis also started rising from May. With declining rainfall from October to December, leptospirosis cases also showed gradual decline. Over the years, the average relative humidity showed an increasing trend during the leptospirosis outbreak period.

4.5 ANOVA TEST FOR CLIMATIC FACTORS

The average monthly rainfall all these years during winter, pre monsoon, southwest monsoon and post monsoon were 12.982, 106.49, 392.55 and 147.03 respectively. The overall variation was significant (ANOVA $P<0.001$). Rainfall during southwest monsoon was significantly higher than the other three periods.

The average relative humidity recorded during the winter period was 61.7 per cent, 69.94 per cent during pre-monsoon ,82.3 per cent during the southwest monsoon period and 72.99 per cent during the post-monsoon period. The overall variation was statistically significant (ANOVA $P<0.001$).

Average highest and lowest temperatures were 34.1^0C and 22.60^0C, respectively during winter. During pre-monsoon and south west monsoon periods the temperature rose to 34.6 and 25.05^0C and fell to 30.9^0C and 23.69^0C, respectively. During post monsoon the maximum temperature and minimum temperature were 32.63^0C and 23.4^0CThe overall variation of both maximum and minimum temperatures across the four seasons was found to be significant ($P<0.001$).

The average mean vapour pressure recorded during the winter period was 26.32 per cent, 31.53 per cent during pre-monsoon ,31.4 per cent during the southwest monsoon period and 29.49 per cent during the post-monsoon period. The overall variation was statistically significant (ANOVA $P<0.001$).

The average mean wind speed recorded during the winter period was 6.5 per cent, 7.05 per cent during pre-monsoon ,5.93 per cent during the southwest monsoon period and 5.28 per cent during the post-monsoon period. The overall variation was statistically significant (ANOVA $P<0.001$).

Chapter 5

MODEL FORMULATION

5.1 MULTIPLE REGRESSION MODEL

The dependent variable dengue is positively skewed (±4.17347) and so is the dependent variable leptospirosis (5.096069). So, log transformation was done to normalize data for linear regression. Initially the Independent variables pooled were rainfall, relative humidity, minimum temperature, maximum temperature, mean vapor pressure, mean wind speed and number of days with rainfall.

5.1.1 MULTIPLE REGRESSION MODEL FOR DENGUE

Minimum temperature and maximum temperature has low correlation with dependent variable.

Correlations

		dengue	rainfall
dengue	Pearson Correlation	1	.422**
	Sig. (2-tailed)		<.001
	N	120	120
rainfall	Pearson Correlation	.422**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.1

Correlations

		dengue	relative_humidity
dengue	Pearson Correlation	1	.561**
	Sig. (2-tailed)		<.001
	N	120	120
relative_humidity	Pearson Correlation	.561**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.2

Correlations

		dengue	max_temp
dengue	Pearson Correlation	1	.024
	Sig. (2-tailed)		.792
	N	120	120
max_temp	Pearson Correlation	.024	1
	Sig. (2-tailed)	.792	
	N	120	120

Figure 5.3

Correlations

		dengue	min_temp
dengue	Pearson Correlation	1	.024
	Sig. (2-tailed)		.792
	N	120	120
min_temp	Pearson Correlation	.024	1
	Sig. (2-tailed)	.792	
	N	120	120

Figure 5.4

Correlations

		dengue	mvp
dengue	Pearson Correlation	1	.281**
	Sig. (2-tailed)		.002
	N	120	120
mvp	Pearson Correlation	.281**	1
	Sig. (2-tailed)	.002	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.5

Correlations

		dengue	mwsp
dengue	Pearson Correlation	1	-.133
	Sig. (2-tailed)		.148
	N	120	120
mwsp	Pearson Correlation	-.133	1
	Sig. (2-tailed)	.148	
	N	120	120

Figure 5.6

Correlations

		dengue	no_of_days
dengue	Pearson Correlation	1	.581**
	Sig. (2-tailed)		<.001
	N	120	120
no_of_days	Pearson Correlation	.581**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.7

Since VIF (Variation Inflation factor) >10, no of days with rainfall excluded from the model. This was confirmed by stepwise regression and automatic linear modeling and the best predictors were found out as relative humidity, mean vapour pressure, rainfall, and mean wind speed.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	humidity	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	mvp	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	rainfall	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	mwsp	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: dengue

Figure 5.8 Stepwise Regression

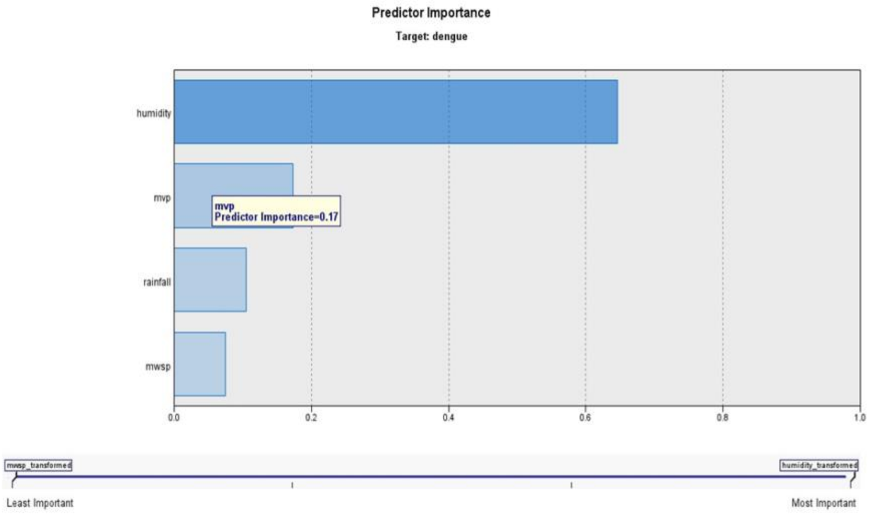


Figure 5.9 Automatic Linear Modeling

A multiple regression model was formulated using data from 2012 to 2021

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	mwsp, mvp, rainfall, max_temp, humidity[b]	.	Enter

a. Dependent Variable: dengue

b. Tolerance = .000 limit reached.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.626[a]	.391	.365	.34758	.391	14.660	5	114	<.001	.481

a. Predictors: (Constant), mwsp, mvp, rainfall, max_temp, humidity

b. Dependent Variable: dengue

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-.784	1.010		-.776	.439	-2.786	1.217					
	rainfall	-.001	.000	-.334	-2.199	.030	-.001	.000	.422	-.202	-.161	.231	4.333
	humidity	.064	.010	1.193	6.161	<.001	.043	.084	.561	.500	.450	.142	7.019
	max_temp	.023	.065	.051	.354	.724	-.105	.151	.024	.033	.026	.254	3.932
	mvp	-.076	.033	-.412	-2.320	.022	-.141	-.011	.281	-.212	-.170	.170	5.897
	mwsp	.077	.044	.169	1.756	.082	-.010	.163	-.133	.162	.128	.578	1.730

a. Dependent Variable: dengue

Figure 5.10 Model Summary

The model had an explanatory capacity of 39% thus explaining that much of the total variations in the occurrence of dengue cases. Relevant assumptions of the model were met. So, since the value of sig column in the ANOVA table is less than 0.5, we know that this value of R-squared is significantly greater than 0 and that means our predictors can account for a significant amount of variance in dengue cases. Therefore, the Regression model is significant.

5.1.2 MULTIPLE REGRESSION MODEL FOR LEPTOSPIROSIS

Min temp and max temp excluded from model due to low correlation with dependent variable.

Correlations

		leptospirosis	rainfall
leptospirosis	Pearson Correlation	1	.382**
	Sig. (2-tailed)		<.001
	N	120	120
rainfall	Pearson Correlation	.382**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.11

Correlations

		leptospirosis	relative_humidity
leptospirosis	Pearson Correlation	1	.495**
	Sig. (2-tailed)		<.001
	N	120	120
relative_humidity	Pearson Correlation	.495**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.12

Correlations

		leptospirosis	min_temp
leptospirosis	Pearson Correlation	1	-.121
	Sig. (2-tailed)		.189
	N	120	120
min_temp	Pearson Correlation	-.121	1
	Sig. (2-tailed)	.189	
	N	120	120

Figure 5.13

Correlations

		leptospirosis	max_temp
leptospirosis	Pearson Correlation	1	-.121
	Sig. (2-tailed)		.189
	N	120	120
max_temp	Pearson Correlation	-.121	1
	Sig. (2-tailed)	.189	
	N	120	120

Figure 5.14

Correlations

		leptospirosis	mvp
leptospirosis	Pearson Correlation	1	.196*
	Sig. (2-tailed)		.032
	N	120	120
mvp	Pearson Correlation	.196*	1
	Sig. (2-tailed)	.032	
	N	120	120

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 5.15

Correlations

		leptospirosis	mwsp
leptospirosis	Pearson Correlation	1	-.343**
	Sig. (2-tailed)		<.001
	N	120	120
mwsp	Pearson Correlation	-.343**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.16

Correlations

		leptospirosis	no_of_days
leptospirosis	Pearson Correlation	1	.407**
	Sig. (2-tailed)		<.001
	N	120	120
no_of_days	Pearson Correlation	.407**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.17

Since VIF>10, no of days with rainfall excluded from the model.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	relative_humidity	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	mvp	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: leptospirosis

Figure 5.18 Stepwise Regression

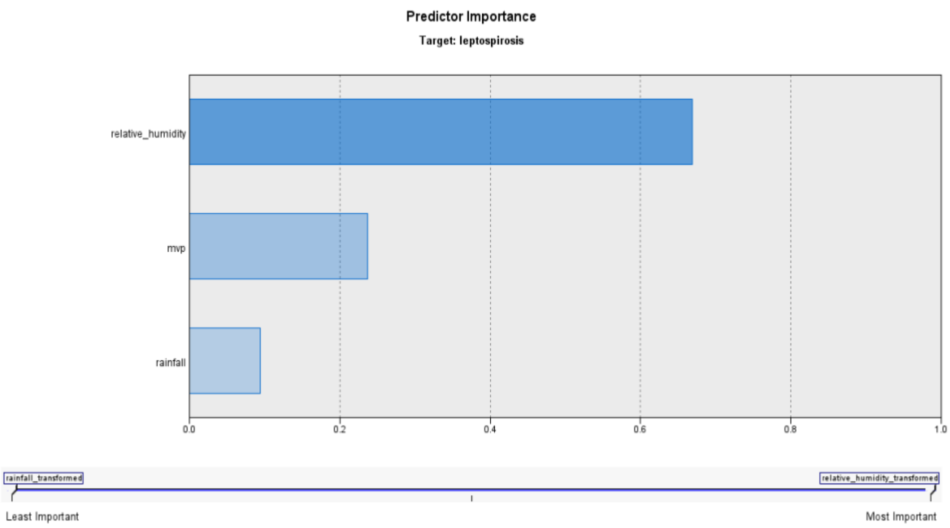


Figure 5.19 Automatic Linear Modelling

This was confirmed by stepwise regression and automatic linear modelling and the best predictors were found out as relative humidity, mean vapour pressure and rainfall.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.555[a]	.308	.290	.23678	.308	17.218	3	116	<.001	.795

a. Predictors: (Constant), rainfall, mvp, relative_humidity

b. Dependent Variable: leptospirosis

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.896	3	.965	17.218	<.001[b]
	Residual	6.504	116	.056		
	Total	9.400	119			

a. Dependent Variable: leptospirosis

b. Predictors: (Constant), rainfall, mvp, relative_humidity

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	.733	.372		1.969	.051					
	relative_humidity	.034	.007	.992	5.208	<.001	.495	.435	.402	.164	6.086
	mvp	-.040	.013	-.339	-3.047	.003	.196	-.272	-.235	.483	2.070
	rainfall	.000	.000	-.304	-1.917	.058	.382	-.175	-.148	.237	4.225

a. Dependent Variable: leptospirosis

Figure 5.20 Model Summary

The model had an explanatory capacity of 30% thus explaining that much of the total variations in the occurrence of dengue cases. Relevant assumptions of the model were met.

So, since the value of sig column in the ANOVA table is less than 0.5, we know that this value of R-squared is significantly greater than 0 and that means our predictors can account for a significant amount of variance in leptospirosis cases. Therefore, the Regression model is significant. In the coefficients table we can see that the p value of relative humidity and mean vapour pressure is less than 0.05 and these two are the significant predictors of leptospirosis cases. Finally, the factor rainfall is not a significant predictor as its p value is slightly greater than 0.05.

5.2 RIDGE REGRESSION

Linear Ridge in SPSS employs the Python learn linear model. Ridge module to construct linear regression models that are regularized using L2 or squared loss. These models predict a dependent variable based on one or more independent variables. The module offers additional functionalities, such as displaying trace plots and facilitating the selection of the alpha hyperparameter through cross-validation. In situations where a single model is trained or alpha is chosen via cross-validation, a portion of the data can be set aside as holdout data to evaluate the model's out-of-sample performance. During the process of selecting alpha through cross-validation, a grid search with cross-validation is conducted to assess various models. The best alpha is determined by evaluating the models based on the highest average R^2 across the validation folds.

5.2.1 RIDGE REGRESSION FOR DENGUE

Case Processing Summary

		N	Percent
Sample	Training	84	70.0%
	Holdout	36	30.0%
Valid		120	100.0%
Excluded		0	
Total		120	

Best Model Summary[a,b]

Alpha	Number of Crossvalidation Folds	Training R Square	Average Test Subset R Square	Holdout R Square
1.000	5.000	1.000	.003	-.110

a. Dependent Variable: dengue

b. Model: VAR00004, VAR00005, VAR00008, VAR00002, VAR00003, VAR00006, VAR00007

Figure 5.21

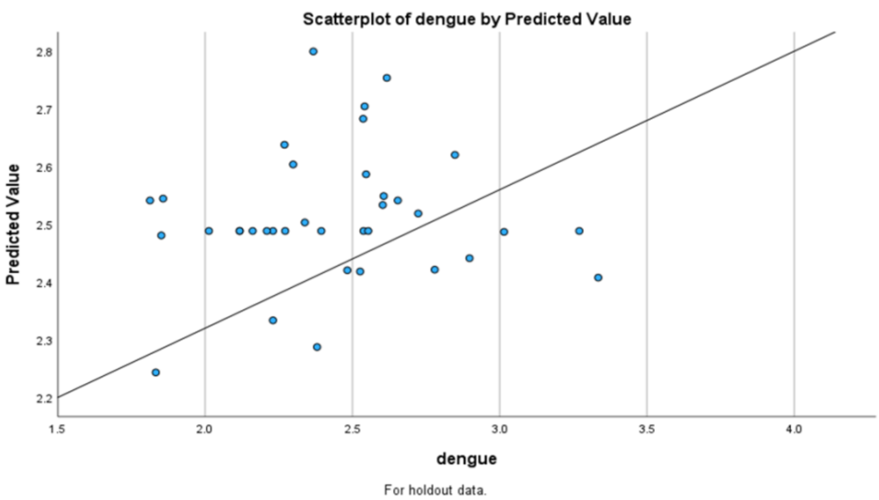


Figure 5.22

There is a linear relationship between the observed dengue cases and the predicted values. R^2 value is 68.8. Mean absolute error is 0.12

5.2.2 RIDGE REGRESSION FOR LEPTOSPIROSIS

Case Processing Summary

		N	Percent
Sample	Training	89	74.2%
	Holdout	31	25.8%
Valid		120	100.0%
Excluded		0	
Total		120	

Figure 5.23

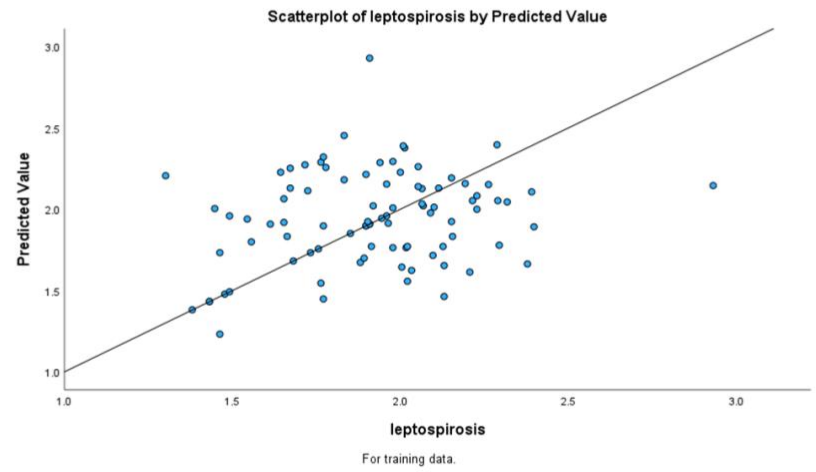


Figure 5.24

There is a linear relationship between the observed leptospirosis cases and the predicted values R^2 value is 78.1. Mean absolute error is 0.053.

Chapter 6

MODELING SEASONAL TIME SERIES DATA (DENGUE) WITH SARIMA

A SARIMA (p, d, q) (P, D, Q) S model was fitted, where p is the order of autoregression, d is the order of integration, q is the order of moving average, P is the order of seasonal autoregression, D is the order of seasonal integration, Q is the order of seasonal moving average and s is the length of seasonal period. The analyses were done in Google Colaboratory.

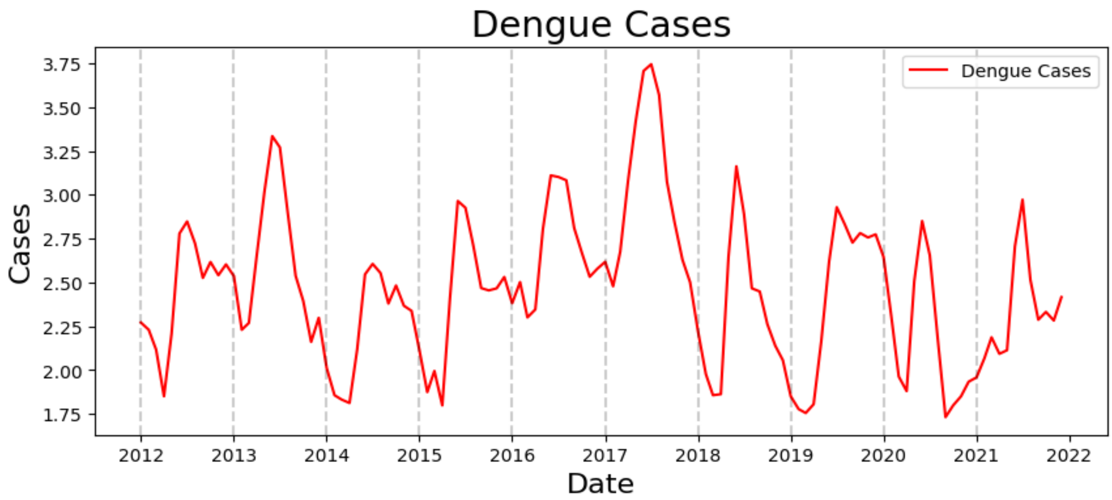
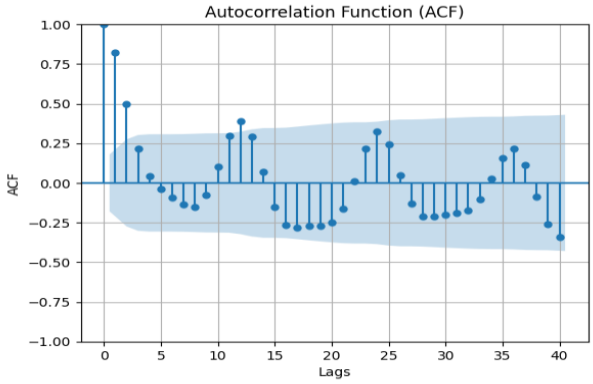


Figure 6.1

To check stationarity and to estimate the d parameter and speed up the process Augmented Dickey–Fuller test was conducted. Since p-value was lower than 0.05, therefore, we can assume that our data is stationary. The ACF and PACF plots were plotted to confirm the presence of seasonality and the seasonal period.



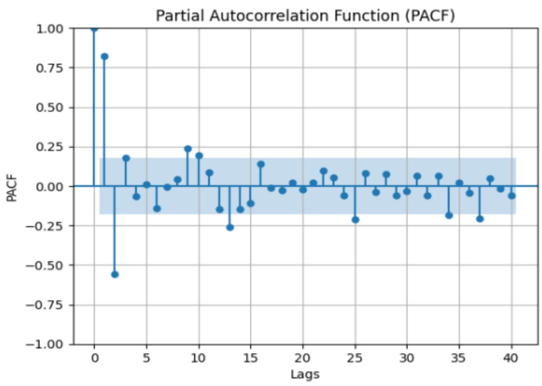


Figure 6.2

A code used essentially automates the process of SARIMA model selection by iterating through different combinations of parameters, fitting SARIMA models, and selecting the best models based on information criteria (AIC and BIC). It helps to identify the SARIMA model that best fits the given time series data.

Best SARIMA model based on AIC:

param	(0, 1, 2)
param_seasonal	(1, 0, 1, 12)
AIC	-72.83524

The non-seasonal autoregressive order (p) is 0. The non-seasonal differencing order (d) is 1. The non-seasonal moving average order (q) is 2. The seasonal autoregressive order (P) is 1. The seasonal differencing order (D) is 0. The seasonal moving average order (Q) is 1. The seasonal period (s) is 12, indicating monthly seasonality. Thus the model consists of non-seasonal ARIMA(0,1,2) and seasonal SARIMA (1,0,[1],12) components.

The data is split into training and testing data. The training data includes all observations from the beginning of the time series up to November 2018. The testing data includes observations from December 2018, up to November 2021. A plot showing the actual dengue data and the corresponding predictions over time is generated.

SARIMAX Results

==

Dep. Variable: dengue No. Observations: 84

Model: SARIMAX(0, 1, 2)x(1, 0, [1], 12) Log Likelihood 26.797

AIC -43.594

BIC -31.500

Sample: 01-01-2012 HQIC -38.735

- 12-01-2018

Covariance Type: opg

==

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.2247	0.104	2.159	0.031	0.021	0.429

ma.L2	-0.1377	0.138	-0.995	0.320	-0.409	0.134
ar.S.L12	0.9993	0.042	23.568	0.000	0.916	1.082
ma.S.L12	-0.9537	1.400	-0.681	0.496	-3.698	1.791
sigma2	0.0223	0.029	0.769	0.442	-0.035	0.079

===

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	7.16
Prob(Q):	0.96	Prob(JB):	0.03
Heteroskedasticity (H):	1.28	Skew:	0.71
Prob(H) (two-sided):	0.52	Kurtosis:	3.26

The high value of log likelihood (26.79), low values of AIC (Akaike Information Criterion) (-43.59), BIC (Bayesian Information Criterion)(-31.5), and HQIC (Hanan-Quinn Information Criterion) (-38.73) indicate better fit. AIC balances the goodness of fit of the model with the number of parameters used. It penalizes model with more parameters to prevent overfitting. Similar to AIC, BIC penalizes models with more parameters but with a more severe penalty compared to AIC.

Coefficient Estimates (ma.L1, ma.L2, ar.S.L12, ma.S.L12) represent the effects of lagged values on the current value of the time series. ma.L1(Moving Average ,Seasonal Lag 12) ma.S.L12 represents the effect of the first lagged value of the moving average component on the current value of the time series. ma.L2(Moving Average, Lag 2) represents the effect of the second lagged value of the moving average component on the current value of the time series. ar.S.L12(Autoregressive, Seasonal Lag 12) represents the effect of the twelfth seasonal lagged value of the autoregressive component on the current value of the time series. ma.S.L12(Moving Average, Seasonal lag 12) represents the effect of the twelfth seasonal lagged value of the moving average component on the current value of the time series. Coefficients with p-values less than 0.05 (ma.L1, ar.S.L12) are considered statistically significant.

The diagnostic tests evaluate various aspects of residuals from SARIME model. Ljung-Box(L1)(Q) test evaluates the presence of autocorrelation in the residuals. The reported statistic is 0.00, which indicates that there is no significant autocorrelation in the residuals at lag 1. Jarque-Bera(JB) test assesses the normality of the residuals. There is evidence against the null hypothesis of normality, suggesting that the residuals may not follow a normal distribution. Heteroskedasticity(H) test examines whether the variance of the residuals is constant over time. There is no significant evidence of heteroskedasticity in the residuals, indicating that the variance remains constant over time.

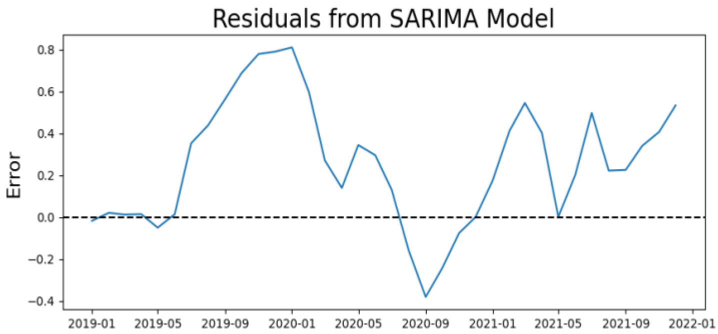


Figure 6.3

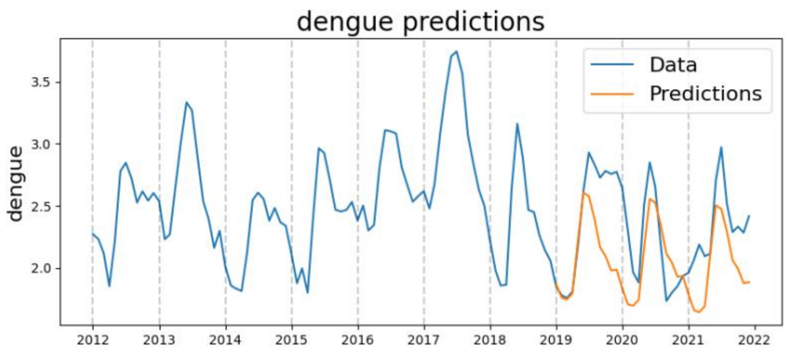


Figure 6.4

Mean Absolute Percent Error is 0.1277. Root Mean Squared Error is 0.3913442825452217. A Mean Absolute Percent Error of 0.1277 suggests that, on average, the absolute percentage difference between the predicted dengue cases and the actual dengue cases is approximately 12.77%. The Root Mean Squared Error of 0.3913 suggests that, on average, the difference between the predicted dengue cases and the actual dengue cases is approximately 0.3913 units.

Chapter 7

RESULTS

Correlations

		dengue	Unstandardized Predicted Value
dengue	Pearson Correlation	1	.605**
	Sig. (2-tailed)		<.001
	N	120	120
Unstandardized Predicted Value	Pearson Correlation	.605**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 7.1 Correlation between actual dengue and predictions of multiple regression

Correlations

		dengue	PredictedValue
dengue	Pearson Correlation	1	.830**
	Sig. (2-tailed)		<.001
	N	120	120
PredictedValue	Pearson Correlation	.830**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 7.2 Correlation between actual dengue and predictions of ridge regression

The Pearson's coefficients for multiple regression and ridge regression are 0.6 and 0.8 respectively.

Correlations

		leptospirosis	Unstandardized Predicted Value
leptospirosis	Pearson Correlation	1	.563**
	Sig. (2-tailed)		<.001
	N	120	120
Unstandardized Predicted Value	Pearson Correlation	.563**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 7.3 Correlation between actual leptospirosis cases and predictions of multiple regression

Correlations

		leptospirosis	PredictedValue
leptospirosis	Pearson Correlation	1	.884**
	Sig. (2-tailed)		<.001
	N	120	120
PredictedValue	Pearson Correlation	.884**	1
	Sig. (2-tailed)	<.001	
	N	120	120

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 7.4 Correlation between actual leptospirosis cases and

predictions of ridge regression

The Pearson's coefficients for multiple regression and ridge regression are 0.56 and 0.88 respectively.

Dengue	R^2 Score	Mean Absolute Error	Pearson's Coefficient
Multiple Regression	39.1	0.27	0.6
Ridge Regression	68.8	0.12	0.8

Fig 7.5

Leptospirosis	R^2 score	Mean Absolute Error	Pearson's Coefficient
Multiple Regression	30.8	0.18	0.56
Ridge Regression	78.1	0.053	0.88

Fig 7.6

The Pearson's correlation coefficient for ridge regression model is higher than multiple regression model for both the diseases. Ridge regression model performed way better than multiple regression model in aspect of Pearson's correlation coefficient, R^2 score and mean absolute error.

SARIMA model was constructed to model the dengue cases. The SARIMA model applied to the dengue cases dataset spanning from January 2012 to December 2018 yielded insightful results. The model, with parameters SARIMAX (0, 1, 2) x (1, 0, [1], 12), demonstrated statistical significance for the moving average term ma.L1 and the autoregressive seasonal term ar.S.L12, implying their importance in capturing the data's temporal dynamics. Diagnostic tests indicated no significant autocorrelation at lag 1 and no evidence of heteroskedasticity, although there was some deviation from normality in the residuals according to the Jarque-Bera test. The Mean Absolute Percent Error (MAPE) of 0.1277 and Root Mean Squared Error (RMSE) of 0.3913 suggest that, on average, the model predicts dengue cases with approximately 12.77% absolute percentage difference and 0.3913 units difference from the actual values, respectively. Despite minor deviations from normality in residuals, the model's statistical significance and relatively low error metrics indicate its potential for accurate forecasting. With further refinement, particularly in addressing the residual normality, the SARIMA model holds great promise for providing valuable insights into dengue incidence trends, aiding in effective public health planning and intervention strategies.

Chapter 8

CONCLUSION

In conclusion, the superiority of the Ridge regression model over the multiple regression model for dengue prediction underscores the significance of regularization techniques in handling multicollinearity and overfitting. The enhanced performance of the Ridge regression model emphasizes its efficacy in capturing the intricate relationships within the data, thereby yielding more accurate predictions. This outcome highlights the importance of leveraging advanced statistical methodologies tailored to the complexities inherent in dengue prediction tasks, ultimately contributing to more robust and reliable predictive models in epidemiological research and public health interventions. While climatic variables serve as essential predictors in forecasting dengue outbreaks, the multifaceted nature of the disease warrants the incorporation of additional factors for more accurate predictions. Beyond climatic conditions, variables such as population density, urbanization levels, human mobility patterns, healthcare infrastructure, and socio-economic status profoundly impact dengue transmission dynamics.

SARIMA models have established themselves as valuable tools in epidemiological research due to their effectiveness in forecasting disease trends. Incorporating additional relevant variables such as temperature, humidity, and rainfall into the modelling process can enhance the predictive accuracy of SARIMA models. These variables, often observed to influence disease transmission dynamics, offer valuable insights that can refine the model's ability to capture and predict epidemiological patterns more comprehensively. Therefore, by integrating such environmental factors alongside the existing time series data, SARIMA models become more robust and tailored to the complexities of epidemiological dynamics, ultimately improving their forecasting performance.

Chapter 9

REFERENCES

---

[1]https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe- dengue

[2]https://main.icmr.nic.in/sites/default/files/icmr_bulletins/april-may06.pdf

[3] Mutheneni, Srinivasa Rao, et al. "Dengue burden in India: recent trends and importance of climatic parameters." *Emerging microbes & infections* 6.1 (2017): 1-10.

[4] Haake, David A., and Paul N. Levett. "Leptospirosis in Humans." Leptospira and Leptospirosis (2015): 65-97.

[5] James, Sruthi, et al. "Outbreak of leptospirosis in Kerala." Nepal journal of epidemiology 8.4 (2018):745.

[6] James, Sruthi, et al. "Outbreak of leptospirosis in Kerala." *Nepal journal of epidemiology* 8.4 (2018): 745.

[7] Karim, Md Nazmul, et al. "Climatic factors influencing dengue cases in Dhaka city: a model for dengue prediction." *The Indian journal of medical research* 136.1 (2012): 32.

[8] Mutheneni, Srinivasa Rao, et al. "Dengue burden in India: recent trends and importance of climatic parameters." *Emerging microbes & infections* 6.1 (2017): 1-10.

[9] Duarte, Juliana Lúcia, et al. "Climatic variables associated with dengue incidence in a city of the Western Brazilian Amazon region." *Revista da Sociedade Brasileira de Medicina Tropical* 52 (2019).

[10] Coelho, Micheline SZS, and Eduardo Massad. "The impact of climate on Leptospirosis in São Paulo, Brazil." *International journal of biometeorology* 56 (2012): 233-241.

[11] Pawar, ShivshaktiD, et al. "Seasonality of leptospirosis and its association with rainfall and humidity in Ratnagiri, Maharashtra." *Int J Health Allied Sci* 7.1 (2018): 37-40.

[12]https://www.scribbr.com/statistics/one-way-anova/

[13] https://www.scribbr.com/statistics/multiple-linear-regression/

[14]https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=features-linear-ridge-regression

[15] https://deepchecks.com/glossary/mean-absolute-error/

[16] Fernando, Jason. "R-squared: Definition, calculation formula, uses, and limitations." Investopedia. Retrieved September 8 (2023): 2023.

[17]https://towardsdatascience.com/time-series-forecasting-with-a-sarima-model-db051b7ae459

[18]https://stats.stackexchange.com/questions/129901/sarima-model-equation