

STATISTICAL ANALYSIS OF RAINFALL IN CHENNAI

Submitted in partial fulfillment of the requirements for the award of
Bachelor of science degree in mathematics (B.sc Mathematics)

BY

V . THANGATHILAKA (38060013)



**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 - 600 119**

APRIL - 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF MATHEMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of

V . THANGATHILAKA(38060013) who carried out the project entitled
"STATISTICAL ANALYSIS OF RAINFALL IN CHENNNAI" under our supervision
from **JANUARY 2021** to **APRIL 2021**

Internal Guide

(Name in Capital letters
Signature)

[H. GIRIJA BAI]

External Guide (if Applicable)

(Name in Capital letters with
with signature)

(Dr. M.NIRMALA)

Head of the Department

(Name in Capital letters with Seal & Signature)

Submitted for Viva voce Examination held on _____

Internal Examiner

(Name in Capital letters
with Signature)

External Examiner

(Name in Capital letters
with signature)

DECLARATION

IV.THANGATHILAKA (38060013) hereby declare that the Project Report entitled
“**STATISTICAL ANALYSIS ON RAINFALL IN**

CHENNAI” done by me under the guidance of **Dr. H. GIRIJA BAI** (Internal) and

_____ (External) at **Sathyabama Institute of Science and**

Technology submitted in partial fulfillment of the requirements for the award of Bachelor of
Engineering / Technology degree in Mathematics.

V. Thangathilaka..

DATE:

PLACE: SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to Dean, School of Science and Humanities and **DR. M. NIRMALA**, Head of the Department, Dept. of Mathematics for providing necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **DR.H.GIRIJA** 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

The objective of the project is to analyse the statistical rainfall in Chennai. It has been of great concern due to the attention given to global climate change. Also, the analysis of rainfall variability at any place is vital for many applications like forecasting, in identifying the reliability or risks, irrigation engineering, designs of hydraulic structures, public health engineering, drainage system design etc.

In this project, the rainfall observed at Chennai from 2016 to 2020. Incidentally, the annual lowest rainfall was 0.10 mm in the year of 2020 and the highest was 548.8 mm in the year of 2020.

The spells of daily rainfall received at Chennai for last five year period (2016-2020) have been analysed. The study reveals that the spells have three distinct patterns with February and March having negligible spells while April to September being 15 to 20 percent and from October to January having very high values.

By the collected rainfall data for past five years (2016-2020), we analysed various statistical tools and predict the rainfall trend for the year 2021.

KEYWORD

- Statistical analysis
- Bivariate
- Karl Pearson's coefficient of correlation
- Regression

CONTENTS

Chapter 1Page no

Introduction

1.1 Introduction	2
1.2 Statistical Analysis	2
1.3 Steps of statistical analysis	3

Chapter 2

Literature survey5

Chapter 3

Research Methodology

3.1 Aim and Scope	
3.11 correlation analysis	9
3.2Assumption ofCorrelation10	
3.3 Bivariate correlation	11
3.4 Types of correlation	11
3.5 Coefficient of correlation and probable error	13
3.6 Condition for the use of probable error	13
3.7 Regression	14

Chapter 4

4.1 Result and discussion 17

4.2 By pearson'scoefficient of correlation for

the average rainfall data 19

4.3 Regression equation of x on y 19

4.4 From the above data using r-programming 20

Chapter 5

5.1 Result and discussion 24

References25

List of tables and figures

	Page no
Table 3.1 difference between correlation and regression	15
Table 4.1 average rainfall data (month wise)	17
Table 4.2 average rainfall during 2016-2020	19
Figure 3.1 positively correlated or negatively correlated	10
Figure 4.1 chart for average rainfall data (monthwise)	18
Figure 4.2 plot(year,rainfall)	20
Figure 4.3 plot(year,rainfall,"h")	20
Figure 4.4 plot (year,rainfall,"h",xlab="year",ylab="average rainfall",main="average rainfall during 2016 to 2020")	21
Figure 4.5 Plot (year,rainfall,"o")	22

Chapter 1

1.1 INTRODUCTION

The rainfall received in a given period at a location is highly variable from one year to another. The study of long-term analysis of rainfall at any place finds use in many applications like irrigation engineering, water harvesting, public health engineering, drainage system design etc.

Also the analysis of the rainfall trend is vital for the purpose of forecasting in identifying the changes and impacts linked to national economy. Extreme rainfall events and the resulting floods usually cause a lot of damage to life and properties of human society also have large negative impacts on many vulnerable aspects of human health, social organization, and nature system and the estimation of extreme values of rainfall parameters is an important aspect of the siting and design of a nuclear facility.

Chennai located at 13° N, 80.2° E on the southeast coast of India, is an important metropolitan city. The study of rainfall over a populated city like Chennai assumes importance especially for the water managers for providing regular water supply to the city population. In this paper statistical analysis of rainfall events observed from 2016 to 2020 at Chennai is done through correlation coefficient and regression is explained.

1.2 STATISTICAL ANALYSIS

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect, summarize data. Statistics is a mathematical body of science that pertains to the collection, analysis, explanation and presentation of data. It is used only to describe the sample or summarize information about the sample and to analyse the sample. The field of statistics touches our life in many ways.

Statistics is used in many sectors such as psychology, geology, sociology, weather forecasting, probability and much more. The goal of statistics is to gain understanding from the data, it focuses on applications, and hence, it is distinctively considered as a mathematical science. Mathematically, statistics is defined as the set of equations, which are used to analyse the things.

1.3 STEPS OF STATISTICAL ANALYSIS

Statistical analysis can be broken down into five discrete steps, as follows:

- Describe the nature of the data to be analyzed.
- Explore the relation of the data to the underlying population.
- Create a model to summarize an understanding of how the data relates to the underlying population .
- Prove (or disprove) the validity of the model.
- Employ predictive analytics to run scenarios that will help guide future actions.

Chapter 2

Literature Survey

S.Natarajan ,B.Ramakrishnan and R.M.A.N Ramanathan has given "An Analytical Study Of Rainfall Pattern Over Chennai".The spells of daily rainfall received at Nungambakkam and Meenambakkam observatories located in Chennai for ten year period have been analysed.

Zulfiqar Ali,S.Bala Bhaskar has proposed" Basic statistical tools in research and data analysis" .The concepts of the basic statistical methods used for conduct of a research study. The basic statistical methods will go a long way in improving the research designs and producing quality medical research which can be utilised for formulating the evidence-based guidelines.

T.Jesan , C.Manonmani, S.Rajaram,P.M. Ravi,R.M.Tripathi,S.Predeepkumarwere analysed "Statistical analysis of rainfall events at Kalpakkam " .In this paper the statistical analysis of rainfall events observed from 1968 to 2015 at Kalpakkam site is done through Mann-Kendall trend test method, Hurst exponent and extreme value distribution analysis with bootstrap estimation.

S. Kokilavani , R. PangayarSelvi , S. Panneerselvam and G.A. Dheebakaran,has given " Trend Analysis of Rainfall Variability in Western Agro Climatic Zone of Tamil Nadu" .The Simple descriptive statistics along with Co-efficient of Variation (CV) were worked to understand the rainfall variability. For Coimbatore and Erode districts, in both the monsoon seasons and annual analysis revealed an increasing trend for long term (100 years-(1916-2015) period).

Mahadevan Palanichamy,RamaswamySankaralingam&Narayanasamy has given "Statistical Studies on Rainfall and Time-based Deviations in Precipitation Trends in Vaigai River Basin".A trend analysis of rainfall was carried out in selected seven rainfall stations of Vaigai river basin in Tamil Nadu state, India,for the period from 1959 to 2016. To determine the trend of rainfall, non-parametric Mann-Kendall test and Spearman Rho tests were used.

S.I. Sridhar and A. Raviraj has proposed "Statistical Trend Analysis of Rainfall in Amaravathi River Basin Using Mann-Kendall Test". In this study, the mean annual and seasonal rainfall in Amaravathi basin have been analysed to determine the trend and its magnitude for the period of 1982-2014. The trend analysis is done using Mann-Kendall and Sen's slope estimator.

M. Mayilvaganan , P. Vanitha were analysed “Correlation Analysis of Meteorological Data in Region of Tamil Nadu Districts Based On K- Means Clustering Algorithm “.The aim of this research work, to focuses the clustering method to analyses the monsoon seasons between twelve months. In this paper to focus the clustering task for grouping the weather data in season wise on correlation method.

M. Nyatuame,V. Owusu-Gyimah and F. Ampiaaw were analysed ”Statistical Analysis of Rainfall Trend for Volta Region in Ghana”. The study was conducted to establish the rainfall trends in Volta Region and also to provide the evidence of climate change by analyzing available rainfall record for 30-year period of 1981 to 2011. Statistical analysis (i.e., LSD, ANOVA) was performed to assess any significant difference among the three zones and within the months and years under study.

G.Arvind, P.Ashok Kumar, S.GirishKarthi and C.R.Suribabu were analysed “ Statistical Analysis of 30 Years Rainfall Data: A Case Study”.The mean, standard deviation and coefficient of variation of monthly and annual rainfall was calculated to check the rainfall variability. From the calculated results, the rainfall pattern is found to be erratic. The best fit probability distribution was identified based on the minimum deviation between actual and estimated values

Muktarun Islam, Tahmid Imran Imon, AtaurGoniPolash has given “Analysis of Rainfall Variability in Sylhet Region of Bangladesh” .The study was conducted to establish the rainfall trends for 35 years period of 1981 to 2015. The mean, standard deviation and coefficient of variation of annual and monthly rainfall was calculated to understand the rainfall variability. Linear regression analysis revealed downward trend in the data in some months and years but statistically insignificant.

V. Rajendran, R. Venkatasubramani& G. Vijayakumar has given ”Rainfall variation and frequency analysis study in Dharmapuri district, India”. The studies have also been carried out to analyze rainfall data using Markov chain model for identifying appropriate cropping systems in different regionsProbability of each event is calculated by Weibull’smethod.

V. Rajendran, A. Nanthini, G. Padmakumar, K. Thopic Ahamed, J. Vijayvarman has proposed "Rainfall variation and frequency analysis study in Namakkal district, Tamil Nadu". The aim of this project is to give a complete study of ten years (2009-2018) daily rainfall data for Namakkal district was collected. The study area taken for this analysis is Namakkal district of Tamil Nadu in this analysis rainfall variability has been calculated to find out the dependability of rainfall over the study area.

M. Vaidheki, U. Arulanandu and A. Raj Kumar were analysed "Statistical Analysis of Rainfall data in Tiruchirappalli District", A rainfall data for Tiruchirappalli district during 1980 to 2014 (34 years). By Kolmogorov-Smirnov test for the goodness of fit of the probability distributions showed that for majority of the data sets on rainfall at different study periods, Log-normal distribution was found to be the best fit. However, all the data sets were shape dominated which indicated that there was not much large variation in the distribution of rainfall.

Arulmozhi. S and Dr. Prince Arulraj. G has given "Rainfall variation and frequency analysis study of Salem district Tamil Nadu". Weibull method was used to analyze the frequency analysis of the 32 years rainfall data. Probability and return periods were calculated for frequency analysis to design the water conservation structures, ground water recharge, water planning and management for proper crop planning during non rainfall periods.

Chapter 3

RESEARCH METHODOLOGY

3.1 AIM AND SCOPE

3.11 CORRELATION ANALYSIS

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

When the fluctuation of one variable reliably predicts a similar fluctuation in another variable, there's often a tendency to think that means that the change in one causes the change in the other. However, correlation does not imply causation. There may be an unknown factor that influences both variables similarly.

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. Although this correlation is fairly obvious your data may contain unsuspected correlations. You may also suspect there are correlations, but don't know which are the strongest. An intelligent correlation analysis can lead to a greater understanding of your data.

If the points plotted were all on a straight line we would have perfect correlation, but it could be positive or negative as shown in the diagrams above,

- a. Strong positive correlation between x and y. The points lie close to a straight line with y increasing as x increases.
- b. Weak, positive correlation between x and y. The trend shown is that y increases as x increases but the points are not close to a straightline
- c. No correlation between x and y; the points are distributed randomly on the graph.
- d. Weak, negative correlation between x and y. The trend shown is that y decreases as x increases but the points do not lie close to a straightline
- e. Strong, negative correlation. The points lie close to a straight line, with y decreasing as x increases

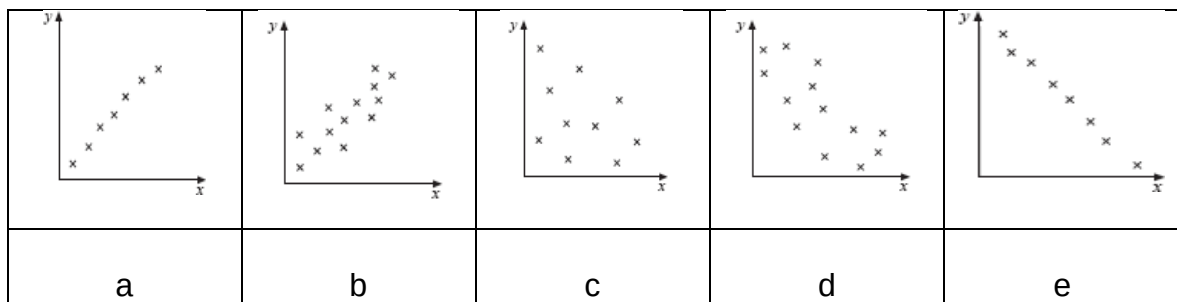


Fig 3.1 positively correlated or negatively correlated

Correlation can have a value:

- i. 1 is a perfect positive correlation
- ii. 0 is no correlation (the values don't seem linked at all)
- iii. -1 is a perfect negative correlation

The value shows how good the correlation is (not how steep the line is), and if it is positive or negative. Usually, in statistics, there are three types of correlations: Pearson correlation, Kendall rank correlation and Spearman correlation.

3.2 ASSUMPTION OF CORRELATION

Employing of correlation rely on some underlying assumptions. The variables are assumed to be independent, assume that they have been randomly selected from the population; the two variables are normal distribution; association of data is homoscedastic (homogeneous), homoscedastic data have the same standard deviation in different groups where data are heteroscedastic have different standard deviations in different groups and assumes that the relationship between the two variables is linear. The correlation coefficient is not satisfactory and difficult to interpret the associations between the variables in case if data have outliers.

An inspection of a scatterplot can give an impression of whether two variables are related and the direction of their relationship. But it alone is not sufficient to determine whether there is an association between two variables. The relationship depicted in the scatterplot needs to be described qualitatively. Descriptive statistics that express the degree of relation between two variables are called correlation coefficients. A commonly employed correlation coefficient are Pearson correlation, Kendall rank correlation and Spearman correlation.

Correlation used to examine the presence of a linear relationship between two variables providing certain assumptions about the data are satisfied. The results of the analysis, however, need to be interpreted with care, particularly when looking for a causal relationship

3.3 BIVARIATE CORRELATION

Bivariate correlation is a measure of the relationship between the two variables; it measures the strength and direction of their relationship, the strength can range from absolute value 1 to 0. The stronger the relationship, the closer the value is to 1. Direction of The relationship can be positive (direct) or negative (inverse or contrary); correlation generally describes the effect that two or more phenomena occur together and therefore they are linked.

3.4 TYPES OF CORRELATION

Pearson's r correlation:

Pearson correlation is widely used in statistics to measure the degree of the relationship between linear related variables. For example, in the stock market, if we want to measure how two commodities are related to each other, Pearson correlation is used to measure the degree of relationship between the two commodities. The following formula is used to calculate the Pearson correlation coefficient r :

$$r = \frac{\frac{\sum XY}{n} - \bar{X}\bar{Y}}{\sqrt{\frac{\sum X^2}{n} - \bar{X}^2} \sqrt{\frac{\sum Y^2}{n} - \bar{Y}^2}}$$

where,

- N - the total number of scores or cases
- $\sum$ - the summation of the items indicated
- $\sum xy$ – the sum of all xy cases
- $\sum x^2$ - each x cases should be squared and then those squares (the sum of the x squared scores)
- $\sum y^2$ – the sum of all y scores
- $\sum y$ – each y scores should be squared and then those squares summed
- $\bar{X}$ - the mean value of x

- $\bar{Y}$ - the mean value of y

Kendall's rank correlation:

Kendall rank correlation is a non-parametric test that measures the strength of dependence between two variables. If we consider two samples, x and y, where each sample size is n, we know that the total number of pairings with x and y is

$n(n-1)/2$. The following formula is used to calculate the value of Kendall rank correlation:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

where,

τ = Kendall rank correlation coefficient

n_c = number of concordant (Ordered in the same way).

n_d = Number of discordant (Ordered differently).

Spearman rank correlation:

Spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. It was developed by Spearman, thus it is called the Spearman rank correlation. Spearman rank correlation test does not assume any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal. The following formula is used to calculate the Spearman rank correlation coefficient:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

where

ρ = Spearman rank correlation coefficient

d_i = the difference between the ranks of corresponding values X_i and Y_i

n = number of value in each data set.

The Spearman correlation coefficient, ρ , can take values from +1 to -1. A ρ of +1 indicates a perfect association of ranks, a ρ of zero indicates no association between ranks and a ρ of -1 indicates a perfect negative association of ranks. The closer ρ to zero, the weaker the association between the ranks.

3.5 COEFFICIENT OF CORRELATION AND PROBABLE ERROR

The probable error of the coefficient of correlation helps in interpreting its value. With the help of probable error it is possible to determine the reliability of the value of the coefficient in so far as it depends on the conditions of random sampling. The probable error of the coefficient of correlation is obtained as follows.

$$P.E._r = 0.6745 \frac{1 - r^2}{\sqrt{N}}$$

Where r is the coefficient of correlation and N the number of pairs of observations.

- If the value of r is less than the probable error there is no evidence of correlation, i.e., the value of r is not at all significant.
- If the value of r is more than six times the probable error, the coefficient of correlation is practically certain, i.e., the value of r is significant.
- By adding and subtracting the value of probable error from the coefficient of correlation we get respectively the upper and lower limits within which coefficient of correlation in the population can be expected to lie. Symbolically,

$$\rho = r \pm P.E$$

Where ρ (rho) denotes correlation in the population .

3.6 CONDITIONS FOR THE USE OF PROBABLE ERROR

The measure of probable error can be properly used only when the following three conditions exist:

- 1.The data must approximate a normal frequency curve (bell-shaped curve).
- 2.The statistical measure for which the P.E. is computed must have been calculated from a sample.
- 3.The sample must have been selected in an unbiased manner and the individual items must be independent.

However, these conditions are generally satisfied and as such the reliability of the correlation coefficient is determined largely on the basis of exterior tests of reasonableness which are often of a statistical character.

3.7 REGRESSION

Is the measure of the average relationship between two or more variables. Regression analysis measures the nature and extent of the relationship between two or more variables, thus enables us to make predictions. Regression coefficient measures the average change in the value of one variable for a unit change in the value of another variable. They represent the slope of regression line. There are two regression coefficients of Y on X and X on Y.

Regression line : Regression lines show the average relationship between two variables. It is also called line of best fit. If two variables x and y are given, then there are two regression lines:

$$\text{Regression line of x on y : } b_{xy} = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$\text{Regression line of y on x : } b_{yx} = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum Y^2) - (\sum Y)^2}$$

Nature of regression lines:

- If $r = \pm 1$, then the two regression lines are coincident.
- If $r = 0$, then the two regression lines intersect each other at 90° .
- The nearer the regression lines are to each other, the greater will be the degree of correlation.
- If regression lines rise from left to right upward, then correlation is positive.

Regression equation

Regression equations are the algebraic formulation of regression lines. There are two regression equations:

❖ Regression equation of y on x

- $Y = a + bX$
- $Y - \bar{Y} = b_{yx}(X - \bar{X})$
- $Y - \bar{Y} = r \cdot \frac{\sigma_y}{\sigma_x}(X - \bar{X})$

❖ Regression equation of x on y

- $X = a + bY$
- $X - \bar{X} = b_{xy}(Y - \bar{Y})$
- $X - \bar{X} = r \cdot \frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$

DIFFERENCE BETWEEN CORRELATION AND REGRESSION	CORRELATION	REGRESSION
DEGREE AND NATURE OF RELATIONSHIP	Correlation is a measure of degree of relationship between x and y.	Regression studies the nature of relationship between the variables so that one may be able to predict the value of one variable on the basis of another.
CAUSE AND EFFECT RELATIONSHIP	Correlation does not always assume cause and effect relationship between two variables.	It clearly expresses the cause and effect relationship between two variables. The independent variable is the cause and dependent variable is effect.
PREDICTION	Correlation doesn't help in making predictions.	Regression enables us to make predictions using regression line.
SYMMETRIC	Correlation coefficients are symmetrical ie., $r_{xy} = r_{yx}$	Regression coefficients are not symmetrical ie., $b_{xy} \neq b_{yx}$

Table 3.1 difference between correlation and regression

CHAPTER 4

4.1 RESULT AND DISCUSSION

The rainfall received in 2016,during January to april is less and during may to December,the rainfall received is high. In 2017,2018 and 2019 ,the rainfall received in the starting of the year(ie., Jan to april) was less and at the end of the year ,it is high. During 2020 ,the rainfall received is less during the starting of the year and it is very high during the end of the year.

By comparing the past five year average rainfall data ,it seem the rainfall received during 2018 is the minimum and during 2020 is the maximum .

Year	2016	2017	2018	2019	2020
January	7.66	10.47	5.28	3.1	48.2
February	0.25	0.46	1.4	10.5	2.7
March	0.13	10.1	1.18	2.3	0.4
April	0.31	0.26	1.6	1	3.8
May	116.59	8.3	10.54	20.8	0.10
June	81.6	37.83	27.14	54	120.1
July	25.37	18.8	5.94	145.9	265.5
August	58.14	78.37	26.14	233	163.7
September	95.94	65.99	47.24	130.2	269.2
October	34.8	133.94	97.44	289.6	211.6
November	15.03	245.53	130.73	360.7	548.8
December	110.77	14.81	28.02	129.8	312.4

Table 4.1 average rainfall data (month wise)

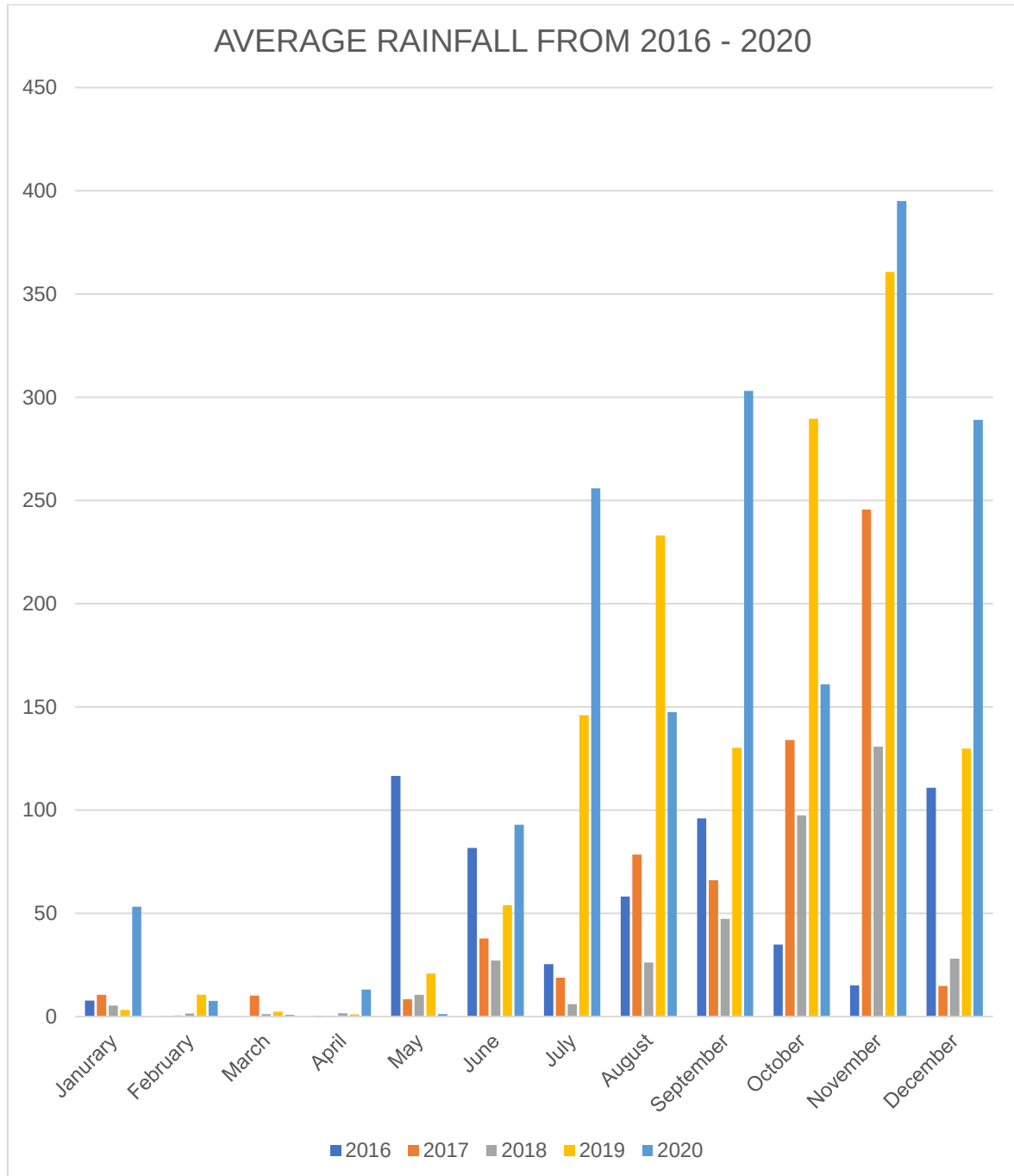


Figure 4.1 chart for average rainfall data (monthwise)

4.2 BY PEARSON'S COEFFICIENT OF CORRELATION FOR THE AVERAGE RAINFALL DATA

- $r_{(2016,2017)} = -0.0816$ [it is negatively correlated]
- $r_{(2017,2018)} = 0.9631$ [it is nearest to +1 then the correlation is strong positive relationship]
- $r_{(2018,2019)} = 0.8817$ [it is nearest to +1 then the correlation is strong positive relationship]
- $r_{(2019,2020)} = 0.3299$ [it is positively correlated]

4.3 REGRESSION EQUATION OF X ON Y

For using regression equation let x be the year and let y be the average rainfall to that year .

X	Y
2016	45.5492
2017	52.0717
2018	31.8875
2019	115.0750
2020	162.2083

Table 4.2 average rainfall during 2016-2020

- By calculating the values required to find b_{xy} and substituting the value in b_{xy}
- We get the values for $b_{xy} = 0.0242$

- After finding the value for b_{xy} , substitute the value required to find the regression equation
- We get the regression equation $x = 0.0242y + 2016.0311$
- Now we are going to predict the average rainfall for this year 2021 by the regression equation We get the average rainfall for the year 2021 is 205.3252 mm

4.4 FROM THE ABOVE DATA USING R-PROGRAMMING

```
year <- c(2016,2017,2018,2019,2020)
```

```
rainfall <- c(45.5492,52.0717,31.8875,115.0750,162.2083)
```

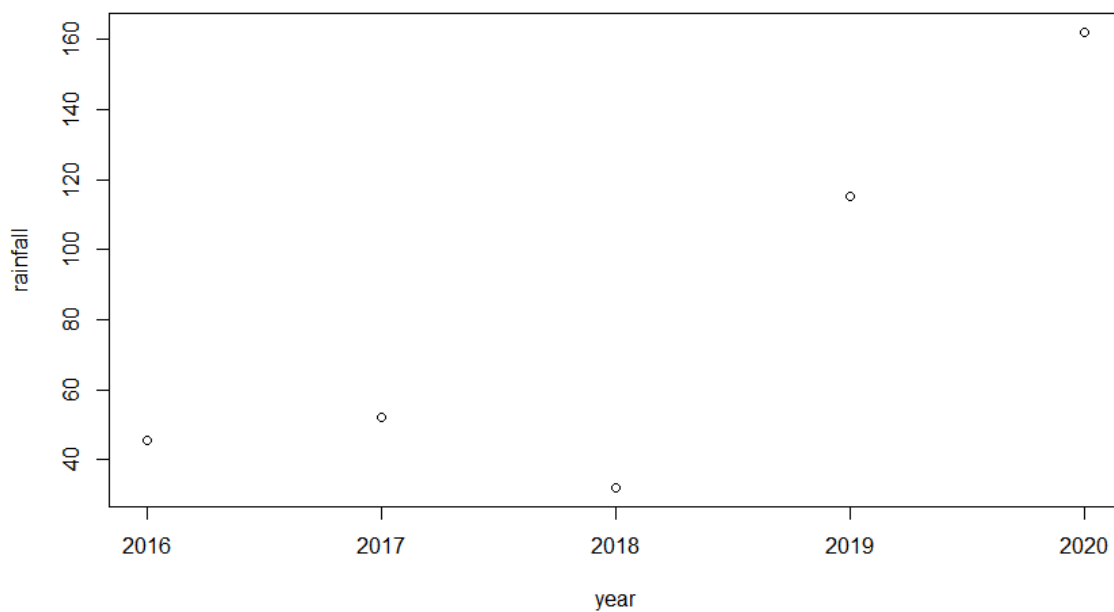


Figure 4.2 `plot(year,rainfall)`

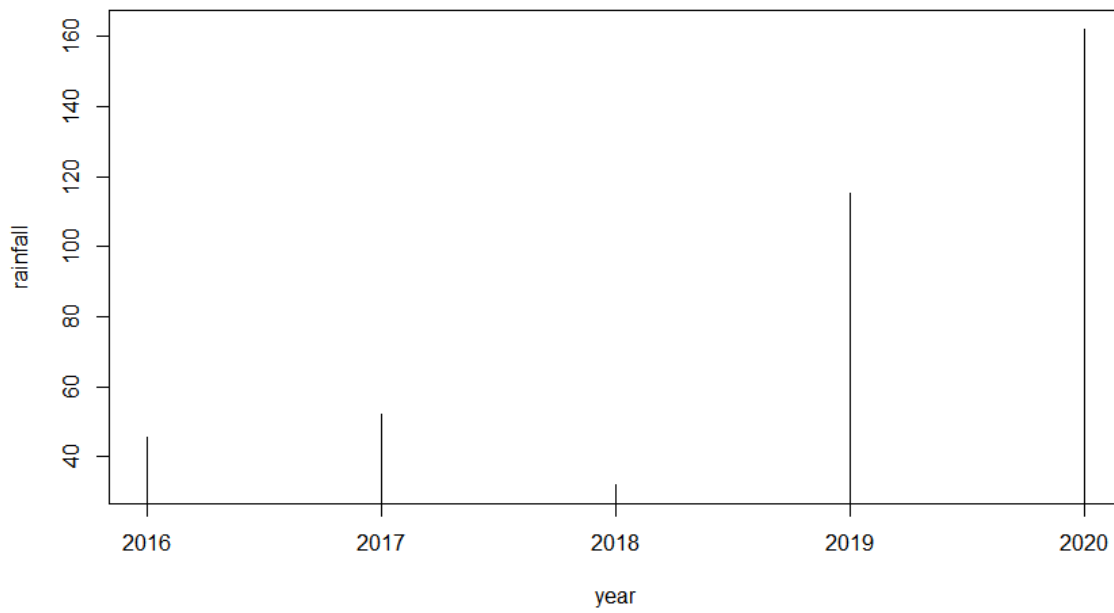


Figure 4.3 `plot(year,rainfall,"h")`

average rainfall during 2016 to 2020

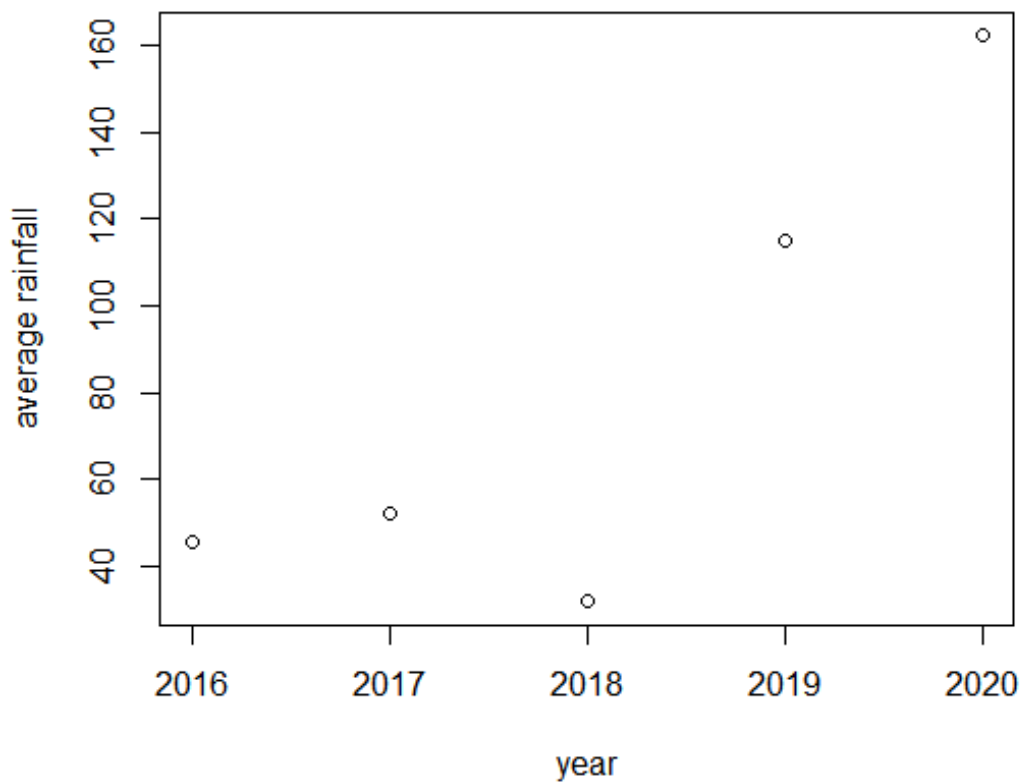


Figure 4.4 `plot(year,rainfall,"h",xlab="year",ylab="average rainfall ",main="average rainfall during 2016 to 2020")`

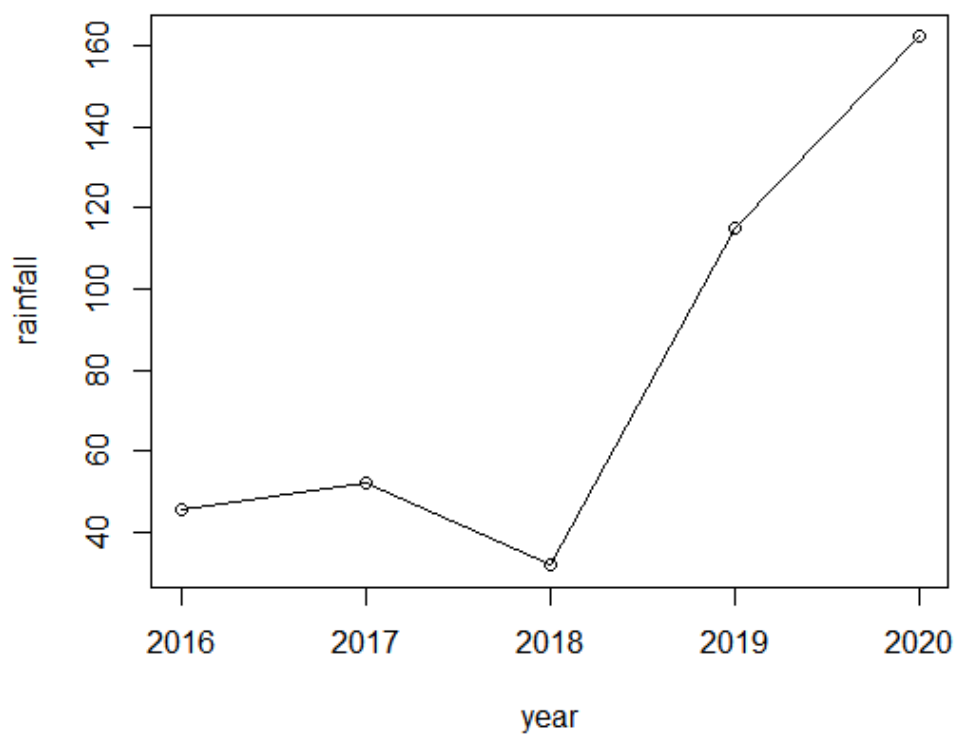


Figure4.5Plot (year,rainfall,"o")

Chapter 5

5.1 CONCLUSION

Statistical analysis of rainfall in Chennai is proposed in this project. correlation coefficient is found for the rainfall data from 2016 to 2020 . for the year 2016 ,the correlation is negatively correlated , for 2017 and 2018 ,it is nearest to +1 then it is positively correlated and for 2020 ,it is positively correlation ,which show the correlation relationship. By regression ,it found regression coefficient and regression equation .Regression lines shows the average relationship between two variable. Therefore by the regression equation ,we predict the average rainfall for the current year 2021 is **205.3252**.

REFERENCE

1. S.Natarajan ,B.Ramakrishnan and R.M.A.N Ramanathan,"An Analytical Study Of Rainfall Pattern Over Chennai",Mausam 55,p.p:- 599-610,2004
2. Zulfiqar Ali,S.Bala Bhaskar," Basic statistical tools in research and data analysis",IP:84.134.164.227,2016
3. T.Jesan , C.Manonmani, S.Rajaram,P.M. Ravi, R.M.Tripathi,S.Predeepkumar, " Statistical analysis of rainfall events at Kalpakkam ",NCRS-2016, December 2016
4. S. Kokilavani , R. PangayarSelvi , S. Panneerselvam and G.A. Dheebakaran, " Trend Analysis of Rainfall Variability in Western Agro Climatic Zone of Tamil Nadu",Department of Physical Science and Information Technology, vol-(12)1,p.p:181-187,2017
5. Mahadevan Palanichamy,RamaswamySankaralingam&Narayanasamy, "Statistical Studies on Rainfall and Time-based Deviations in Precipitation Trends in Vaigai River Basin",Department of Civil Engineering,vol 49(01), p.p:15-23,January 2020
6. S.I. Sridhar and A. Raviraj , "Statistical Trend Analysis of Rainfall in Amaravathi River Basin Using Mann-Kendall Test ",vol 12(1),89-96 ,2017

7. M. Mayilvaganan , P. Vanitha, "Correlation Analysis of Meteorological Data in Region of Tamil Nadu Districts Based On K- Means Clustering Algorithm", Department of Computer Science,vol-3 issue 3 ,may-june 2015
8. M. Nyatuame,V. Owusu-Gyimah and F. Ampaw , "Statistical Analysis of Rainfall Trend for Volta Region in Ghana",vol 2014,IP:203245,2014
9. G Arvind, P Ashok Kumar, S GirishKarthi and C R Suribabu*, "Statistical Analysis of 30 Years Rainfall Data: A Case Study",School of Civil Engineering,2017
10. Muktarun Islam*, Tahmid Imran Imon, AtaurGoniPolash, "Analysis of Rainfall Variability in Sylhet Region of Bangladesh",Dept. of Irrigation and Water Management, ISSN: 2349-4336 (Online) Volume 6, Issue 2,2016
11. V. Rajendran, R. Venkatasubramani& G. Vijayakumar, "Rainfall variation and frequency analysis study in Dharmapuri district, India", Vol. 45 (11), pp. 1560-1565, November 2016.
12. V. Rajendran,A.Nanthini,G.Padmakumar,K.Thopic Ahamed, J.Vijayvarman "Rainfall variation and frequency analysis study in namakkaldistricttamilnadu", volume-6, issue-3, 2019.
13. M Vaidheki, U Arulanandu and A Raj Kumar, " Statistical Analysis of Rainfall data in Tiruchirappalli District ", vol-6(24), 2484-2493 Article CS232048111,2017.
14. Arulmozhi.S and Dr. Prince Arulraj.G, " Rainfall variation and frequency analysis study of Salem district Tamil Nadu",vol 46(01),pp:213-218,January 2017.