ABSTRACT

This project focuses to analyse the monsoonal rainfall (June through September) data over West Bengal. Exploratory and Time Series Analysis is done on this data to understand the overall characteristics as well as to find if is there any trend of monsoonal rainfall which is very crucial for an agro–based economic perspective. The results of this project will be helpful for further study of the prediction aspect.

The exploratory analysis suggests the average daily monsoonal rainfall and tells us the variability in the rainfall between decades and years. Histogram and mean deviation plot, coefficient of variation defines the nature of the distribution of rainfall and the variability between different time period.

Regarding rainfall trend results of different studies shows variation which leads to an unclear picture of rainfall trend. In the study of monsoon rainfall data for several years we can visualize if there is any pattern or not as well as try to understand if there is some trend in the set of data.

There is no as such significance trend in the overall time period of 65 years but there is decadal trends present in the data in different decadal time periods which will be calculated by Mann Kendall test and Sen's Slope. The rough trend is calculated through Moving average method.

Keywords: Rainfall Data, Coefficient of Variation, Mean Deviation Plot, Boxplot, Histogram, Trend Analysis, Mann Kendall Test, Theil Sen's Slope, Moving Average

1] ***INTRODUCTION***

Rainfall is one of the climatological data which is widely analysed for a long time. Analysis of rainfall data is important as it facilitates policy decisions regarding the cropping pattern, sowing date, construction of roads and providing drinking water to urban and rural areas. The areas in the tropical region like West Bengal, '**Monsoon**' is such a season (June-July-August-September) of rainfall that determines the pathway of agricultural economy. The moisture-laden winds reaches the southernmost point of the Indian subcontinent and hit the Western Ghat at the seacoast in Kerela in the **1st week of June**. Due to the topography of Indian peninsula the flow of wind then divide into two parts: The Arabian Sea branch and The Bay of Bengal Branch.

The Arabian Sea branch causes rainfall in Malabar and Konkan coast in the west of the Western Ghats but the east of the Western Ghat still does not receive much rain as wind does not cross the Western Ghats.

Whereas the Bay of Bengal branch floes over the Bay of Bengal, heading towards the West Bengal, picking up more moisture from the Bay of Bengal. The heavy rainfall is observed in the **Indo-Gangetic plain** and eastern Himalayas. The monsoon is widely welcomed and appreciated by city-dwellers as well, for it provides relief from the climax of summer heat in June. However, the roads take a battering every year. Often houses and streets are waterlogged and slums are flooded despite drainage systems.

Actually **most of the annual rainfall occurs during the Monsoon period**. The rainfall during this time has played a major role to boost the agricultural economy of the state. Though the heavy rainfall is caused by another geographical event during Autumn due to the tropical cyclones but most of those rainfalls are destructive in nature and the state's agricultural economy is badly affected by that cyclones.

Here our study focuses mainly on the rainfall occurs during the monsoon period.

The R software is used to do the analysis and to prepare the graphs and charts for interpreting the results.

2] OBJECTIVES OF THE RESEARCH

1. This study aimed to analyse the overall variability and identify the changes in monsoonal rainfall pattern (if any) for the last 6 or 7 decades in the state of West Bengal.
2. For this purpose, an **Exploratory and Time Series Analysis** of the daily average rainfall data during Monsoon in West Bengal across 65 years spanning 1951-2015 has been done in this project.
3. The focus of this study is to analyse the **decadal changes** in the average daily rainfall and the variability in the **7 decades**.(last decade i.e. the analysis for 2011 to 2020 is incomplete due to insufficient data, we have only managed to study over 5 years data in the last decade). Moving average process , Linear Trend and non-parametric test like Mann-Kendall test and Sen's Slope is used to identify and analysed the trend.
4. This project work will helpful for the further study of the prediction aspect.

3] ***DATASET***

For investigating the ***Exploratory and Time Series Analysis of Monsoonal Rainfall data over West Bengal*** the rainfall data is collected from **Indian Meteorological Department**. This daily data is further transformed into seasonal mean rainfall over 1st June to 30th September for 65 years (1951-2015). In other words, we mean the rainfall for 120 days of monsoon period (1st June to 30th September) for each year.

Year	mm
1951	4.56577
1952	5.13797
1953	5.00394
1954	4.68045
1955	4.78166
1956	5.81038
1957	3.92422
1958	4.31816
1959	5.77171
1960	4.34381
1961	4.34126
1962	5.79587
1963	4.49357
1964	4.77004
1965	4.5911
1966	4.0767
1967	4.60275
1968	5.37359
1969	4.60992
1970	5.53956
1971	6.29696
1972	4.02739

Data_Annual_Rain_Tmax_Tmin_1951

Courtesy: Indian Meteorological Department

https://mausam.imd.gov.in/

4] ***METHODOLOGY***

For analysing the Exploratory and Time Series Analysis of Monsoonal Rainfall data we use several tools from descriptive statistics like Mean, Median, Standard Deviation, Maximum and Minimum rainfall and Coefficient of variation. Also various charts and graphs like Boxplot, Histogram, Line Diagram and Mean Deviation Plot is used for understanding and visualise the characteristics of overall dataset. Further we investigate the long term movement of rainfalls over time. Trend analysis is used to investigate such long term movements. To investigate the overall and decadal trend present (if any) in the dataset , the non-parametric test- Mann Kendall test and the Theil-Sen's Slope has been used for long term trend analysis. The last decade have only contained 5 years (2011-2015) data. The relevant details of the foresaid tests and methods are summarized in the following sections

4.A] Exploratory Analysis:

4.A.I] Mean:

The mean, also known as the average, is a basic statistical measure that represents the central tendency of a dataset. It is calculated by summing up all the values in the dataset and dividing the sum by the total number of values.

Mathematically, the mean is computed using the following formula:

mean = (sum of all values) / (total number of values)

The mean is denoted by the symbol μ (mu) for a population mean or $\bar{x}$ (x-bar) for a sample mean. It is commonly used to provide a representative value that summarizes the dataset.

4.A.II] Standard Deviation:

Standard deviation is a measure of the dispersion or variability of a dataset. It quantifies how spread out the values in a dataset are from the mean or average value. In other words, it provides a measure of how much the individual data points deviate from the average.

Mathematically, the standard deviation is calculated as the square root of the variance.

The standard deviation is typically denoted by the symbol σ (sigma) for a population or s for a sample. It has the same unit of measurement as the original data.

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \quad .$$

Interpreting the standard deviation involves understanding the spread or dispersion of the data. A higher standard deviation indicates a greater variability or spread of the data points, while a lower standard deviation indicates less variability and a more clustered dataset.

The standard deviation is an important statistical measure as it provides insights into the variability of the data, allows for comparisons between datasets, and is used in various statistical analyses and modeling techniques.

4.A.III] Mean Deviation Plot:

In statistical analysis, a mean deviation plot is a graphical tool used to visualize the deviations of the data points from a central reference line, typically the mean. It helps to assess the variability and distribution of the data.

<u>Construction:</u>

1. Calculate the mean of the dataset.
2. Compute the absolute deviations of each data point from the mean by subtracting the mean from each data point and taking the absolute value.
3. Plot the absolute deviations on the y-axis against the corresponding data points on the x-axis.
4. Draw a reference line at y=0, representing the mean.

The resulting plot displays the deviations from the mean for each data point. The plot shows whether the data points tend to cluster around the mean or deviate significantly from it. By examining the distribution of the deviations, we can gain insights into the spread, symmetry and potential outliers in the dataset.

4.A.IV] Coefficient of variation (CV%):

The coefficient of variation (CV) statistically measures relative variability in the time series. It shows the individual data position differing from the mean value. Higher values of CV specify the high variability.

$\boldsymbol{C.V} = \frac{\sigma}{\mu} \times \mathbf{100\%}$ **;** where, $\sigma =$ standard deviation of the distribution (data set), $\mu =$ mean of the distribution (dataset).

4.A.V] Histogram:

A histogram is a graphical representation of the distribution of a dataset. It is used to visualised the frequency or count of values within different intervals or bins. The X-axis of a histogram represents the range of values in the dataset, while the Y-axis represents the frequency of values within each bin.

To create a histogram, the dataset is divided into intervals along the X-axis the height of each bar in the histogram represents the frequency of values falling within that particular bin. The width of the bars may vary depending on the range and number of intervals chosen.

Histograms are particularly useful for understanding the shape, central tendency, spread and skewness of the dataset. They provide insights into the distribution and allow us to identify patterns, outliers or, any other interesting characteristics of the data.

4.A.VI] Boxplot:

A boxplot, also known as box whisker plot is a graphical representation of the distribution of a dataset. It provides a visual summary of the minimum, 1^{st} quartile, median, 3^{rd} quartile, maximum of the dataset.

The boxplots consists of several components:

1. Minimum: the smallest value in the dataset, excluding outliers.
2. Maximum: the largest value in the dataset, excluding outliers.
3. 1^{st} quartile (Q1): the 25^{th} percentile of the data, representing the lower boundary of the central 50% of the dataset.
4. 3^{rd} quartile (Q3): the 75^{th} percentile of the data, representing the upper boundary of the central 50% of the dataset.
5. Median(Q2): the middle value of the dataset, separating the lower and upper halves.
6. Inter Quartile Range (IQR): the range between the 1^{st} and 3^{rd} quartile (Q3-Q1), which captures the spread of the central 50% of the data.
7. Whiskers: lines extending from the box representing the range of the data, excluding any outliers. The length of the whiskers typically extends to the minimum and maximum values within a specified range.
8. Outliers: data points that fall outside the whiskers and are considered unusually high or low values compared to rest of the dataset.

Boxplots are usually for visually summarise the distribution of a dataset and identifying key statistical measures such as; spread, symmetry and potential outliers in the data.

By comparing multiple boxplots side by side, we can easily compare the distributions of the different groups or variables and identify any differences or similarities between them.

4.B] Time Series Analysis:

4.B.I] Moving Average Method:

It consists in measurement of trend by smoothing out the fluctuations of the data by means of a moving average. Moving average of extent (or, period) m is a series of successive averages (arithmetic mean) of m terms at a time, starting with 1^{st} , 2^{nd} , 3^{rd} term etc. Thus, the first average is the mean of the 1^{st} m terms; 2^{nd} average is the mean of the n terms from 2^{nd} to $(m+1)^{th}$ term, the third is the mean of the m terms from 3^{rd} to $(m+2)^{th}$ term and so on. If m is odd i.e. $= (2k + 1)$ say, moving averages is placed against the mid-value of the time interval it covers i.e. , against $t = k + 1$ and if m is even $= 2k$ (say), it is placed between the two middle values of the time interval it covers, i.e., between $t = k$ and $t = k + 1$.

The resulting sequence of Moving Average values represent the smoothed values of the original time series. The moving average acts as a filter, reducing noise and short-term fluctuations, thus revealing the underlying trends or patterns in the data.

4.B.II] Linear Trend:

A linear trend is a type of trend observed in a time series dataset where the values change steadily and consistently over time in a linear fashion. It indicates a constant rate of change in the variable being measured.

In a linear trend, the relationship between the independent variable (such as time) and the dependent variable (e.g. rainfall) is linear and can be expressed by a straight line equation. The equation typically takes the form: $\boldsymbol{y = a + b * x}$, where :

- y represent the dependent variable.
- x represent the independent variable (usually time).
- a represent the y interpret, which is the value of y when x is 0.
- b represent the slope of the line, indicating the rat of change of y per unit change in x.

To identify and analyse a linear trend in a time series or dataset, we can perform the following steps:

1. plot the data points on a scatter plot with time (or, the independent variable) on the x-axis and the variable of the interest (dependent variable) on the y- axis.

2. Visually inspect the scatter plot for a general linear pattern or trend in the data points. Look for a clear upward or downward progression.
3. Calculate the least squares regression line that best fits the data. This line represents the linear trend.
4. Assess the strength and significance of the linear trend by analysing the coefficient of determination (R-squared value) and performing hypothesis testing on the slope coefficient.
   - The R-squared value indicates the proportion of the variance in the dependent variable that can be explained by the linear trend.
   - Hypothesis testing on the slope coefficient can help determine if the linear trend is statistically significant.

Interpreting the linear trend involves considering the slope and its significance, as well as the direction of the trend. A positive slope indicates an increasing trend, while a negative slope indicates a decreasing trend. The magnitude of the slope represents the rate of change of the variable over time.

4.B.III] Mann Kandel Test for Monotonic Trend:

<u>Background of the test:</u>

The purpose of the Mann-Kendall (MK) test (Mann 1945, Kendall 1975, Gilbert 1987) is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (downward) trend means that the variable consistently increases (decreases) through time, but the trend may or may not be linear. The MK test can be used in place of a parametric linear regression analysis, which can be used to test if the slope of the estimated linear regression line is different from zero. The regression analysis requires that the residuals from the fitted regression line be normally distributed; an assumption not required by the MK test, that is, the MK test is a non-parametric (distribution-free) test.

Hirsch, Slack and Smith (1982, page 107) indicate that the MK test is best viewed as an exploratory analysis and is most appropriately used to identify stations where changes are significant or of large magnitude and to quantify these findings.

<u>Assumptions:</u>

The following assumptions underlie the MK test:

- When no trend is present, the measurements (observations or data) obtained over time are independent and identically distributed. The assumption of independence means that the observations are not serially correlated over time.

- The observations obtained over time are representative of the true conditions at sampling times.
- The sample collection, handling, and measurement methods provide unbiased and representative observations of the underlying populations over time.

There is no requirement that the measurements be normally distributed or that the trend, if present, is linear. The MK test can be computed if there are missing values and values below the one or more limits of detection (LD), but the performance of the test will be adversely affected by such events. The assumption of independence requires that the time between samples be sufficiently large so that there is no correlation between measurements collected at different times.

Calculation:

The MK test tests whether to reject the null hypothesis (H_0) and accept the alternative hypothesis (H_a), where

H_0: No monotonic trend

H_a: Monotonic trend is present

The initial assumption of the MK test is that the H_0 is true and that the data must be convincing beyond a reasonable doubt before H_0 is rejected and H_a is accepted.

The MK test is conducted as follows (from Gilbert 1987, pp. 209-213) :

1. List the data in the order in which they were collected over time $x_1, x_2, \dots\dots, x_n$, which denote the measurements obtained at times 1, 2, 3,……. ,n respectively.

2. Determine the sign of all $n(n-1)/2$ possible differences $x_j - x_k$, where $j > k$. These differences are:

$$x_2 - x_1, x_3 - x_1, \dots\dots\dots, x_n - x_1, x_3 - x_2, x_4 - x_2, \dots\dots, x_n - x_{n-2}, \quad x_n - x_{n-1}$$

3. Let $\boldsymbol{Sgn}(x_j - x_k)$ be an indicator function that takes on the values 1, 0, or -1 according to the sign of $(x_j - x_k)$, that is,

$$\boldsymbol{Sgn(x_j - x_k)} \quad = \mathbf{1}\ ; \boldsymbol{if}, \boldsymbol{x_j - x_k} > \mathbf{0}$$

$$= \mathbf{0}\ ; \boldsymbol{if}, \boldsymbol{x_j - x_k} = \mathbf{0}\ or\ if\ the\ sign\ of\ x_j - x_k\ cannot\ be\ determined\ due\ to\ non\ detects\ .$$

$$= -\mathbf{1}\,; \boldsymbol{if}, \boldsymbol{x_j} - \boldsymbol{x_k} < \mathbf{0}$$

For example, if $x_j - x_k > 0$, that means that the observation at time j , denoted
by x_j, is greater than the observation at time k, denoted by x_k.

4. Compute

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} Sgn(x_j - x_k) \qquad (1)$$

which is the number of positive differences minus the number of negative differences. If S is a positive number, observations obtained later in time tend to be larger than observations made earlier. If S is a negative number, then observations made later in time tend to be smaller than observations made earlier.

5. If $n \le 10$, follow the procedure described in Gilbert (1987, page 209, Section 16.4.1) by looking up S in a table of probabilities (Gilbert 1987, Table A18, page 272) . If this probability is less than α (the probability of concluding a trend exists when there is none), then reject the null hypothesis and conclude the trend exists. If n cannot be found in the table of probabilities (which can happen if there are tied data values), the next value farther from zero in the table is used. For example, if $S = 12$ and there is no value for $S = 12$ in the table, it is handled the same as $S = 13$.

If $n > 10$, continue with steps 6 through 10 to determine whether a trend exists. This follows the procedure described in Gilbert (1987, page 211, Section 16.4.2).

6. Compute the variance of S as follows:

$$VAR(S) = \frac{1}{18}\left[n(n-1) - \sum_{p=1}^{g} t_p(t_p - 1)(2t_p + 5)\right] \qquad (2)$$

Where, g is the number of tied groups and t_p is the number of observations in the p th group. For example, in the sequence of measurements in time {23, 24, 29, 6, 29, 24, 24, 29, 23} we have $g = 3$ tied groups, for which $t_1 = 2$ for the tied value 23, $t_2 = 3$ for the tied value 24, and $t_3 = 3$ for the tied value 29. When there are ties in the data due to equal values or non-detects, $VAR(S)$ is adjusted by a tie correction method described in Helsel (2005, p. 191) .

7. Compute the MK test statistic, Z_{MK} , as follows:

$$
\begin{aligned}
Z_{MK} &= \frac{S-1}{\sqrt{VAR(S)}} \; ; if\ S > 0 \\
&= 0 \qquad ; if\ S = 0 \\
&= \frac{S+1}{\sqrt{VAR(S)}} \; ; if\ S < 0
\end{aligned} \tag{3}
$$

A positive (negative) value of Z_{MK} indicates that the data tend to increase
(decrease) with time.

8. Suppose we want to test the null hypothesis

H_0: No monotonic trend

versus the alternative hypothesis

H_a:Upward monotonic trend

at the Type I error rate α, where $0 < \alpha < 0.5$. (Note that α is the Level of Significance that the MK test will falsely reject the null hypothesis.) Then H_0 is
rejected and H_a is accepted if $Z_{MK} \geq Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)^{\text{th}}$
percentile of the standard normal distribution. These percentiles are provided in
many statistics book (for example Gilbert 1987, Table A1, page 254) and
statistical software packages.

9. To test H_0 above versus

H_a : Downward monotonic trend

at the Type I error rate α , H_0 is rejected and H_a is accepted
if $Z_{MK} \leq -Z_{1-\alpha}$.

10. To test the H_0 above versus

H_a : Upward or downward monotonic trend

at the Type I error rate α , H_0 is rejected and H_a is accepted
if $|Z_{MK}| \geq Z_{1-\alpha/2}$,
where the vertical bars denote absolute value.

4.B.IV] Sen's Slope:

Sen's slope, also known as the Sen's Estimator or the Sen's Method, is a non-parametric method used to estimate the slope or trend in a dataset. It is particularly useful for analysing Time Series or Spatial data when the underlying distribution may not be known or when the data may contain outliers.

The Sen's slope is calculated by determining the median of all possible slopes between pairs of the data points. This approach makes it robust to outliers and resistant to extreme values. The Sen's slope provides an estimate of the direction and magnitude of the trend in the dataset, regardless of the shape of the distribution.

Here is a simplified explanation of how Sen's slope is calculated:

1 Consider the dataset with n observations, represented as $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ where x_i represents the independent variables (e.g. time) and y_i represent the dependent variable (e.g. Rainfall).
2 Calculate the differences in the dependent variables (Δy) and the differences in the independent variables (Δx) between all pairs of data points.
3 Calculate the slopes (S) between all pairs of the data points using the formula: $S = {(y_i - y_j)}/{(x_i - x_j)}$ $, for\ i < j$. Where i & j represent the indices of two data points.
4 Determine the median of all calculated slopes. This median value represents the Sen's slope, which provides an estimate of the overall trend in the dataset.

The Sen's slope is a robust estimator that is not effected by extreme values or outliers. It is widely used in various fields, including environmental science, hydrology and climatology, to assess trends in variables such as rainfall, temperature, river flow or, pollutant concentrations.

The Sen's slope is often used in conjunction with other statistical methods such as the Mann Kendall test to analyse trends and assess their statistical significance in Time Series or Spatial data.

<u>Interpretation of Sen's Slope:</u>

1 Direction of the Trend: the sign (+ or, -) of the Sen's slope indicates the direction of the trend. A positive slope suggests an increasing trend, while a negative slope suggests a decreasing trend. For example, a positive Sen's Slope for rainfall data would indicate an increasing trend in rainfall over time.

2 Magnitude of the trend: The magnitude of the Sen's slope provides an estimate of the rate or magnitude of the trend. The slope value represents the change in the dependent variable (e.g. rainfall) per unit change in the independent variable (e.g. time). For instance, if the Sen's slope for rainfall data is 10 mm/year, it implies that an average increase of 10 millimetres of rainfall per year.

3 Comparisons with other methods: When interpreting Sen's Slope, it can be useful to compare it with other trend analysis methods, such as linear regression. If the Sen's slope and the slope estimated by linear regression are similar, it provides additional confidence in the trend estimate.

4 Statistical Significance: While Sen's slope provides an estimate of the trend, it does not inherently provide information about the statistical significance of the trend. To determine the statistical significance, additional tests such as the Mann-Kendall test can be performed.

5] ***Analysis & Implementation***

The Monsoonal rainfall for 65 years of the state has been analysed. The statistical parameters like mean, median, standard deviation (S.D) and other statistical tests have been estimated using R software.

5.A] Analysing Rainfall characteristics by Exploratory Analysis during year 1951-2015:

[The relevant codes and their outputs can be found in Appendix-I]

The average Monsoonal rainfall of the state is found as 5.00168 mm/day with a minimum value of 3.70508 mm/day and maximum value of 6.29696 mm/day. The maximum daily average monsoonal rainfall was observed in the year 1971 and minimum daily average Monsoonal rainfall was observed in the year 2012. The standard deviation of the daily average monsoonal rainfall is 0.6438. This indicates that there is less variability i.e. the data points (which is the daily average Monsoonal rainfall for each year) are closer to the mean value 5.00168 mm/day.

To simplify the analysis and compare the variability in daily average Monsoonal rainfall over the period of 65 years (1951 to 2015) we divide the whole data set into 7 groups each containing data for 10 years. Thus we now investigate the decadal exploratory analysis for each 10 years period.

[Note: 7^{th} decade has only consists of 5 observation i.e. 5 years data (2011 to 2015)]

The decades are grouping as follows:

1. First decade – 1951 to 1960.
2. Second decade - 1961 to 1970.
3. Third decade - 1971 to 1980.
4. Fourth decade - 1981 to 1990.
5. Fifth decade - 1991 to 2000.
6. Sixth decade - 2001 to 2010.
7. Seventh decade - 2011 to 2015.

5.A.I] The detailed decadal exploratory analysis of daily average Monsoonal rainfall data is shown in **Table 1**

Table 1

Sl no	Decade	Mean	Median	Standard Deviation	Min Value	Max Value	Coefficient of Variation
1	Overall 65 years	5.0016 8	5.0596	0.6438	3.7050	6.2969	12.8731
2	1951 to 1960	4.8338	4.7311	0.6133	3.9242	5.8103	12.6896
3	1961 to 1970	4.8194	4.6063	0.5588	4.0767	5.7958	11.5961
4	1971 to 1980	5.0886	5.1985	0.7674	4.0103	6.2969	15.0820
5	1981 to 1990	5.2342	5.2387	0.6625	3.7311	6.2406	12.6586
6	1991 to 2000	5.2892	5.2979	0.7317	4.1106	6.2557	13.8345
7	2001 to 2010	4.9719	5.0457	0.4117	4.2259	5.6609	8.2812
8	2011 to 2015	4.5474	4.3292	0.6404	3.7050	5.2071	14.0839

[the detailed code and output is in Appendix-I]

By this table we are able to compare the decadal measurement of statistical parameters with the measurement of parameters of overall time period of 65 years.

Analyse the standard deviation of various decadal periods we can quantify the spread or dispersion in the observations among these seven decades with the overall time period. This allows us to compare between 7 different decades. The third decade shows us the highest standard deviation with 0.7674 that means there is greater dispersion or variability in the daily average Monsoonal Rainfall in this particular decade. Whereas in case of sixth decade i.e. the time period of 2001 to 2011 we have seen a lowest standard deviation with 0.4117 which indicates that there is low dispersion or variability in daily average Monsoonal rainfall in between 2001 to 2011. The other decades shows more or less a homogeneous dispersion or variation in Daily average Monsoonal rainfall as compare to the standard deviation of overall time period of 65 years .

We are also interested to investigate about the relative variability. For which we go for the coefficient of variation. It is typically expressed as a percentage and provides information about the spread of data in relation to the mean. From the table we conclude that the third decade i.e. in 1971 to 1980 we have seen a greater CV that means the greater variability of Daily average rainfall in the monsoonal season of 1971 to 1980. One more important observation can be made from the Coefficient of Variation column of the above table . that is, the CV's are not similar i.e. there is no consistency between any two CV's of corresponding two decades which means there is lack of stability in the monsoonal rainfall between any two decades. A lower CV value indicates more stable rainfall whereas the higher CV value indicates the less stable Monsoonal rainfall in the corresponding decade.

5.A.II] Analysis with Histogram:

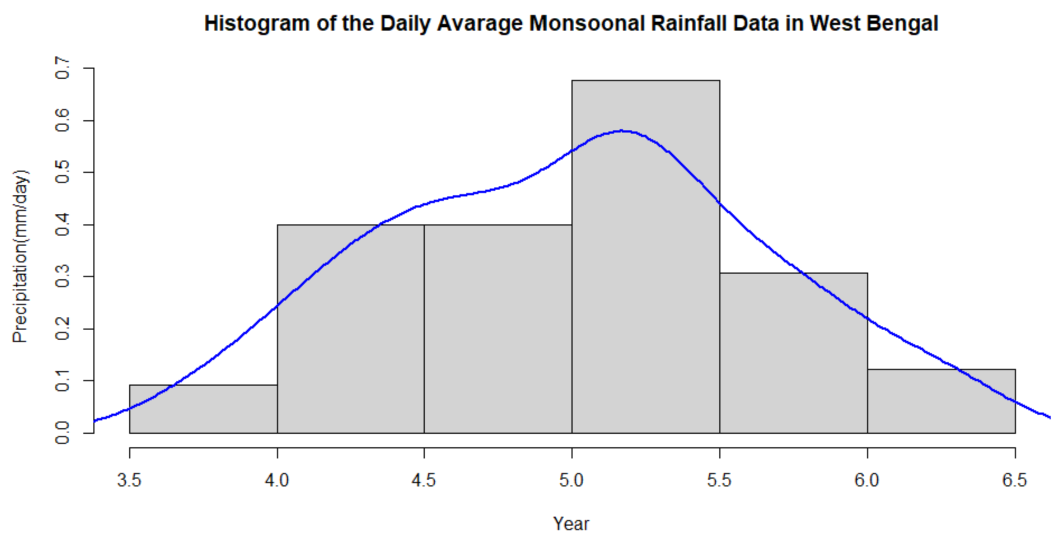

Figure - 1

[Relevant codes are found in Appendix-I]

After investigating the histogram of the dataset of distribution of Daily Average Monsoonal Rainfall in West Bengal over 65 years we gain the following observations:

- First of all the distribution of monsoonal rainfall data of West Bengal over 65 years has a unique mode value hence it is a unimodal distribution.
- The distribution doesn't show symmetry around the center as if it does not resemble like a bell shaped curve
- We can see more or less a positive skewness of the distribution of Daily average monsoonal rainfall over 65 years by the Histogram also.
- The distribution does not show any relative flatness which suggesting that the data is not evenly distributed across the range.

5.A.III] Analysing the Mean Deviation Plot:

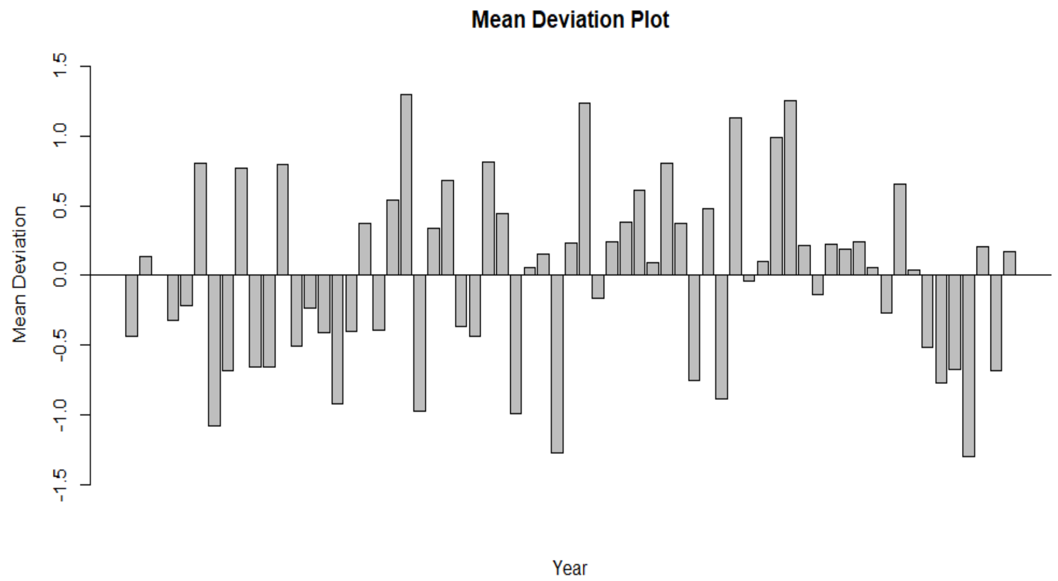

Figure - 2

[codes are found in Appendix-I]

After investigating the mean deviation plot more specifically Mean Absolute Deviation we can conclude the following understandings over the distribution of Daily Average Monsoonal rainfall on West Bengal over 65 years.

The conclusions are as follows:

- Each data points in the plot represent the absolute deviation of an individual observation from the mean value of the dataset. This is nothing but the distance of each point from the zero line (which is mean line) that indicates the magnitude of the deviation.
- We can see the spread of the data points is very high. This wider spread suggest a higher degree of variation indicating that the data points deviate more from the mean. That means we see a huge variation in the Daily average Monsoonal rainfall in West Bengal on the Monsoon Season among the period of 65 years.
- Most importantly we can do a rough estimate that whether there is any presence of trend or not by this mean deviation plot. By examining the pattern of the datapoints we see that there is a wider variation which leads us to a random scatterness of the data points without a clear pattern indicates a lack of systematic variation or simply trend.
  [Note: it is a crude measurement of trend by analysing the mean deviation plot through naked eyes. The detailed analysis of trend is done in the next chapter.].

5.A.IV] ***Analysis with Boxplot***

[Relevant code is found in Appendix-I]

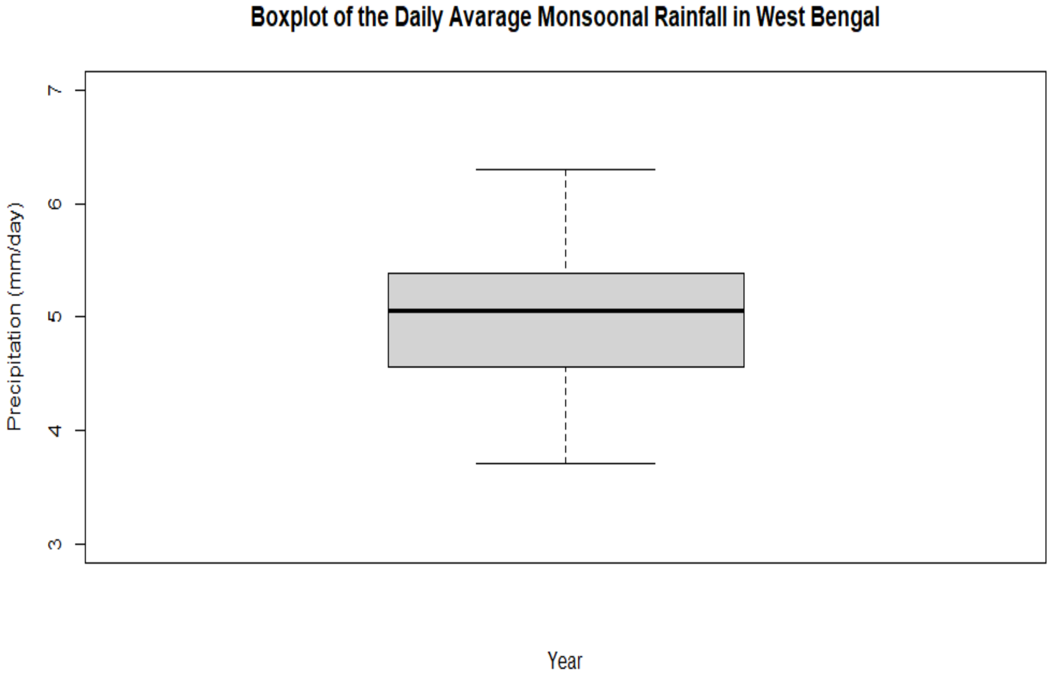

Figure - 3

A boxplot also known as box whisker plot, is a graphical representation of the distribution of a dataset.

By this diagram shown above we can see that the line inside the box represent the median which is the middle value of the whole dataset when the data is sorted in ascending order i.e. the center value of the daily average Monsoonal Rainfall over West Bengal for 65 years.

We can find the summary statistics from the diagram.

The data investigated from the boxplot are shown below:

- The lower extreme value of the dataset of Daily average Monsoonal Rainfall over West Bengal is 3.71 mm/day. i.e. the minimum daily average rainfall recorded is 3.71 mm/day.
- The 1st quartile value is 4.57 mm/day. That means the 75% of the observation of daily average monsoonal rainfalls over 65 years lie above this value.
- The 2nd quartile or the median is 5.06 mm/day. That means the middle most value of the distribution of daily average monsoonal rainfall over 65 years is 5.06 mm/day.
- The 3rd quartile is 5.38 mm/day which means 25% of the total observation of the distribution of daily average monsoonal rainfall over 65 years lie above the value 5.38 mm/day.

- The upper extremum value or upper whisker is 6.30 mm/day i.e. that maximum daily average monsoonal rainfall recorded over 65 years is 6.30 mm/day which is equal to the maximum rainfall given in Table 1.
- Here the number of observation is 65.
- The 4.90 mm/day & 5.22 mm/day suggests that the confidence interval of the median observation. i.e. the median should lie in the range of 4.90 to 5.22.
- According to the diagram there is no outliers in the distribution.[it is a value that lies outside the overall distribution pattern and thus can affect the overall data series. These anomalies are treated as abnormal values that can distort the final insights.]

The box itself represents the Inter Quartile Range (IQR) which is 5.38-4.57=0.81 mm/day. So, the length of the box is 0.81 and this is the measure of the spread of the data more specifically the extent of spread of data.

The whiskers represent the range of the data outside the IQR.

As the median is not in the approximate centre of the box so we can conclude that there is lack of symmetry in this data set of Daily Average Monsoonal Rainfall over 65 years.

The length of upper whisker from the 3rd quartile (0.92) is greater than the length of the lower whisker from the 1st quartile (0.86). That indicates the distribution of daily average rainfall data is positively skewed.

As we know that boxplots are useful for comparing multiple datasets, therefore we divide the whole dataset i.e. the whole time period into 7 decadal time periods and thus we get 7 simultaneous boxplots for 7 decades [Note: last decade contains only 5 observation i.e. data of 2011 to 2015].

5.A.V] The multiple boxplot diagram is shown below:

[codes are in Appendix-I]

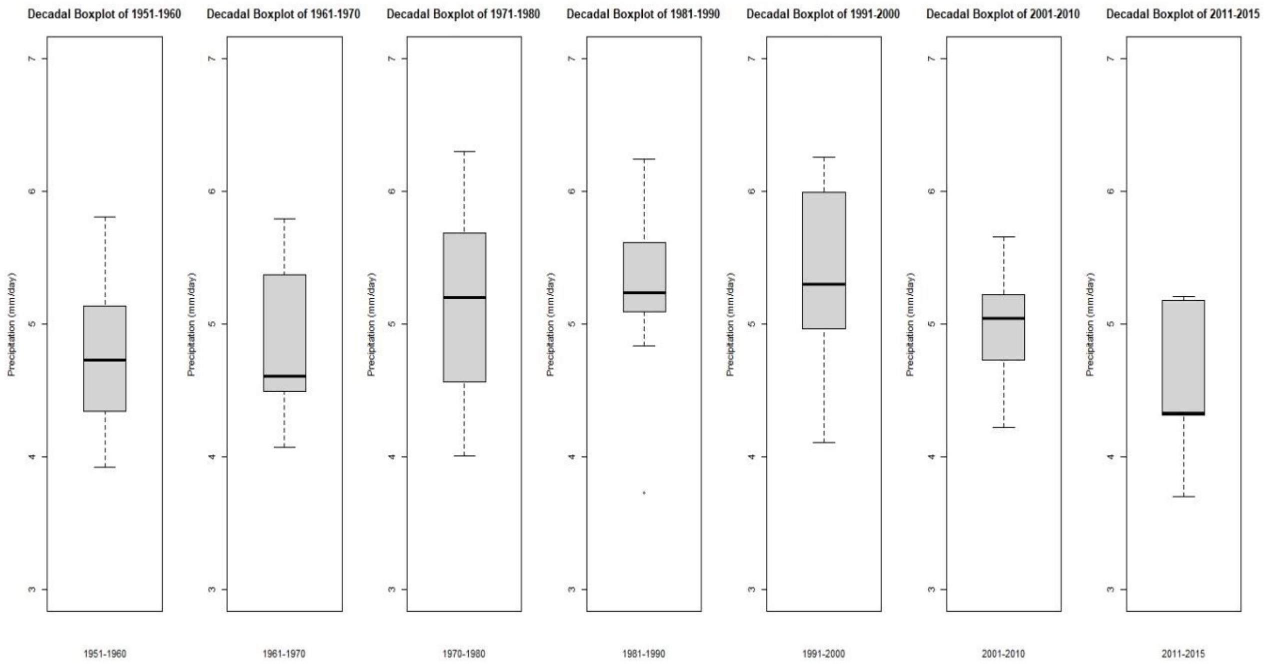

Figure-4

After investigate and comparing the 7 decadal boxplot shown in the figure we can conclude that:

- After comparing the medians of 7 decadal boxplots we see that boxplot consisting data of 1971-80 , '81-'90 & 91-2000 have greater median with respect to other decades. That indicates a higher central tendency in those decades in on the daily average monsoonal rainfall.
- Similarly it can be found that the decades 1971-1980 and 1991-2000 have the larger length of the boxes which suggests the more spread in their distribution of Daily rainfall average in monsoonal season. i.e. in those decades there should be potentially more variability in the Daily average rainfall in monsoons.

5.B] Time Series Analysis over the year 1951-2015

5.B.I] Moving Average analysis for trend in Rainfall data:

[codes are found in Appendix-I]

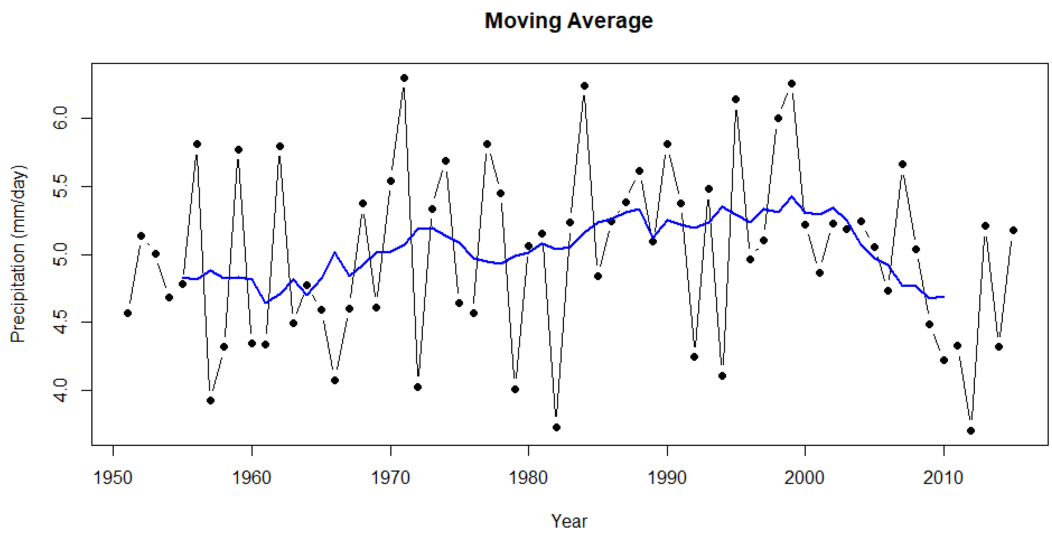

Figure-5

The primary purpose of the moving average is to smooth out the data by reducing short term fluctuations. By this smoothing moving average helps us to identify the trend by removing random noise. After plotting moving average line we can see that there is no straight forward slope or we can't get any decresing or increasing trend line into the overall trend. And even the line is not flat to conclude the lack of trend. So we conclude that there is a certain trend but in many different time period intervals. To find the more specific trend we should go for linear trend analysis and know about the exact magnitude and location of the trend in daily average rainfall data of monsoon period we have to check MK Test and Sen's Slope.

5.B.II] Monsoonal Rainfall Trend Analysis of West Bengal:

To calculate the rough estimate whether there is any presence of any trend or not we first draw a line diagram & then fit a trend equation. This leads us to this results:

[Relevant codes are in Appendix-I]

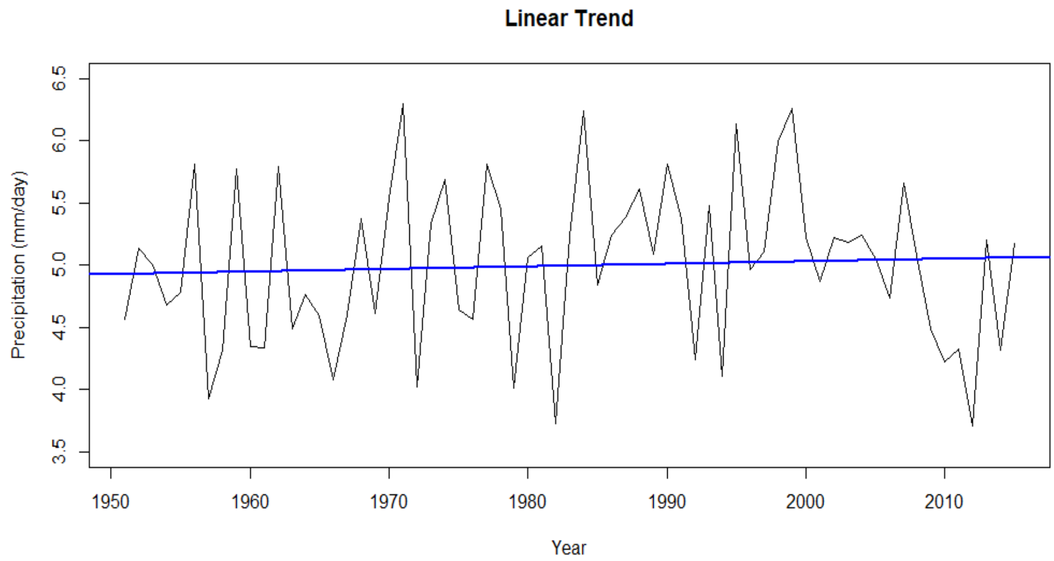

Figure - 6

From **Figure-6 -** the line diagram and fitted regression line we can conclude that there is no presence of any kind of trend in the given dataset of Daily Average Monsoonal Rainfall over West Bengal for 65 years. But this is again a rough conclusion as if the trend analysis of this data tells us a certain different story.

- This fitted regression line has a intercept of 0.8284 and the regression coefficient between two variables i.e. Precipitation (mm/day) and time period (Year) is 0.0021045 which is though very small but we say that there is a certain correlation between daily average monsoonal rainfall (Precipitation(mm/day)) and time periods (Year).
- Hence the fitted regression equation is given by: $\boldsymbol{y = 0.82846 + 0.0021045 * x}$; where y is the dependent variable i.e. average monsoonal rainfall [precipitation(mm/day)] , x is the dependent variable i.e. the time period (Year)
- For justification of this equation just put the value 1951 in place of x and then we get 4.924 mm/day rainfall which is approximately equal to the average monsoonal rainfall recorded in the year 1951.

To get a more specific result for gaining the trend equation we go for moving average method. For analysing moving average method using R software we have to install a pakage named “zoo”.

5.B.III] Decadal & Overall Monsoonal Rainfall Trend analysis by Mann Kendall Test & Sen's Slope:

From the previous trend analysis by Moving average method we can see that there is a trend in the daily average monsoonal rainfall data over 65 years time period but we are unable then to take any further dicision as what is the magnitude of the trend or where is actually the trend starts from. More precisely we can't get the direction of the trend by all the previous measures.

To measure the strength and direction of the trend we now going to use the Mann Kendall Test as well as Theil Sen's Slope one by one for overall 65 years time period data set and also we analyze the decadal trend by these two techniques.

To analyse the decadal trend and overall trend by using MK Test & Sen's Slope using R software we have to download another pakage named – "trend".

Analysis of trend for overall data for time period 1951-2015 using MK Test & Sen's Slope:

To test whether there is any presence of trend or not we assume the non-parametric inference.

Let H_0 : there is no significant trend in the data set.

Vs. H_a : there is the presence of significant trend.

We now test the hypothesis by the test statistics given in (3).

[The Codes and outputs are in Appendix-I]

The value of the test statistic $Z_{MK} = 0.39064$ and the p-value = 0.6961.

Decision Rule: alternative hypothesis is true and S is not equal to 0 under 5% level of significance. [denoted by α].

And this result makes us understand that though there is no significance trend is detected during the investigation through naked eye, there is actually a trend present hiding behind the rough estimates. [this is just like Simpson Paradox].

Now we want to estimate the magnitude of the trend using Sen's Slope;

[codes for calculating Theil Sen's Slope are in Appendix-I]

The Sen's Slope is 0.002322 which is approximately equal to the regression coefficient we got in the previous analysis.

The positive value indicates the direction i.e. an increasing trend of monsoonal rainfall.

The magnitude of the Sen's slope provides the estimate of the rate or magnitude of the trend. i.e. it indicates an average increase of 0.002322 mm/day rainfall per year in the monsoon season.

[Note: the Sen's Slope and the slope estimated by the linear regression are approximately equal, this gives the additional confidence in favour of an increasing trend in Daily Average Monsoonal rainfall data over West Bengal.]

We now go for the decadal trend calculation in a similar way. To do so we have to chose the null and alternative hypothesis and test statistics as done before. The values of test statistics , p-value , decision rule , Sen's Slope and conclusion for 7 different decades with the overall calculation are shown in the **Table 2** below:

Table - 2

Sl. No	Time Period	Test Statistic Value	p-value	Decision Rule	Sen's Slope	conclusion
1	1951 to 1960	-0.3577	0.7205	Accept null hypothesis; there is no significant trend at 95% confidence interval.	-0.0353	An average decrease of 0.0353 mm rainfall per Monsoon. (an insignificant negative trend).
2	1961 to 1970	1.0733	0.2831	-do-	0.04358	An average increase of 0.04353 mm rainfall / monsoon. (an insignificant positive trend).
3	1971 to 1980	-0.7154	0.4743	-do-	-0.0805	An average decrease of 0.0805 mm rainfall / monsoon. (an insignificant negative trend).

4	1981 to 1990	1.4311	0.1524	-do-	0.0764	An average increase of 0.0764 mm rainfall / monsoon. (an insignificant positive trend).
5	1991 to 2000	1.0733	0.2831	-do-	0.1101	An average increase of 0.1101 mm rainfall / monsoon. (an insignificant positive trend).
6	2001 to 2010	-1.4311	0.1524	-do-	-0.0661	An average decrease of 0.0661 mm rainfall / monsoon. (an insignificant negative trend).
7	2011 to 2015	0.2449	0.8065	-do-	0.2593	An average increase of 0.2593 mm rainfall / monsoon. (an insignificant positive trend).
Overall data	1951 to 2015	0.3906	0.6961	-do-	0.0023	An average increase of 0.0023 mm rainfall / monsoon. (an insignificant positive trend)

[Codes for creating this table can be found in Appendix-I]

5.B.IV] GRAPHICAL REPRESENTATION OF DECADAL TREND

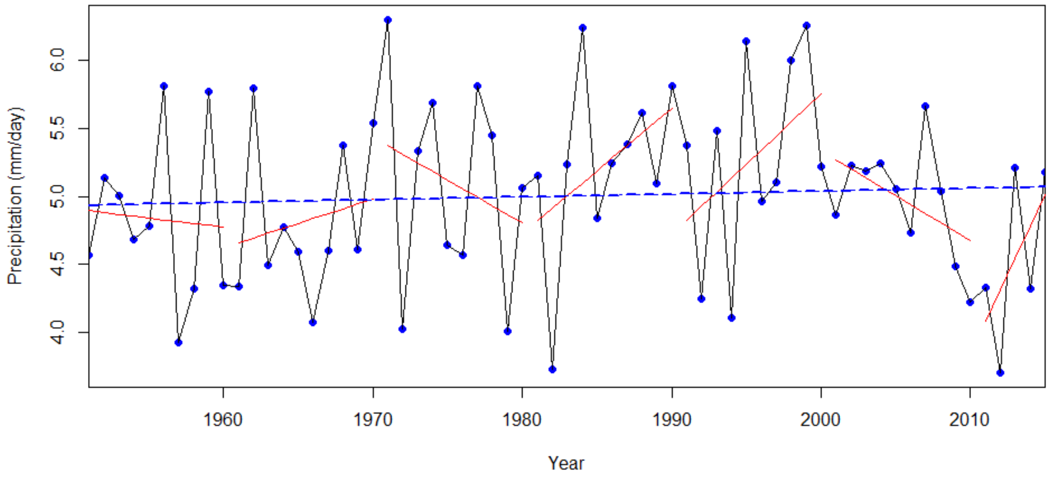

Figure-7

The above diagram shows the all decadal trend lines with overall trend line of the Daily Average Monsoonal Rainfall Data of West Bengal Over 65 years.

[The relevant codes are in Appendix-I]

6] *CONCLUSION*

The following are some important conclusion derived from the study.

- The average monsoonal rainfall of the state is 5.00168 mm / day with standard deviation ± 0.6438 mm/day.
- The maximum monsoonal rainfall was recorded in 1971 which is 6.29696 mm/day and the minimum monsoonal rainfall was recorded in the year 2012 which is 3.70508 mm/day.
- Coefficient of Variation and standard deviation among seven decades shows that there was wide variability in the monsoonal rainfall.
- The boxplot suggests inter quartile range is within 1.5 mm/day.
- The multiple boxplot for each decade suggests the wider spread more precisely the greater variation in monsoonal rainfall among every decades as well as overall time period of 65 years.
- The histogram of monsoonal rainfall data shows us that the distribution of daily average monsoonal rainfall data of West Bengal is a unimodal distribution.
- Mean deviation plot also suggest for a huge variation in rainfall during monsoonal season for different years.
  By simple linear regression analysis and trend fitting we see that there is a negligible trend presence in the overall data but we can't sure about the magnitude and location of the trend though we predict the Monsoonal rainfall by regression analysis and linear trend equation which is $y = 0.82846 + 0.0021045 * x$; where y is the dependent variable i.e. average monsoonal rainfall [precipitation(mm/day)] , x is the dependent variable i.e. the time period (Year) .
- Thus we go for moving average method to get detailed idea about trend curve and saw that there is several location where the trend curve has fluctuated.
- Hence Mann Kendall Test and Sen's Slope suggest there is insignificance trend in the distribution. And the magnitude of the trend is given by 0.002322 by using Sen's slope formula.
- The Sen's Slope and the slope estimated by the linear regression are approximately equal, this gives the additional confidence in favour of an increasing trend in Daily Average Monsoonal rainfall data over West Bengal.
- The decadal trend analysis is given in a similar way by using MK test and Sen's Slope , the magnitude of the trend are given in the table, and the graphs shows the decadal trends as well as overall trend.
- Though there was very high decadal trends (increasing or decreasing) shown in the figures but the overall trend is not so high.

This is just because of the very high trend for some decades balancing the whole trend with some very low trends for some decades. Results the overall trend is to small in magnitude.

As the monsoonal rainfall plays the most vital role in the hydrological analysis and water balance studies, the results of this study may provide useful inputs for the planning and management in context of agricultural and water resource areas.

7] APPENDIX – I

Code & Output: [for exploratory analysis]

```
> Monsoonal_Rainfall_Data<-read.csv(file="E:\\EDUCATION\\6th
Semestar\\PROJECT\\Data_Annual_Rain_Tmax_Tmin_1951-2015 (2).csv",head=TRUE)
> Monsoonal_Rainfall_Data
ï..Year     mm
1    1951 4.56577
2    1952 5.13797
> M<-mean(Monsoonal_Rainfall_Data$mm)
> M.D<-median(Monsoonal_Rainfall_Data$mm)
> STD<-sd(Monsoonal_Rainfall_Data$mm)
> Minimum_rainfall<-min(Monsoonal_Rainfall_Data$mm)
> Maximum_rainfall<-max(Monsoonal_Rainfall_Data$mm)
> Range<- range(Monsoonal_Rainfall_Data$mm)
> cv<-STD/M*100
> M
[1] 5.001682
> M.D
[1] 5.05964
> STD
[1] 0.643874
> Minimum_rainfall
[1] 3.70508
> Maximum_rainfall
[1] 6.29696
> Range
[1] 3.70508 6.29696
> cv
[1] 12.87315
```

Code and output [for decadal exploratory analysis]

#1st decadal analysis

```
> M1<-mean(Monsoonal_Rainfall_Data$mm[1:10])
> M.D1<-median(Monsoonal_Rainfall_Data$mm[1:10])
> STD1<-sd(Monsoonal_Rainfall_Data$mm[1:10])
> Minimum_rainfall1<-min(Monsoonal_Rainfall_Data$mm[1:10])
> Maximum_rainfall1<-max(Monsoonal_Rainfall_Data$mm[1:10])
> cv1<-STD1/M1*100
> M1
[1] 4.833807
> M.D1
[1] 4.731055
> STD1
[1] 0.613391
> Minimum_rainfall1
[1] 3.92422
```

```
> Maximum_rainfall1
[1] 5.81038
> cv1
[1] 12.6896
```

#2nd decadal analysis

```
> M2<-mean(Monsoonal_Rainfall_Data$mm[11:20])
> M.D2<-median(Monsoonal_Rainfall_Data$mm[11:20])
> STD2<-sd(Monsoonal_Rainfall_Data$mm[11:20])
> Minimum_rainfall2<-min(Monsoonal_Rainfall_Data$mm[11:20])
> Maximum_rainfall2<-max(Monsoonal_Rainfall_Data$mm[11:20])
> cv2<-STD2/M2*100
> M2
[1] 4.819436
> M.D2
[1] 4.606335
> STD2
[1] 0.5588692
> Minimum_rainfall2
[1] 4.0767
> Maximum_rainfall2
[1] 5.79587
> cv2
[1] 11.59615
```

#3rd decadal analysis

```
> M3<-mean(Monsoonal_Rainfall_Data$mm[21:30])
> M.D3<-median(Monsoonal_Rainfall_Data$mm[21:30])
> STD3<-sd(Monsoonal_Rainfall_Data$mm[21:30])
> Minimum_rainfall3<-min(Monsoonal_Rainfall_Data$mm[21:30])
> Maximum_rainfall3<-max(Monsoonal_Rainfall_Data$mm[21:30])
> cv3<-STD3/M3*100
> M3
[1] 5.088605
> M.D3
[1] 5.19851
> STD3
[1] 0.767464
> Minimum_rainfall3
[1] 4.01034
> Maximum_rainfall3
[1] 6.29696
> cv3
[1] 15.08201
```

#4th decadal analysis

```
> M4<-mean(Monsoonal_Rainfall_Data$mm[31:40])
> M.D4<-median(Monsoonal_Rainfall_Data$mm[31:40])
> STD4<-sd(Monsoonal_Rainfall_Data$mm[31:40])
> Minimum_rainfall4<-min(Monsoonal_Rainfall_Data$mm[31:40])
> Maximum_rainfall4<-max(Monsoonal_Rainfall_Data$mm[31:40])
> cv4<-STD4/M4*100
> M4
```

```
[1] 5.23426
> M.D4
[1] 5.238745
> STD4
[1] 0.6625863
> Minimum_rainfall4
[1] 3.73117
> Maximum_rainfall4
[1] 6.24032
> cv4
[1] 12.65864
```

#5th decadal analysis

```
> M5<-mean(Monsoonal_Rainfall_Data$mm[41:50])
> M.D5<-median(Monsoonal_Rainfall_Data$mm[41:50])
> STD5<-sd(Monsoonal_Rainfall_Data$mm[41:50])
> Minimum_rainfall5<-min(Monsoonal_Rainfall_Data$mm[41:50])
> Maximum_rainfall5<-max(Monsoonal_Rainfall_Data$mm[41:50])
> cv5<-STD5/M5*100
> M5
[1] 5.28921
> M.D5
[1] 5.29797
> STD5
[1] 0.7317407
> Minimum_rainfall5
[1] 4.11068
> Maximum_rainfall5
[1] 6.25576
> cv5
[1] 13.83459
```

#6th decadal analysis

```
> M6<-mean(Monsoonal_Rainfall_Data$mm[51:60])
> M.D6<-median(Monsoonal_Rainfall_Data$mm[51:60])
> STD6<-sd(Monsoonal_Rainfall_Data$mm[51:60])
> Minimum_rainfall6<-min(Monsoonal_Rainfall_Data$mm[51:60])
> Maximum_rainfall6<-max(Monsoonal_Rainfall_Data$mm[51:60])
> cv6<-STD6/M6*100
> M6
[1] 4.971905
> M.D6
[1] 5.04572
> STD6
[1] 0.411734
> Minimum_rainfall6
[1] 4.22592
> Maximum_rainfall6
[1] 5.66095
> cv6
[1] 8.281212
```

#7th decadal analysis

```
> M7<-mean(Monsoonal_Rainfall_Data$mm[61:65])
> M.D7<-median(Monsoonal_Rainfall_Data$mm[61:65])
```

```
> STD7<-sd(Monsoonal_Rainfall_Data$mm[61:65])
> Minimum_rainfall7<-min(Monsoonal_Rainfall_Data$mm[61:65])
> Maximum_rainfall7<-max(Monsoonal_Rainfall_Data$mm[61:65])
> cv7<-STD7/M7*100
> M7
[1] 4.547418
> M.D7
[1] 4.32924
> STD7
[1] 0.6404575
> Minimum_rainfall7
[1] 3.70508
> Maximum_rainfall7
[1] 5.20716
> cv7
[1] 14.08398

> ##decadal_exploratory_analysis
> Mean<-c(M1,M2,M3,M4,M5,M6,M7,M)
> Median<-c(M.D1,M.D2,M.D3,M.D4,M.D5,M.D6,M.D7,M.D)
> Standard_Deviation<-c(STD1,STD2,STD3,STD4,STD5,STD6,STD7,STD)
> Max_value<-
c(Maximum_rainfall1,Maximum_rainfall2,Maximum_rainfall3,Maximum_rainfall4,Maximum_rainfall
5,Maximum_rainfall6,Maximum_rainfall7,Maximum_rainfall)
> Min_value<-
c(Minimum_rainfall1,Minimum_rainfall2,Minimum_rainfall3,Minimum_rainfall4,Minimum_rainfall5,
Minimum_rainfall6,Minimum_rainfall7,Minimum_rainfall)
> Coefficient_of_Variation<-c(cv1,cv2,cv3,cv4,cv5,cv6,cv7,cv)
> data.frame(Mean,Median,Standard_Deviation,Max_value,Min_value,Coefficient_of_Variation)
     Mean   Median Standard_Deviation Max_value Min_value Coefficient_of_Variation
1 4.833807 4.731055          0.6133910   5.81038   3.92422                12.689604
2 4.819436 4.606335          0.5588692   5.79587   4.07670                11.596154
3 5.088605 5.198510          0.7674640   6.29696   4.01034                15.082011
4 5.234260 5.238745          0.6625863   6.24032   3.73117                12.658643
5 5.289210 5.297970          0.7317407   6.25576   4.11068                13.834593
6 4.971905 5.045720          0.4117340   5.66095   4.22592                 8.281212
7 4.547418 4.329240          0.6404575   5.20716   3.70508                14.083982
8 5.001682 5.059640          0.6438740   6.29696   3.70508                12.873150
```

Codes for analysis of histogram

```
> hist(Monsoonal_Rainfall_Data$mm,probability =
TRUE,xlab="Year",ylab="Precipitation(mm/day)",main="Histogram of the Daily Avarage Monsoonal
Rainfall Data in West Bengal")
> lines(density(Monsoonal_Rainfall_Data$mm),lwd=2,col="blue")
```

Codes for analysis of mean deviation plot

```
> Mean_Deviation<-Monsoonal_Rainfall_Data$mm-M
> barplot(Mean_Deviation,width = 1,ylim=c(-1.5,1.5),xlab="Year",ylab="Mean
Deviation",main="Mean Deviation Plot")
> abline(0,0)
```

Codes for analysis of boxplot and multiple boxplot

```
> boxplot(Monsoonal_Rainfall_Data$mm,ylim=c(3,7),xlab="Year", ylab="Precipitation
(mm/day)",main="Boxplot of the Daily Avarage Monsoonal Rainfall in West Bengal")
```

Decadal analysis of multiple boxplot

```
> par(mfrow=c(1,7))
> boxplot(Monsoonal_Rainfall_Data$mm[1:10],ylim=c(3,7),xlab="1951-1960", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 1951-1960")
> boxplot(Monsoonal_Rainfall_Data$mm[11:20],ylim=c(3,7),xlab="1961-1970", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 1961-1970")
> boxplot(Monsoonal_Rainfall_Data$mm[21:30],ylim=c(3,7),xlab="1970-1980", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 1971-1980")
> boxplot(Monsoonal_Rainfall_Data$mm[31:40],ylim=c(3,7),xlab="1981-1990", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 1981-1990")
> boxplot(Monsoonal_Rainfall_Data$mm[41:50],ylim=c(3,7),xlab="1991-2000", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 1991-2000")
> boxplot(Monsoonal_Rainfall_Data$mm[51:60],ylim=c(3,7),xlab="2001-2010", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 2001-2010")
> boxplot(Monsoonal_Rainfall_Data$mm[61:65],ylim=c(3,7),xlab="2011-2015", ylab="Precipitation
(mm/day)",main="Decadal Boxplot of 2011-2015")

> info<-boxplot(Monsoonal_Rainfall_Data$mm,ylim=c(3,7),xlab="Year", ylab="Precipitation
(mm/day)",main="Boxplot of the Daily Avarage Monsoonal Rainfall in West Bengal")
> info
$stats
        [,1]
[1,] 3.70508
[2,] 4.56566
[3,] 5.05964
[4,] 5.38378
[5,] 6.29696

$n
[1] 65

$conf
         [,1]
[1,] 4.899309
[2,] 5.219971

$out
numeric(0)

$group
numeric(0)

$names
[1] ""
```

Codes and outputs of Trend Analysis

Codes for analysis of moving average

```
> install.packages("zoo")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and
install the appropriate version of Rtools before proceeding:
```

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/Nilanjan Barik/Documents/R/win-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/zoo_1.8-12.zip'
Content type 'application/zip' length 1032683 bytes (1008 KB)
downloaded 1008 KB

package ‘zoo’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\Nilanjan Barik\AppData\Local\Temp\RtmpmuXNDF\downloaded_packages
> library(zoo) ##install library 'Zoo'

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

  as.Date, as.Date.numeric

Warning message:
package ‘zoo’ was built under R version 4.1.3

> MA <- rollmean(Monsoonal_Rainfall_Data$mm, k = 10)
>
plot(Monsoonal_Rainfall_Data$ï..Year,Monsoonal_Rainfall_Data$mm,type="b",bg="blue",pch=16,xlab="Year",ylab="Precipitation (mm/day)",main="Moving Average")
> lines(Monsoonal_Rainfall_Data$ï..Year[5:60],MA, col="blue",type = "l", lwd=2)

Codes for linear trend fitting and analysis

> fit <- glm(Monsoonal_Rainfall_Data$mm~Monsoonal_Rainfall_Data$ï..Year)
> co <- coef(fit)
>
plot(Monsoonal_Rainfall_Data$ï..Year,Monsoonal_Rainfall_Data$mm,,type="l",xlim=c(1951,2015),ylim=c(3.5,6.5),xlab="Year",ylab="Precipitation (mm/day)",main="Linear Trend")
> abline(fit, col="blue", lwd=2)
> co
    (Intercept) Monsoonal_Rainfall_Data$ï..Year
    0.828463546   0.002104497

Codes and outputs for Mann Kendall Test

> install.packages("trend")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/Nilanjan Barik/Documents/R/win-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/trend_1.1.5.zip'
Content type 'application/zip' length 705999 bytes (689 KB)
downloaded 689 KB

package ‘trend’ successfully unpacked and MD5 sums checked

```
The downloaded binary packages are in
        C:\Users\Nilanjan Barik\AppData\Local\Temp\RtmpmuXNDF\downloaded_packages
> library(trend)
Warning message:
package ‘trend’ was built under R version 4.1.3
> mk.test(Monsoonal_Rainfall_Data$mm, alternative = c("two.sided"), continuity = TRUE)

        Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm
z = 0.39064, n = 65, p-value = 0.6961
alternative hypothesis: true S is not equal to 0
sample estimates:
           S         varS          tau
7.000000e+01 3.120000e+04 3.365385e-02
```

Codes and outputs for Sen’s Slope

```
> sens.slope(Monsoonal_Rainfall_Data$mm)

        Sen's slope

data:  Monsoonal_Rainfall_Data$mm
z = 0.39064, n = 65, p-value = 0.6961
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.007368649  0.011885789
sample estimates:
Sen's slope
0.002322014
```

For decadal trend analysis

```
> ## First decade trend
> fit1 <- glm(Monsoonal_Rainfall_Data$mm[1:10]~Monsoonal_Rainfall_Data$ï..Year[1:10])
> co1 <- coef(fit1)
> mk.test(Monsoonal_Rainfall_Data$mm[1:10], alternative = c("two.sided"), continuity = TRUE)

        Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[1:10]
z = -0.35777, n = 10, p-value = 0.7205
alternative hypothesis: true S is not equal to 0
sample estimates:
          S        varS         tau
 -5.0000000 125.0000000  -0.1111111

> sens.slope(Monsoonal_Rainfall_Data$mm[1:10])

        Sen's slope

data:  Monsoonal_Rainfall_Data$mm[1:10]
z = -0.35777, n = 10, p-value = 0.7205
alternative hypothesis: true z is not equal to 0
```

```
95 percent confidence interval:
 -0.1371560  0.1681025
sample estimates:
Sen's slope
-0.03537286

> ## Second decade trend
> fit2 <- glm(Monsoonal_Rainfall_Data$mm[11:20]~Monsoonal_Rainfall_Data$ï..Year[11:20])
> co2 <- coef(fit2)
> mk.test(Monsoonal_Rainfall_Data$mm[11:20], alternative = c("two.sided"), continuity = TRUE)

	Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[11:20]
z = 1.0733, n = 10, p-value = 0.2831
alternative hypothesis: true S is not equal to 0
sample estimates:
        S        varS         tau 
 13.0000000 125.0000000   0.2888889 

> sens.slope(Monsoonal_Rainfall_Data$mm[11:20])

	Sen's slope

data:  Monsoonal_Rainfall_Data$mm[11:20]
z = 1.0733, n = 10, p-value = 0.2831
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.070380  0.176004
sample estimates:
Sen's slope 
 0.04358167 

> ## Third decade trend \
> fit3 <- glm(Monsoonal_Rainfall_Data$mm[21:30]~Monsoonal_Rainfall_Data$ï..Year[21:30])
> co3 <- coef(fit3)
> mk.test(Monsoonal_Rainfall_Data$mm[21:30], alternative = c("two.sided"), continuity = TRUE)

	Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[21:30]
z = -0.71554, n = 10, p-value = 0.4743
alternative hypothesis: true S is not equal to 0
sample estimates:
    S  varS   tau 
 -9.0 125.0  -0.2 

> sens.slope(Monsoonal_Rainfall_Data$mm[21:30])

	Sen's slope

data:  Monsoonal_Rainfall_Data$mm[21:30]
z = -0.71554, n = 10, p-value = 0.4743
```

```
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.2858275  0.1345675
sample estimates:
Sen's slope
-0.08054333

> ## Fourth decade trend
> fit4 <- glm(Monsoonal_Rainfall_Data$mm[31:40]~Monsoonal_Rainfall_Data$ï..Year[31:40])
> co4 <- coef(fit4)
> mk.test(Monsoonal_Rainfall_Data$mm[31:40], alternative = c("two.sided"), continuity = TRUE)

        Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[31:40]
z = 1.4311, n = 10, p-value = 0.1524
alternative hypothesis: true S is not equal to 0
sample estimates:
          S        varS         tau
 17.0000000 125.0000000   0.3777778

> sens.slope(Monsoonal_Rainfall_Data$mm[31:40])

        Sen's slope

data:  Monsoonal_Rainfall_Data$mm[31:40]
z = 1.4311, n = 10, p-value = 0.1524
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.0504600  0.2597262
sample estimates:
Sen's slope
   0.076474

> ## Fifth decade trend
> fit5 <- glm(Monsoonal_Rainfall_Data$mm[41:50]~Monsoonal_Rainfall_Data$ï..Year[41:50])
> co5 <- coef(fit5)
> mk.test(Monsoonal_Rainfall_Data$mm[41:50], alternative = c("two.sided"), continuity = TRUE)

        Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[41:50]
z = 1.0733, n = 10, p-value = 0.2831
alternative hypothesis: true S is not equal to 0
sample estimates:
          S        varS         tau
 13.0000000 125.0000000   0.2888889

> sens.slope(Monsoonal_Rainfall_Data$mm[41:50])

        Sen's slope

data:  Monsoonal_Rainfall_Data$mm[41:50]
```

```
z = 1.0733, n = 10, p-value = 0.2831
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.067840  0.328325
sample estimates:
Sen's slope
  0.1100713

> ## Sixth decade trend
> fit6 <- glm(Monsoonal_Rainfall_Data$mm[51:60]~Monsoonal_Rainfall_Data$ï..Year[51:60])
> co6 <- coef(fit6)
> mk.test(Monsoonal_Rainfall_Data$mm[51:60], alternative = c("two.sided"), continuity = TRUE)

	Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[51:60]
z = -1.4311, n = 10, p-value = 0.1524
alternative hypothesis: true S is not equal to 0
sample estimates:
         S        varS         tau
-17.0000000 125.0000000  -0.3777778

> sens.slope(Monsoonal_Rainfall_Data$mm[51:60])

	Sen's slope

data:  Monsoonal_Rainfall_Data$mm[51:60]
z = -1.4311, n = 10, p-value = 0.1524
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.1510667  0.0469350
sample estimates:
Sen's slope
   -0.06611

> ## Last decade trend
> fit7 <- glm(Monsoonal_Rainfall_Data$mm[61:65]~Monsoonal_Rainfall_Data$ï..Year[61:65])
> co7 <- coef(fit7)
> mk.test(Monsoonal_Rainfall_Data$mm[61:65], alternative = c("two.sided"), continuity = TRUE)

	Mann-Kendall trend test

data:  Monsoonal_Rainfall_Data$mm[61:65]
z = 0.24495, n = 5, p-value = 0.8065
alternative hypothesis: true S is not equal to 0
sample estimates:
       S     varS      tau
 2.00000 16.66667  0.20000

> sens.slope(Monsoonal_Rainfall_Data$mm[61:65])

	Sen's slope
```

```
data:  Monsoonal_Rainfall_Data$mm[61:65]
z = 0.24495, n = 5, p-value = 0.8065
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
 -0.88833  1.50208
sample estimates:
Sen's slope
   0.25938
```

Codes for fitting trend lines for 7 decadal trend analysis

```
> plot(Monsoonal_Rainfall_Data$ï..Year, Monsoonal_Rainfall_Data$mm, xlab = "Year", ylab =
"Precipitation (mm/day)",
+      col = "blue",bg = "blue", pch = 16, cex=1,xaxs = "i")
> lines(Monsoonal_Rainfall_Data$ï..Year, Monsoonal_Rainfall_Data$mm, col = "black")
> lines(Monsoonal_Rainfall_Data$ï..Year[1:10], predict(fit1), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[11:20], predict(fit2), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[21:30], predict(fit3), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[31:40], predict(fit4), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[41:50], predict(fit5), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[51:60], predict(fit6), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year[61:65], predict(fit7), col = "red")
> lines(Monsoonal_Rainfall_Data$ï..Year, predict(fit), col = "blue", lty = 2,lwd=2)
```