**Methodology**

**3.1 Introduction**

In this chapter, the technique that was utilized to address each of the study questions is broken out into very specific detail. The challenges surrounding the acquisition of secondary data, as well as their sources, are discussed, and an in-depth explanation of the empirical models that were used to evaluate the data is provided. In light of this, the following is the structure of this chapter: In Section 3.2, the approach that was utilized to measure the effect that precipitation and its shifting pattern had on the environment is outlined. Data collection, the unit root test, descriptive statistics, the SARIMA model, and all of the methodologies and tools that we utilized in our study are covered in this section.

**3.2 Materials and Methods for analyzing rainfall**

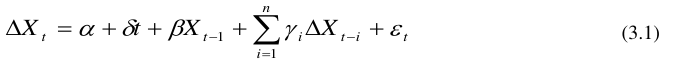
**3.2.1 Time series data and its sources**

The climatic data on daily total precipitation for the timeframe of 2008–2020 were acquired from the Bangladesh Meteorological Department (BMD) for all of the nation's meteorological stations. These data were used to create a forecast for the country.

**3.2.2 Unit roots and stationarity**

In order to use any type of regression analysis, the time series of the variables in question need to be stationary. This indicates that the mean and variance of each variable should not change in a predictable manner during the course of the investigation. When non-stationary data are used directly in the econometric estimation process, it is possible for erroneous findings to be produced (Gujrati 2004). Before beginning the regression analysis, it is essential to determine whether or not the time series of the variables are stationary in order to avoid any unexpected results. A time series variable is said to be non-stationary (or stationary) if its mean, variance, and autocovariance (at various lags) do not remain constant (or remain constant) during the course of the time series. It is said that a non-stationary series is integrated of order d, also written as I, if in order for it to become stationary, it has to have d differences performed on it (d). For the purpose of determining whether or not unit roots are stationary, the Augmented Dickey–Fuller (1979; ADF) test is utilized.

The equation for the ADF test looks like this for each series that is being investigated:



where t is the time or trend variable, *εt*is a pure white noise term and *ΔXt-1­ = (Xt-1 –* *Xt-2)*, *ΔXt-2­ = (Xt-2 – Xt-3)*, and so on. In the test for a unit root, the null hypothesis states that *β* = 0. If there is a statistically significant difference between the coefficient and 0, then the hypothesis that *Xt* has a unit root may be rejected.

**3.2.3 Descriptive statistics**

Statistics that are descriptive in nature, such as moving averages, standard deviations, and coefficients of variation, are used so that changes in rainfall over time may be understood.

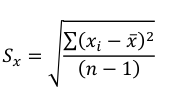
**3.2.3.1 Moving average**

The calculation of an average value requires the use of just a certain set of data; for instance, an average may be derived by utilizing only the most recent five values (Waller 2008).

**3.2.3.2 Standard deviation**

Utilizing the standard deviation approximation as a method of assessing dispersion is one of the most straightforward approaches of determining the variability of rainfall.

The following is an illustration of the sample standard deviation, often known as Sx (Waller 2008):



where 𝑆𝑥 = the estimator of the standard deviation 𝜎𝑥 of the amount of rainfall

𝑥 = sample mean

𝑛 = sample size

𝑥𝑖 = 𝑖th observation of X.

The scattering of any given dataset may be measured by examining how near the value of the standard deviation is to the number zero. This suggests that the data values are nearer to the average value of the sample, and that as a result, the data is more dependable for use in analytical endeavors.

**3.2.3.3 Mean**

The usage of the mean is an estimation of the point estimates that are used for the amount of typical rainfall that occurs in a given month. in order for it to be possible to utilize it to make predictions about future values.

x = xi/n

xi = ith observation of the average rainfall in a particular month

n = the number of year observation of a particular month

**3.2.4** **Seasonal Autoregressive Integrated Moving Average (SARIMA)**

One component separates an autoregressive integrated moving average (ARIMA) model from a seasonal autoregressive integrated moving average (SARIMA) model, which is built on the basis of seasonal patterns. There are regular seasonal impacts at play in many different types of time series data. Take for instance the typical amount of rainfall recorded in a region that experiences all twelve seasons. There will be a cyclical influence on an annual basis, and the amount of precipitation that falls during this specific timespan will almost certainly have a significant association with the amount of precipitation that was observed during the same period the previous year.

The general form of SARIMA(p, d, q)(P, D, Q)m is given by:

ΦP(Bm)φ(B)ΔDmΔdXt = ΘQ(Bm)θ(B)εt

In this case, {εt} denotes a nonstationary time series, and the Gaussian white noise process is referred to as {εt} . The cycle of the time series is denoted by the letter m. Polynomials φ(B) and θ(B) of orders p and q, respectively, are used to express the normal autoregressive and moving average elements, respectively. ΦP(Bm) and ΘQ(Bm) are the components of the seasonal autoregressive and moving averages, respectively; P and Q are the orders of these elements. The normal and seasonal difference components are represented by Δd and ΔDm respectively. The backshift operator is denoted by the letter B.

The time series of monthly rainfall is the primary focus of this investigation. In the event where the annual cycle of the series m equals 12.

**3.2.4.1 Model Identification**

During the phase of preliminary formulation, also known as model identification, the objective is to use computationally straightforward methods with the intention of reducing the number of plausible alternative explanations.

First, one should generate a time plot of the data and then investigate the chart to see whether or not there are any irregularities (Cryer and Chan, 2008). In the event that the variation continues to increase over time, then it will be required to stabilize the variance. The subsequent stage is to determine initial estimates of autoregressive order P, the order of differencing d, the moving average order q, and their respective cyclical parameters P, D, and Q. This will allow us to go on to the following phase. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the two components that have the utmost significance in this context (Stoffer and Dhumway, 2010). The degree of linear dependency that exists between samples in a time series that are divided by a lag q is what the ACF analyzes and quantifies. The PACF is used to assist in determining the required number of autoregressive terms, which is denoted by p. The ordering of the difference in frequency when comparing non-stationary time series to stationary time series is represented by the quantity d. In addition, a time series diagram and an ACF of the data will often be able to indicate whether or not any differencing is required. If there is a need for differencing, the time plot will demonstrate a consistent pattern of some form.

As soon as the initial numbers of D and d have been determined, the next step is to assess the ACF and PACF of ΔDmΔdXt in order to figure out the values of P, Q, P, and q. In addition, we have the option of selecting parameters with the help of Akaike's Information Criterion (AIC), which enables us to measure the values of parameters (Stoffer and Dhumway, 2010).

**3.2.4.2 Parameter Estimation**

After the model has been provisionally developed, statistical methods like Maximum Likelihood (ML) may be used to make estimates of the model's parameters as well as the standard errors associated with those parameters.

**3.2.4.2 Diagnostic Checking**

In most cases, this stage involves the examination of model evaluations in addition to the study of residuals. Standardized residuals should act as an independent and identically distributed (I.I.D.) sequence with a mean of 0 and a variance of 1 if the model fits the data properly (Cryer and Chan, 2008). According to Stoffer and Dhumway (2010), a normalized residuals plot or a Q-Q plot may be helpful in determining whether or not the data is normally distributed. Both the parametric test and the diagnostic check ought to go well for the model.

**3.2.5 Data analyzing methods**

The methods of data analysis make use of a variety of methodologies, some of which are comparable to one another. Because of this, we will examine each one individually.

|  |  |  |
| --- | --- | --- |
| **Statistical tools** | **Purpose of use** | **Software used** |
| Descriptive statistics such as mean, standard deviation, maximum, minimum | For figuring out the average, standard deviation, minimum, and maximum monthly and annual rainfall | R (Version 4.1.2) |
| Mean Procedure | To compare the average amount of rain that falls each month and each year in different districts. | R (Version 4.1.2) |
| Line chart | To keep track of rainfall patterns. | R (Version 4.1.2) |
| Augmented Dickey-Fuller test | To check the stationarity of time series data | R (Version 4.1.2) |
| Ljung-Box test | For checking the autocorrelation in rainfall data | R (Version 4.1.2) |
| SARIMA model | To establish a model for the monthly rainfall data | R (Version 4.1.2) |
| Autocorrelation function plot (ACF), Partial Autocorrelation function plot (PACF) | For figuring out the SARIMA model's seasonal and non-seasonal orders | R (Version 4.1.2) |
| Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) | For the purpose of selecting the ideal model from among a number of alternatives. | R (Version 4.1.2) |
| QQ plot | To check the validity of the model using residuals | R (Version 4.1.2) |
| Shapiro-Wilk test | To check the normality of the model's residuals | R (Version 4.1.2) |

Table 3.1 Statistical techniques for analyzing rainfall data

**3.3 Conclusion**

The primary purpose of this section was to provide an overview of the conceptual approach that would be used in this investigation. As a result, the process by which we simulate the precipitation data for each of the 5 districts (Dhaka, Rajshahi, Khulna, Chittagong, and Sylhet) in Bangladesh and the effect that this has on the country have been discussed. The mean, standard deviation, median, maximum, and lowest values for each year and month across all of the districts make up the essential approaches for the first stage. In the second step, we will make use of a line chart to depict the pattern of rainfall from year to year. At the third and final level, the SARIMA model and tests such as the Augmented Dickey-Fuller test, the Ljung-Box test, and the Shapiro-Wilk test were employed.