International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

# Linear Regressions of Predicting Rainfall over Kalay Region

**Ohnmar Myint**

Department of Engineering Mathematics, Technological University, Kalay, Myanmar

***How to cite this paper:*** Ohnmar Myint "Linear Regressions of Predicting Rainfall over Kalay Region" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1681-1685, **https://doi.org/10.31142/ijtsrd26711**

IJTSRD26711



**ABSTRACT**

Regression analysis is a statistical technique for investigating the relationship between variables. In this paper, rainfall and water level prediction models are discussed with the use of empirical statistical technique, Simple Linear Regression and analyzed the development of the predictive power of Linear Regression model to forecast the predicting rainfall and water level over Kalay in Sagaing Region for 10 years (2008-2017).The data of the monthly rainfall and water level used in this study were obtained from Meteorology and Hydrology Department of Kalay, Myanmar. In July 2015, Kalay was affected by the floods. So the rainfall and water level are predicted for next five years in this paper.

***KEYWORDS:*** *empirical statistical technique, linear regression, predicting, rainfall, water level*

## I. INTRODUCTION

The accurate and timely prediction of rainfall is a challenging task. Rainfall information is important for food production plan, water resource management and all activity plans in the nature. The occurrence of prolonged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduce crop yield. Myanmar is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural countries like Myanmar.

A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally, there are two approaches to predict rainfall. They are empirical and dynamical methods. The empirical approach is based on analysis of historical data of the weather and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely use empirical approaches used for climate prediction are regression, artificial neural network, stochastic, fuzzy logic and group method of data handling. In dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions. The dynamical approaches are implemented using numerical weather forecasting method [2].

In this paper, rainfall prediction model over Kalay region for 10 years is analyzed with the use of empirical statistical technique, Linear Regression. Kalay is situated at the western bank of MyitThar River in Sagaing Division, in the upper of Myanmar. Kalay township is located the north latitude of 23° 12′ and the east longitude of 94° 19′.The length of township is 12 miles from east to west and the breadth of it is 72 miles from North to South. So, we also focus on water level of MyitThar river for 10 years and analyze with the use of Linear Regression.

## II. SIMPLE LINEAR REGRESSION MODELS

Regression is a statistical empirical technique of data mining which has wide range of application in many fields like business, biological sciences and climate prediction. It is becoming a promising tool in the society. Regression in simple terms is defined as predicting one variable from

The simple linear regression model is as follows.

$Y = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n.$

In this equation, $Y$ is value of dependent variable or say predicted variable, $\beta_0$is constant value and $\beta_1$ is the predictor coefficient,the slope of regression line how much$Y$ changes for one unit change in $X$, $x_i$arepredictor variables or independent variables explaining the value of $Y$and $\varepsilon$ is an error term.As the number of dependent variable increases the accuracy in predicted value also increases.

The estimated regression line or least squares regression line is

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i, i = 1, 2, \dots, n$

where

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, is regression coefficient and

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$

$\bar{x} = \frac{1}{n}\sum x_i$ ,

$S_{xx} = \sum x_i^2 - n\bar{x}^2$

$S_{yy} = \sum y_i^2 - n\bar{y}^2$

$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y}$ ,is sample covariance.

The residual; $e_i = y_i - \hat{y}_i$ and the following relations can easily be derived.

SST (Total sum of square) = SSR (Regression sum of square) +SSE (Error sum of square) where

$$SST = \sum (y_i - \bar{y})^2 = S_{yy},$$

$$SSE = SST(1 - R^2).$$

The square of the sample correlation coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

Is called the coefficient of determination, $R^2$. It plays an important role in model checking. It tells us the goodness of fit of the model [4], [5].

## III. CASE STUDY

### A. Rainfall of Kalay Region

### A1: Mean Rainfall of Kalay Region for 10 years (2008-2017)

Table 1 shows the mean rainfall of Kalay Region for 10 years from 2008 to 2017.

Table1. Mean Rainfall & Estimated Mean Rainfall

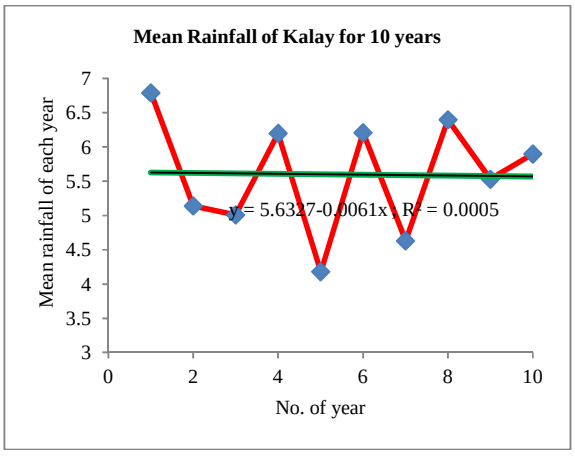
| No. of year X | No. of year | Mean rainfall Y | Estimated mean rainfall | Residual |
|---|---|---|---|---|
| 1 | 2008 | 6.79 | 5.63 | 1.16 |
| 2 | 2009 | 5.14 | 5.62 | -0.48 |
| 3 | 2010 | 5.01 | 5.61 | -0.6 |
| 4 | 2011 | 6.2 | 5.61 | 0.59 |
| 5 | 2012 | 4.18 | 5.6 | -1.42 |
| 6 | 2013 | 6.21 | 5.6 | 0.61 |
| 7 | 2014 | 4.63 | 5.59 | -0.96 |
| 8 | 2015 | 6.4 | 5.58 | 0.82 |
| 9 | 2016 | 5.53 | 5.58 | -0.05 |
| 10 | 2017 | 5.9 | 5.57 | 0.33 |

**Simple Linear Regression Model for Mean Rainfall**

Consider mean rainfall of Kalay Region for 10 years as shown in Table 1 and plot the scatter diagram shown in Figure 1a and 1b where number of years is represented along X-axis and mean rainfall of each year is represented along Y-axis.

According to above data, regression coefficient can be determined as follows. Here regression coefficient $\hat{\beta}_1$ is $-0.0061$ and $\hat{\beta}_0$ is 5.6327. Therefore estimated regression model is $\hat{Y} = 5.6327 - 0.0061x$ and which is expressed in Figure 1a. Here, $R^2 = 0.0005$.

The interpretation of $\hat{\beta}_1 = -0.0061$ is that the mean rainfall expected to decrease, since the slope is negative.



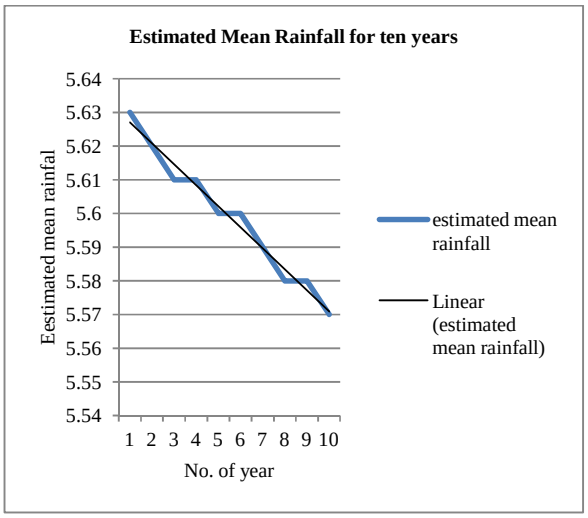




Figure1b. Regression Line for Estimated Mean Rainfall

### A2: Mean Rainfall in Rainy Season for 10years (2008-2017)

Table 2 shows the mean rainfall of rainy season in Kalay Region for 10 years from 2008 to 2017.

Table 2 Mean Rainfall of Rainy Season

| No of Year | X | Y |
|---|---|---|
| 16.64 | 1 | 2008 |
| 10.76 | 2 | 2009 |
| 10.28 | 3 | 2010 |
| 14.28 | 4 | 2011 |
| 10.56 | 5 | 2012 |
| 14.15 | 6 | 2013 |
| 11.15 | 7 | 2014 |
| 14.99 | 8 | 2015 |
| 12.32 | 9 | 2016 |
| 13.68 | 10 | 2017 |

**Simple Linear Regression Model for Mean Rainfall in Rainy Season**

According to above data, regression coefficient can be determined as follows. Here regression coefficient $\hat{\beta}_1$ is 0.0123 and $\hat{\beta}_0$ is 12.813. Therefore estimated regression model $is\ \hat{Y} = 12.813 + 0.0123x$ and which is expressed in Figure 2. Here $R^2 = 0.0003$. The mean rainfall of Kalay region in rainy season can be expected to increase since the slope $\hat{\beta}_1 = 0.0123$ is positive.

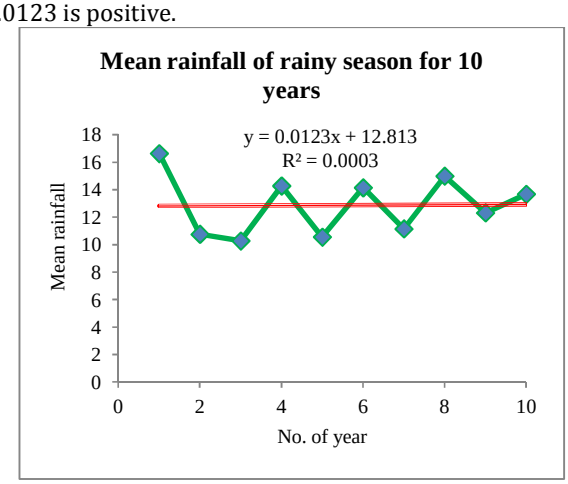



Figure2. Estimated Regression Line for Mean Rainfall in

**B. Water Level of MyitThar River**

**B1: Mean Highest Water Level of MyitThar River for 10 years (2008-2017)**

Table 3 shows the mean highest water level of Mithras River for 10 years from 2008 to 2017.

Table 3 Mean Highest Water Level

| No of Year | X | Y |
|---|---|---|
| 527.5 | 1 | 2008 |
| 499 | 2 | 2009 |
| 471.7 | 3 | 2010 |
| 552.8 | 4 | 2011 |
| 419.2 | 5 | 2012 |
| 554.7 | 6 | 2013 |
| 426.4 | 7 | 2014 |
| 706.9 | 8 | 2015 |
| 600.8 | 9 | 2016 |
| 578.6 | 10 | 2017 |

**Simple Linear Regression Model for Highest Water Level**

Consider mean highest water level of MyitThar River for 10 years and plot the scatter diagram shown in Figure 3 where number of years is represented along X-axis and mean highest water level of each year is represented alongY-axis. Here $r$egression coefficient $\hat{\beta}_1$is 12.759 and $\hat{\beta}_0$ is 463.57. Therefore estimated regression model is$\hat{Y} = 463.57 + 12.759x$and which is expressed in Figure 3. Here, $R^2 = 0.2003$.The mean highest water level of MyitThar river for 10 years can be expected to increase since the slope $\hat{\beta}_1 = 12.759$ is positive.

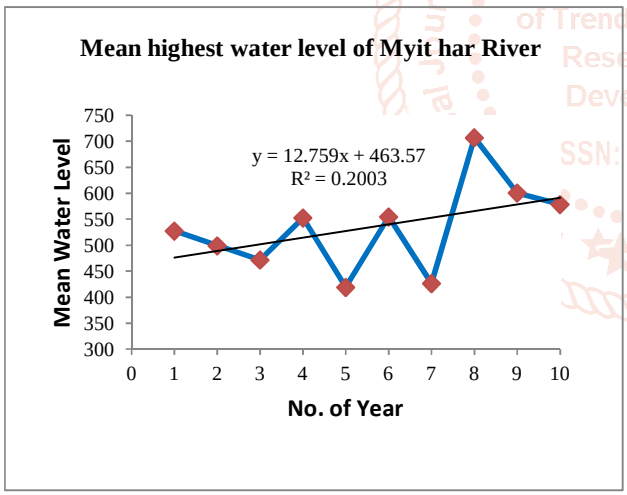



Figure3. Estimated Regression Line for Mean Highest Water Level of Mithras River

**B2: Mean Highest Water Level of Mithras River in Rainy Season**

Table 4 shows the mean highest water level of Mithras River in rainy season, from Jun to September, for 10 years from 2008 to 2017.

Table 4.Mean Highest Water Level in Rainy Season

| No of Year | X | Y |
|---|---|---|
| 845.5 | 1 | 2008 |
| 623 | 2 | 2009 |
| 632.5 | 3 | 2010 |
| 902.25 | 4 | 2011 |
| 633.25 | 5 | 2012 |
| 868.75 | 6 | 2013 |
| 650.25 | 7 | 2014 |
| 1243.75 | 8 | 2015 |
| 783.25 | 9 | 2016 |
| 878 | 10 | 2017 |

**Simple Linear Regression Model for Mean Highest Water Level in Rainy Season**

According to above data, regression coefficient can be determined as follows. Here regression coefficient $\hat{\beta}_1$ is 24.194 and $\hat{\beta}_0$ is 673.58. Therefore estimated regression model is $\hat{Y} = 673.58 + 24.194x$and which is expressed in Figure 4.Here, $R^2 = 0.1465$.The mean highest water level of MyitThar river in rainy season can be expected to increase since the slope $\hat{\beta}_1 = 24.194$ is positive.

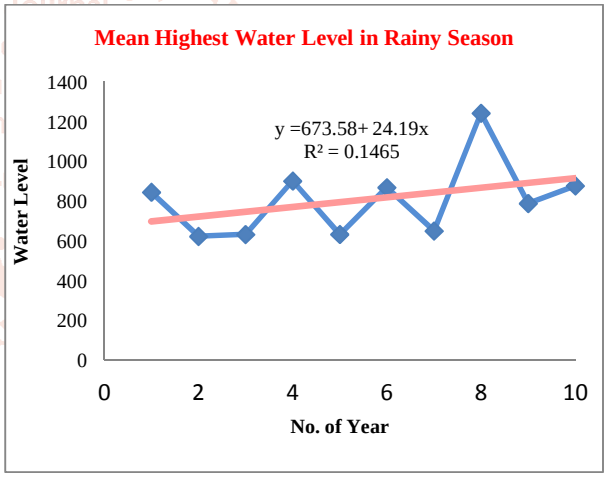



Figure 4 Estimated Regression Line for Mean Highest Water Level in Rainy Season

**C: Rainfall of Kalay Region and Highest Water Level of Mithras River**

Table 5 shows the rainfall of Kalay region and highest water level of MyitThar river in July for 10 years from 2008 to 2017.

Table 5.Rainfall of Kalay Region and Highest Water Level of Mithras River in July

| | No. of year | Rain-fall | Highest water level | | No. of year | Rain-fall | High-est water level |
|---|---|---|---|---|---|---|---|
| 1 | 2008 | 27.91 | 1024 | 6 | 2013 | 15.91 | 663 |
| 2 | 2009 | 8.39 | 442 | 7 | 2014 | 13.34 | 547 |
| 3 | 2010 | 9.18 | 609 | 8 | 2015 | 31.61 | 1942 |
| 4 | 2011 | 15.52 | 912 | 9 | 2016 | 16.64 | 691 |
| 5 | 2012 | 11.57 | 416 | 10 | 2017 | 17.08 | 866 |

The rainfall of Kalay region and highest water level of Mithras River in July for 10 years are analyzed as shown in Figures 5a and 5b respectively. We can obviously see that among them, in year 2015, both the rainfall and water level are maximum and

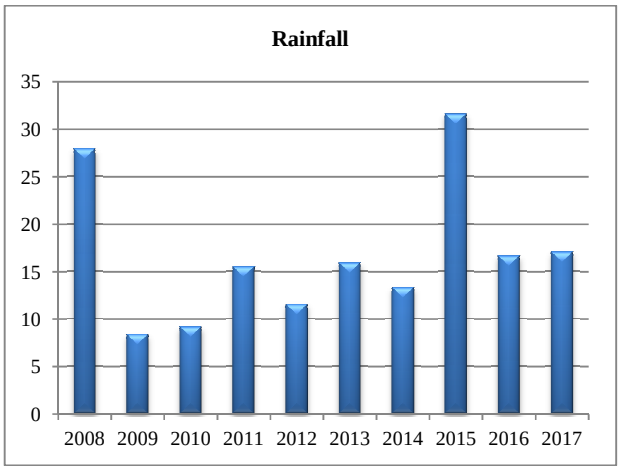



Figure 5a.Rainfall of Kalay Region in July

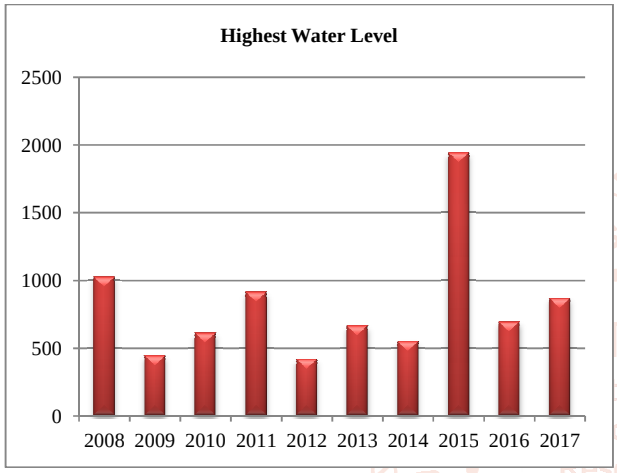



Figure 5b Highest Water Level of Mithras river in July

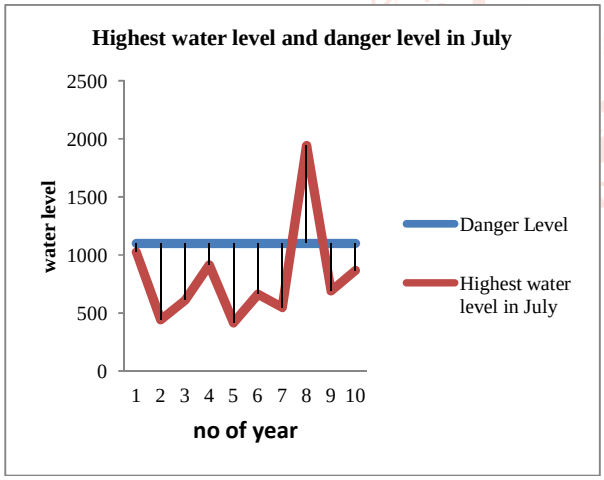



Figure 5c Highest Water Level and Danger Level of Mithras River in July

Table 6.Rainfall and Highest Water Level in 2008 and 2015

| | | 2008 | 2015 |
|---|---|---|---|
| Rainfall | in July | 27.91 | 31.61 |
| | in rainy season(mean) | 16.64 | 14.99 |
| | for a year(mean) | 6.79 | 6.4 |
| Highest Water Level | in July | 1024 | 1942 |
| | in rainy season(mean) | 845.5 | 1243.75 |
| | for a year(mean) | 527.5 | 706.92 |

According to the previous data analysis and trends of regression lines, we can forecast for the estimated mean rainfall of Kalay region and mean highest water level of MyitThar river for next 5 years from 2018 to 2022 as shown in Table 7.

Table 7 Estimated Rainfall and Highest Water Level in next 5 years (from 2018 to 2022)

| | | Rainfall | | Highest water level | |
|---|---|---|---|---|---|
| | | in rainy season (Mean) | For a year (Mean) | in rainy season (Mean) | for a year (Mean) |
| 1 | 2018 | 12.9483 | 5.5656 | 939.714 | 603.919 |
| 2 | 2019 | 12.9606 | 5.5595 | 963.908 | 616.678 |
| 3 | 2020 | 12.9729 | 5.5534 | 988.102 | 629.437 |
| 4 | 2021 | 12.9852 | 5.5473 | 1012.3 | 642.196 |
| 5 | 2022 | 12.9975 | 5.5412 | 1036.49 | 654.955 |

## IV. RESULTS AND DISCUSSIONS

In years 2008 and 2015, the mean rainfall of Kalay Region is maximum based on previous 10 years data and it will be expected to decrease since the slope of the regression line is negative. Furthermore, in those years, the mean rainfall of Kalay Region in rainy season is also maximum but it will be expected to increase since the slope of the regression line is positive.

Similarly, in year 2015, the mean highest water level of MyitThar river is maximum based on previous 10 years data and it will be expected to increase since the slope of the regression lines are positive, so does the mean highest water level of MyitThar river in rainy season.

In July 2015, a monsoon rain triggered a natural disaster, and a state of emergency was declared in four regions of the country. The disaster caused a flash flood in Kalay and surrounding areas. Kalay was devastated by the disaster. The reason is because of the maximum rainfall and water level of MyitThar river. In July 2015, both the rainfall and water level are maximum and these are above the danger level 1100cm. According to data analysis for next 5 years, the rainfall will be maximum and water level can be reached to the danger level. So preparations will be needed to prevent from flash flood in Kalay and surrounding areas.

## V. CONCLUSION

The values of rainfall are calculated using the data collected over ten years. Mean rainfall of Kalay region and highest water level of MyitThar river for 10 years and rainy season as well are used as a dependent variables in simple linear regression, which can improve the efficiency of prediction of rainfall and water level.

## VI. ACKNOWLEDGEMENTS

The author would also like to thank the Department of Hydrology and Meteorology, Kalay for providing important and valuable data for the research. The author specially would like to Dr Khaing Khaing Aye, Professor and Head of Yangon Technological University for her giving helpful advice and commands throughout the research.

**REFERENCES**

[1] MAINavid and NH Niloy, "Multiple Linear Regressions for Predicting Rainfall for Bangladesh." *Communications, Science PG,* vol. 6, No. 1, pp. 1-4,1990

[2] Wint Thida Zaw and Thinn Thu Naing, "Empirical Statistical Modeling of Rainfall Prediction over Myanmar." *International Journal of Computer and Information Engineering*, vol. 2, No. 10, pp. 3418-3421, 2008

[3] Paras and Sanjay Mathur, "A Simple Weather Forecasting Model using Mathematical Regression." Indian Research Journal of Extension Education, Special Issue, vol. 1, pp. 161-168, 2012

[4] Douglas C. Montgomery, Elizabeth A. Peck and G. Geoffrey Vining, "Introduction to Linear Regression Analysis." Fifth Edition, Wiley Series in Probability and Statistics,2012

[5] John O. Rawlings, Sastry G. Pantula and David A. Dickey, "Applied Regression Analysis." Second Edition, Springer, 1998