CHAPTER - VIII

RELATIONSHIP BETWEEN RAINFALL AND RICE YIELD

CHAPTER - VIII

RELATIONSHIP BETWEEN RAINFALL AND RICE YIELD.

8.1. INTRODUCTION

Agriculture is considered as the primary sector for economic development of our country .In tern agricultural productivity can be defined in different senses of which agriculture yield per hectare is one .

The influences of weather element,particularly rainfall on the agricultural operations and crop yield,can affect favourable and unfabourble conditions during the different stages of development. So,rainfall is the factor determining the success or failure of agricultural enterprises. This is because cultivators have no control over this natural random phenomenon.

In this area farming is done under unirrigated conditions as such it depends mostly on weather and climate. The impact of weather and climate on agricultural productivity is of vital importance. Rainfall is the most important weather parameter that affects paddy crop in this region where irrigation plays a minor role. Paddy is the principal crop of the kharif season. It is mostly rain fed crop. The total absolute rainfall of a particular season and the total production of crop are considered for most of the economic analysis on food grain production. But the nature of intensity of daily rainfall has much influence on variations of yield of this particular crop. These effects are manifested through plant characteristics like height, number of tillers,leaf area and number of ear heads which ultimately affect the yield of the crop.For the purpose of our thesis ,we concentrate on the effect of this climatic factor to the yield of rice.And it has been considered in this study as economic factor in agriculture .

This chapter attempts to establish the relationship between rice yield and the nature of daily rainfall in this district.

8.2. VARIABLES

A time series of crop yield may be divided into three major sources of variation in the yield of rice over any region. These are identified as (i) technological change,(ii) meteorological variability and (iii) random "noise". The technological change in yields creeps in due to the recent advancement in agricultural technology. A suitable linear time-scale variable is introduced in crop-yield model to account for the technological trend. Among meteorological parameters rainfall is one of the most important elements which affect crop yield significantly. The third component has the impact and contribution of random noise which is very little.

Of all the factors affecting agricultural production, availability of water is undoubtedly the most important one. For rainfed crops,rainfall constitutes water gain while the nature of rainy days are of much importance during different stages of development of crop. The nature of daily rainfall data,derived from it,are used in this study. The classification is made by comparing the amount of rainfall received during the past 24 hours. The daily derived rainfall parameters are classified into the following categories :

i) Rainy Days (RD) Days when the rainfall is equal to or more then 2.5 mm.

ii) Restricted Rainy Days (RRD). Number of rainy days except when preceded by a day having 50 mm or more rainfall.

iii) Crop Rainy Days (CRD) Number of days when the rainfall is 5 mm or more. But a day having rainfall 3.8 mm is considered as a crop rainy day with a gap of two non rainy days.

The daily derived parameters as described above are based on rainfall,during the period 1972 to 1988 at Cooch Behar. The monthly derived values from June to the middle of October of each year have been computed. The nature of the day's rainfall for the months of sensitive period of crop life time are considered as the explanatory variables for this study. The average yield of winter rice in this district is used as the dependent variable in this study. A data series of average yield of winter rice for the same period i.e.1972 to 1988 has been used in the present study to develop the multiple regression model.

The yield rate of winter rice has been considered as the dependent variable throughout this chapter. The other variables are potential explanatory variables.

Now we first consider the model with fourteen explanatory variables namely :

I) Rainy Days (RD) :

X_1 = Rainy days during July.

X_4 = Rainy days during August.

X_7 = Rainy days during September.

X_{10} = Rainy days during October.

X_{11} = Rainy days during June

II) Restricted rainy days (RRD)-

X_2 = Restricted rainy days during July

X_5 = Restricted rainy days during August

X_8 = Restricted rainy days during September

X_{12} = Restricted rainy days during June

III) Crop rainy days (CRD) -

X_3 = Crop rainy days during July

X_6 = Crop rainy days during August

X_9 = Crop rainy days during September

X_{13} = Crop rainy days during June

X_{14} = Technological trend in time-scale variable takes from 1 to 17.

8.3. DATA MATRIX OF EXPLANATORY VARIABLES

The individuals of 17 years and 14 variables are mainly related to the derived variables from the daily rainfall data. These variables are separated by rainy months included in the life cycle of the crop. We denote the number of individuals by n, the number of variables by k. So,there are nk variable values.The individual variables are denoted by the subscripts : $X_1, \ldots\ldots X_k$. Mathematically,a set of values are arranged in rows and columns in the matrix form (data matrix). The single symbol X stands for the numerical information in this data matrix as a whole . The data,in the form of a matrix are given in the Appendix.

A preliminary summarisation of the data is provided by calculating the mean and standard deviation of each variable. These basic statistics of the variables are given in Table 8.1. This table also shows the standard error of mean of each variable.

It is also seen from the table that except the dependent variable, the means of the independent variables are in the range from 3.765 to 21.412 days while their standard deviations arefrom 2.322 to 3.704. But those of the dependent variables are 1194.882 and 140.265 respectively. It is also observed that the standard errors of the means of the variables, X_{11} and X_{14} are greater than one while those of the other variables are less than unity. The standard error of the means of the dependent variable is 34.019. So, there are high variability on the observations of the yield of winter rice in this district.

Besides these basic statistics,the other measurements of association between each pair of explanatory variables are correlation coefficients. The Table 8.2 reports the correlation matrix between the explanatory variables as well as the dependent variable with each explanatory variable. The correlation matrix is a symmetric matrix.

The most strongly and positively correlated pair of variables are X_{11} and X_{19} with a correlation coefficient 0.90 and the weakest are X_5 and X_9 with a correlation coefficient .005, virtually zero.

Table **8.2** also shows the correlation between the dependent variable and each of the explanatory variables in which the dependent variable is positively correlated with X_3, X_4, X_5, X_9, X_{10}, X_{13} and X_{14} whereas it has negative correlation with the variables X_1, X_2, X_6, X_7, X_8, X_{11} and X_{12}.

8.4. MULTIPLE REGRESSION MODEL.

An agroclimatic study of the relationship between the crop yields and weather parameters is carried out with the help of empirical statistical multiple regression model. This model is also generally employed for making quantitative crop yield forecast on operational basis.

The technique has been utilized in the present study to develop linear multiple regression model to establish the crop weather relationship. The relationship between yield per hectare (in Kg) of paddy crop and the nature of daily rainfall is to be established employing the following functional form :

$$y = b_o + \sum_{i=1}^{k} b_i x_i + u_i$$

Where y is the dependent variable i.e. yield of winter rice in Cooch Behar.

X_i's are the derived parameters from the nature of daily rainfall for the sensitive period by month. These are also known as nonstochastic components of the model.

u_i is the stochastic component. This is also known as disturbance term.

8.4.1. ASSUMPTIONS

The linear multiple regression model is based on some assumptions . Some of them are related to the distribution of stochastic variables, some to the relationship between nonstochastic and stochastic variables and finally some refer to the relationship between these nonstochastic variables.

a) Assumptions about the stochastic variables.

Assumption -I.Normality-the stochastic variables are normally distributed with zero mean and constant variance. Regarding constant variance,for all values of X, u' s would show the same dispersion round their mean.

i.e.

$$E\ (u_i) = 0$$

and $E\ (u_i^2) = \sigma_u^2$ (Constant)

Assumption 2. Nonautocorrelation or serial independence of the u's. This assumption states that any two values of disturbance term (stochastic) u_i and u_j are not linearly correlated i.e.

$$E\ (u_i u_j) = 0 \quad i \neq j.$$

(b) Assumption 3.Assumption for stochastic and nonstochastic variables.

Independence of u_i and x_i. This assumption states that every stochastic term is independent of the explanatory variables i.e.

$$E\ (u_i\ x_{ij}) = 0$$

Where X_{ij} is the element of the data matrix,i.e.,in the ith variable and the jth individual year.

This assumption is not very critical. It is automatically fulfilled if the explanatory variable is nonstochastic.

Assumption 4. Assumption about the relationship between the explanatory variables themselves.

This assumption states the closeness of explanatory variables to each other. This is the case of multicollinearity.

8.4.2. ESTIMATION OF COEFFICIENT OF REGRESSION MODEL

We try to use the rules of thumb by which we can derive

(a) The regression co-efficients ,

(b) The variance of the co-efficients

and

(c) The coefficient of multiple determination,

The population regression space involving dependent variable Y and k explanatory variables :

X_1, X_2,................,X_k may be described as -

$$y_i = B_o + B_1 X_{1i} + B_2 X_{2i} \ldots\ldots\ldots\ldots B_k X_{ki} + u_i$$

Where B_o is the intercept , B_1 to B_k are the regression (slope) coefficients,u_i is the disturbance term the suffix i indicates the ith observation of the population which comprises n observations. We should therefore,have n number of linear equations. These equations can be put in a compact form using matrix notation :

$$\underset{\sim}{y} = \underset{\sim}{x} \; \underset{\sim}{B} + \underset{\sim}{u} \; .$$

where

$\underset{\sim}{y}$= (nx1) column vector of observations on the dependent variables y_i

$\underset{\sim}{X}$= nx(k+1) matrix giving n observations on k explanatory variables. The first column of the matrix is unity representing the intercept term.

$\underset{\sim}{B}$= $\{(K+1)x1\}$column vector of the regression coefficient.

$\underset{\sim}{u}$= (nX1) column vector of n observations on disturbance term i.e.,

$$\underset{\sim}{Y} = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix}$$

$$\underset{\sim}{X} = \begin{vmatrix} 1 & X_{11} & X_{21} & X_{k1} \\ 1 & X_{12} & X_{22} & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & X_{kn} \end{vmatrix}$$

$$\underset{\sim}{B} = \begin{vmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{vmatrix} \qquad \underset{\sim}{u} = \begin{vmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{vmatrix}$$

The ordinary least square method has been applied to estimate the regression coefficients. This is given by

$$\underset{\sim}{b} = (X'X)^{-1} \underset{\sim}{X}'\underset{\sim}{Y}$$

Here the estimators follow the Gauss-Markov property.

In this model the ordinary least squares estimators are those which minimize Σe_i^2 with respect to b_i where e_i is the n element column vector of n residuals in the sample regression space and it can be measured by

$$e_i^2 = (Y_i - b_o - b_1 x_{1i} - \ldots\ldots\ldots\ldots b_k X_{ki})^2$$

The total variation in the dependent variable (Y) can be partitioned into variation due to chosen regressors and the residual variance. The partitioning is obtained as follows :

$$\sum_{i}^{n} (Y_i - \overline{Y})^2 = \Sigma (\hat{Y}_i - \overline{Y})^2 + \Sigma e^2$$

$$\text{i.e.,} \Sigma y_i^2 = \Sigma \hat{y}_i^2 + \Sigma e_i^2$$

Where,

Σy_i^2 = Total sum of square (TSS)

$\Sigma \hat{y}_i^2$ = Regression sum of square (RSS)

and

Σe_i^2 = Error sum of square

The coefficient of multiple determination or the squared multiple correlation coefficient is the ratio of regression sum of square to total sum of square and is denoted by R^2 .

The estimate of R^2 can be obtained as

$$R^2 = \frac{\text{Regression SS}}{\text{TSS}}$$

$$= \frac{\Sigma \hat{y}_i^2}{\Sigma y_i^2}$$

When the explanatory variables in the function are large with comparison to the number of observations,the multiple determination can never reduce but it will rise usually because the numerator will increase in the expression of R^2 while the denominator remains the same. To correct this defect we adjust R^2 by considering the degree of freedom. So,the adjusted coefficient of multiple determination is

$$\bar{R}^2 = 1 - (1 - R^2)\frac{(n - 1)}{(n - k)}$$

$$= 1 - \frac{\Sigma e^2/(n - k)}{\Sigma y^2/(n - 1)}$$

Where R^2 is the unadjusted multiple correlation coefficient, n is the number of sample observations and k is the number of parameters estimated from the samples.

The adjusted value of multiple correlation coefficient $\bar{R}^2$ when multiplied by 100 gives the percentage of total variation in Y explained by the regression.

The method outlined in this sub-section is concerned to develop the suitable multiple regression model. Multiple regression analysis,in practice,require the use of appropriate computer programme. The methods employed in our analysis are supported by the software package of Data matric. The computer out puts of multiple regression analysis are presented there step by step.

The multiple regression equation can be written as

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \;\ldots\ldots\ldots + b_{14} \; x_{14} + u$$

The extract form of the computer output for the fourteen explanatory variables model is shown in Table **8.3.** For this model, the explanatory variables, X_1, X_2, X_6, X_7, X_8, X_{11} and X_{12} have marginal negative effects and X_3, X_4, X_5, X_9, X_{10}, X_{13} and X_{14} have the marginal positive effect on the dependent variable. The table also shows the value of t-statistics i.e., $b_i / SE(b_i)$ of each of the explanatory variables as well as the constant i.e. the intercept.

In this equation the degree of freedom of t-statistic is 2, i.e, [17 - (14 + 1)]. It is observed that the regression coefficients b_4, b_{11} and b_{14} have the t-value less than 2 i.e. the hypothetical value of standard-error test which takes the constant value 2 to compare the calculated value. But we cannot consider this procedure to detect the unwanted variables in the equation.

However, to get the accuracy of the predictive model we have to omit the variable which may contribute little to predict the dependent one.

Now, the step-down method has been applied to solve this problem. One of the rules of this method is to omit the unwanted variable (x_u) if

$$T = b_u^2 \;/\; s^2 < 1$$

Where b_u is the regression coefficient of the unwanted explanatory variables and s^2 is square of standard error of that coefficient.

Now,we consider the multiple regression equation of fourteen explanatory variables. Here the variable x_{14} is detected as unwanted variable where the value of T for the variable x_{14} is .02224 (<1). The variable x_{14} i.e. time scale variable is omitted from our multiple regression equation.

As the technological trend i.e. time-scale variable is omitted from the multiple regression equation, it is indicated that the advances in agricultural technology have very little contribution on the yield of winter rice in the district of Cooch Behar.

Moreover,the technological trend caused by the use of high yielding varieties,large scale use of fertilizers,better irrigation facilities,use of insecticides and improved management practices has no significant contribution on the production of winter rice at the district of Cooch Behar.

The calculated value of R^2,adjusted $\bar{R}^2$ and F-value of the multiple regression equation and the regression coefficients along with respective values of standard error and t-values are given in Table **8.3**.

Now,the regression of y on the remaining (14-1) i.e.,thirteen explanatory variables are to be computed. The computed new multiple regression equation is also given in Table **8.4** ,and the same rule is followed to identify the unwanted variables.

Here the value of T for the value X_{11} is observed less than one i.e.,

$$T\ (x_{11}) = 0.0627 \quad (<1)$$

The values of R^2, adjusted $\bar{R}^2$ and F- value and model of multiple regression equation with the values of regression coefficient corresponding to their standard errors and the t-values are given in Table **8.4**.

So, this variable X_{11} is omitted from the equation.

This is indicated that the rainy days during the month of June have no significant effect on the production of winter rice in Cooch Behar district.

Here we also note, that the value of R^2 is decreased by a small quantity whereas the value of adjusted $\bar{R}^2$ and the F value have been increased due to the omission of the variable X_{14} that is the variable of technological change.

By following the same method we now compute the multiple regression equation by eliminating the variable X_{11} i.e. rainy days during the month of June.

We next compute the regression :

$$y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{12}, X_{13})$$

Now, omitting the variables X_{11} and X_{14}, the regression equation of y on the remaining twelve explanatory variables is computed. The latest model of multiple regression equation is given in Table **8.5** which finally consists of one dependent variable and twelve explanatory variables. The variable X_{11}, rainy days during the month of June and the variable X_{14}, the technological trend are excluded from the multiple regression equation by adopting the step-down method. But in the new equation, there is no variable which we can omit by applying the step-down method. So, this equation now has been employed to develop the relation between the winter rice yield and the derived variables of rainfall of different sensitive months.

So, finally, we fit multiple linear regression model on yield of winter rice.

However the calculated values of R^2, adjusted $\overline{R}^2$ and F-statistic of the equation and the required regression coefficients along with their standard errors and the corresponding t-values are given in Table 8.5.

The estimates of the multiple correlation coefficient and regression coefficient which are defined through least squares estimators, are essential to examine the reliability precision of the estimates.

8.4.3. TEST FOR THE SIGNIFICANCE OF THE MULTIPLE CORRELATION COEFFICIENT

The test performed in regression analysis is a test concerning the overall explanatory power of the regression as measured by R^2.

The null hypothesis of this test is that the multiple correlation coefficient in population is zero and the alternative hypothesis is that it is greater than zero.

The F ratio is a test of significance of R^2, the statistic :

$$F^* = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k}$$

Confirms to F-distribution with (k,n-k-1) degrees of freedom.

So,we can apply the test of this hypothesis under consideration,knowing only the value of observed multiple correlation.

The values of R^2,adjusted $\overline{R}^2$ and F^* ratio are given in Table 8.5.

The calculated R^2,the squared coefficient of multiple correlation is 0.9832 and the adjusted $\overline{R}^2$ is 0.9327. The calculated value of F^* ratio is 19.4674. The tabulated value of F-ratio with the degrees of freedom 12 and 4 is 14.37 at 1 percent level of significance. Hence,it is statistically highly significant. Therefore ,we consider that multiple regression equation is well fitted to the derived parameters of rainfall in Cooch Behar district.

Now consider the value of adjusted $\overline{R}^2$ which is 0.9327. We multiply the value of $\overline{R}^2$ by hundred,we get it 93 percent.

The model,therefore, can explain 93 percent of total variation of rice yield with multiple correlation coefficient as 0.98. Hence,all the explanatory variables may be emerged significantly at 1 percent level.

8.4.4. THE MEAN AND VARIANCE OF THE PARAMETER ESTIMATES

Since the least square estimators are linear combinations of independent variables. The intercept and slopes are normally distributed with

$$\text{Mean} : E(b_o) = B_o$$

$$\text{and} : E(b_i) = B_i$$

And the variance of any estimator may be derived from the principal diagonal of dispersion matrix $(X'X)^{-1}$ and then multiplied by σ^2.

An unbiased estimator of σ^2 is denoted by $\hat{\sigma}^2$ and is given by

$$\hat{\sigma}^2 = \frac{\Sigma e_i^2}{(n-k-1)}$$

Where k is the total number of regressors,or independent explanatory variables of the model.

so,standard error of $b_i = \sqrt{\hat{\sigma}^2 C_{ii}}$

where C_{ii} is the diagonal element of the dispersion matrix $(X'X)^{-1}$.The procedure described above has been suggested by Rao (1984).

The value of b_i,s and their standard error are given in Table **8.5** which also shows our best fitted multiple regression model.

8.4.5. TEST OF SIGNIFICANCE OF PARAMETER ESTIMATES

When the sample is small $(n < 30)$ and the parent population is normal,we may apply a test statistic which is based on Student's t-distribution. To perform a two tailed test of this test criterion we must adopt the null and alternative hypothesis with desired level of significance and suitable degree of freedom. We assume that the least squares estimates are normally distributed.

The null hypothesis is that there is no linear relationship between the explanatory variables and the dependent variable y i.e.

$$H_o : B_i = 0$$

and is tested against the alternative hypothesis,

$$H_1 : B_i \neq 0$$

In this case the t-statistic reduces to

$$t^* = \frac{b_i}{\hat{\sigma}(b_i)}$$

The sample value of t^* is estimated by dividing the estimate b_i by its standard error.

Where

b_i = least squares estimate of B_i

$\sigma^2(b_i)$ = estimated variance of B_i

n = Sample size

k = Total no. of estimated parameters.

The value of t^* is compared to the theoretical (tabular) values of t which defines the critical region in two-tailed test with {n − (k + 1)} degrees of freedom.

If the calculated values of t^* are greater than the tabular value of t i.e. $t^* > t$,we may reject the null hypothesis. This means that we may accept the estimated b_i which is statistically significant at chosen level of significance.

Therefore,it is essential to examine the reliability or the precision of the least square estimators.

Table 8.5 reports the value of the estimates of the parameters and their respective standard errors and the computed t^*-value of each of the variables.

The observed value of the t^*-ratio would be compared to the theoretical value of t, obtainable from the t-table with degree of freedom {n − (k + 1)} = 17 − (12 + 1) = 4 which is the suitable degree of freedom for the fitted model.

The statistical reliability of the estimated parameters of the equation as shown in Table 8.5 is discussed here one by one parameter.

The calculated t-value of the regression coefficient of b_1 is 5.995 which is greater than the tabulated value of t with 4 degree of freedom at one percent level of significance (4.604). Since we reject the null hypothesis with 99 percent confidence the variable X_1, the rainy days during July, has a significant effect on the yield of winter rice in the district of Cooch Behar. But the regression coefficient bears negative sign and the negative coefficient is also statistically significant. Hence the yield of winter rice is found to be inversely related to the rainy days in the month of July.

The same result has also been observed on the coefficient of variable X_2, the restricted rainy days in the month of July but the coefficient is significant at 5 percent level (t = 2.776). This suggests that this variable has a significant and negative influence of the yield of winter rice at this place.

The coefficient of the variable X_3 i.e., the crop rainy days in the months of July has positive sign and it is statistically highly significant at 5 percent level. The result shows that the variable has a positive influence on the rice yield.

The coefficient of the variable X_4, rainy days in the month of August, has a positive sign but it is not statistically significant, showing that this variable is not a determinant of the yield of winter rice at Cooch Behar.

The estimated coefficient of the variable X_5, the restricted rainy days in the month of August, has a positive sign and it is found to be highly significant, in influencing the yield of rice. It is significant at one percent level of significance.

The coefficient of the variable X_6, the crop rainy days in the month of August, has a negative sign and it is also statistically significant at 5 percent level. Hence, it is inversely related to the yield of winter rice.

The estimated coefficients of the variables x_7 and x_8, the rainy days and the restricted rainy days in the month of September respectively, have negative sign and these are found to be highly significant at one percent level of significance. So, these variables are also inversely related to the yield of winter rice.

The estimated coefficients of the variables X_9 and X_{10}, the crop rainy days in the month of September and the rainy days in the month of October, have positive sign and are statistically highly significant. These are significant at one percent level of significance. The results show that the variables have positive influence to the yield of winter rice at this place.

The estimated coefficient of the variable X_{12}, restricted rainy days in the month of June, has a statistically highly significant at 0.1. percent level (t=8.61) it has a negative influence on the rice yield.

The coefficient of the variable X_{18} , the crop rainy days in the month of June,has a positive sign and it is found to be highly significant at 0.1 percent level. This suggests that the variable has a significant influence on the yield of winter rice in Cooch Behar.

8.5. SECOND ORDER TEST OF FITTED EQUATION

If the assumptions of an econometric method are violated in any application,the estimates obtained from the method do not possess some or all of the optimal properties. Now we will develop the economic criteria or second order test for judging the goodness of the estimates. These criteria provide some evidence about the validity of the assumptions of the classical linear regression model. The major problems likely to be encountered at this stage of analysis are autocorrelation and multicollinearity. While multicollnearity is the problem concerning closeness of explanatory variables to each other, the other pertains to the behaviour of estimated residuals.

8.5.1. AUTOCORRELATION.

One of our assumptions of ordinary least squares specified earlier is that the successive values of disturbance term (u) are independent which is stated as

$$E(u_i u_j) = 0 \qquad i \neq j, i \,\&\, j = 1,2,\ldots\ldots,n$$

When this assumption is not satisfied,the problem of autocorrelation arises. For testing the presence of autocorrelated errors,we use the residual terms by estimating the ordinary least squares.

Darbin-Watson have suggested a test for autocorrelated errors which is applicable to small samples also.

The test may be outlined as follows :

The null hypothesis is

$Ho : \rho = 0$

or , Ho: the U's are not autocorrelated. This hypothesis is tested against the alternative hypothesis

$$H_1 : \rho \neq 0$$

i.e. H_1 : the U's are autocorrelated

To test the null hypothesis,we use the Durbin-Watson statistic :

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Where e_t are sample residuals.

It can be shown that the value of d lies between 0 and 4 and that when d = 2,then ρ = 0,

Thus,testing Ho : ρ = 0 is equivalent to testing Ho : d = 2. Expanding the 'd' statistic we,therefore,obtain

$$d \approx 2(1 - \rho)$$

From this expression,it is clear that the values of d lies between 0 and 4.

If there is no autocorrelation i.e,

$$\rho = 0 \text{ i.e. } d = 2$$

Thus if from the sample data we get d = 2 and we can not reject the null hypothesis that there is no autocorrelation in the function.

So,it should be clear that in the Durbin-Watson test,the null hypothesis of zero autocorrelation is carried out indirectly when the explanatory variables are large,by testing the equivalent null hypothesis, Ho : d = 2

The computed D -W value (d) is 2.042 which is likely equivalent to 2. Thus,from the sample data it is found that

$$d = 2$$

Hence,we accept the null hypothesis that there is no autocorrelation in the multiplied regression equation. The calculated values of D-W statistic is given in Table -**8.5.**

Therefore, the autocorrelation is not found to be significant and hence,it may be concluded that statistically,the yields in winter rice in Cooch Behar district do not persist from year to year during the time frame of this study.

8.5.2. MULTICOLLINEARITY

The multicollinearity in econometrics is defined as a situation of high correlation between the explanatory variables. Thus,the problem of multicollinearity is always present except for an extreme case where the explanatory variables are called orthogonal. By very definition,two variables are called orthogonal if the correlation between them is zero. But such a case is never encountered where the real world data are handled.If multicollinearity is always present,we have to ascertain whether it is serious i.e.,whether it is of such magnitude as to bother the analyst.

To ascertain the presence and extent of multicollinearity in the data simple zero order correlation matrix of all the explanatory variables are worked out. If none of the correlation coefficient is high,the problem of multicollinearity is not considered serious. In general,if the correlation coefficient is 0.8 or more,it is to be considered high but if it is less than 0.70,it can be considered as low.

In this connection Klein has suggested that multicollinearity should not be considered serious if the simple correlation between a pair of variables is less than the multiple correlation coefficient. That is Klein argues that collinearity is harmful if

$$r^2_{x_i x_j} \geq R^2_{y.x_1 \ldots\ldots\ldots\ldots x_k},$$

where $r^2_{x_i x_j}$ is the simple correlation coefficient between any two explanatory variables (X_i, X_j) and R^2 is the multiple correlation coefficient of the function.

In using Klein's method of detecting multicollinearity,we should consider the value of the estimated multiple correlation coefficient. The simple correlation matrix of explanatory variables are given in Table **8.2.** The multiple correlation coefficient (R^2) has been estimated as 0.98. Then,according to Klein,the simple correlation coefficient *between explanatory variables cannot be considered high because* these are all less than 0.98.

Hence,multicollinearity is not a problem in the fitted model of multiple regression.

8.6 CONCLUSION :

The linear form of the multiple regression equation is found to be empirically appropriate as it possesses high explanatory power. The results show that the derived parameters of rainfall in the months,June to October (middle) have been found to be highly significant in influencing the yield of winter rice in Cooch Behar. The model explains more than 93 percent of total *variation in yield with multiple correlation coefficient* exceeding 0.98.

All the independent variables except the variable X_4,the rainy days in the month of August,have emerged significant ,at least,at 5% level and reveal stability of regression coefficients. This model may forecast the rice yields moderately within tolerable limits of statistical significance. From the result of Durbin-Watson test for autocorrelations,it may be concluded that the successive values of the disturbance term i.e. stochastic term are independant. Hence,the assumption of ordinary least squares estimate that the successive values of the random variable i.e. disturbance term,are independent,has been established. From Klien's approach,multicollinearity is not a problem in the fitted multiple regression equation. Hence a crucial condition for the application of least squares that the explanatory variables are not perfectly linearly correlated is established.

The satisfactory performance of well fitted model of multiple regression equation are developed to establish the crop-weather relationship in the district of Cooch Behar.

T A B L E – 8.1

MEAN, STANDARD DEVIATION AND STANDARD ERROR OF MEAN OF INDEPENDENT AND DEPENDENT VARIABLES

VARIABLES	MEAN	STD.DEV	STD.ER.
X_1	21.412	3.374	0.818
X_2	16.706	2.823	0.685
X_3	18.706	3.496	0.848
X_4	15.421	3.432	0.832
X_5	11.529	2.322	0.563
X_6	13.706	3.704	0.898
X_7	15.882	3.238	0.785
X_8	13.235	2.705	0.656
X_9	14.059	2.680	0.650
X_{10}	3.705	2.705	0.656
X_{11}	17.294	4.469	1.084
X_{12}	14.235	3.419	0.829
X_{13}	15.235	3.632	0.881
X_{14}	9.000	5.050	1.225
Y	1194.882	140.265	34.019

X_i's–Independent variables (Nature of rainfall) Y–dependent variable.

TABLE – 8.2.

CORRELATION MATRIX

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	Y
X_1	.16	.62	.38	-.05	.33	.38	.40	.18	-.07	-.12	-.12	-.07	.41	-.396
X_2	+	.37	.33	.07	.23	-.03	-.02	-.05	.16	.02	.18	.09	-.13	-.005
X_3	+	+	.30	-.01	.15	.32	.38	.03	-.31	.28	.35	.31	.12	.076
X_4	+	+	+	-.02	.70	.05	.09	.15	-.24	-.05	-.14	-.02	.30	.082
X_5	+	+	+	+	.03	.03	.03	.005	-.31	.13	.46	.21	-.37	.059
X_6	+	+	+	+	+	.09	.08	.20	-.14	-.27	-.32	-.21	.31	-.075
X_7	+	+	+	+	+	+	.64	.69	-.38	-.18	-.18	.04	.32	-.038
X_8	+	+	+	+	+	+	+	.64	-.45	-.06	-.07	.15	.27	-.213
X_9	+	+	+	+	+	+	+	+	-.30	-.21	-.17	-.03	.29	.140
X_{10}	+	+	+	+	+	+	+	+	+	-.07	-.11	-.22	-.13	.106
X_{11}	+	+	+	+	+	+	+	+	+	+	.85	.90	-.21	-.162
X_{12}	+	+	+	+	+	+	+	+	+	+	+	.69	-.35	-.030
X_{13}	+	+	+	+	+	+	+	+	+	+	+	+	-.26	.240
X_{14}	+	+	+	+	+	+	+	+	+	+	+	+	+	.400

Y-dependent variable and X_i-explanatory variables.

T A B L E - 8.3.

MULTIPLE REGRESSION EQUATION WITH FOURTEEN EXPLANATORY VARIABLES

Parameter	Estimate Co-efficients	Std.error of coefficients	t-Value(calculated)
b_0	849.1006	183.949	4.616
b_1	-34.3120	9.230	3.717
b_2	-17.1002	8.231	2.078
b_3	70.4048	11.836	5.948
b_4	30.0341	21.970	1.367
b_5	66.1988	15.083	4.389
b_6	-48.3085	19.483	2.479
b_7	-66.0784	19.271	3.429
b_8	-42.8111	10.267	4.170
b_9	104.2897	18.419	5.662
b_{10}	44.3591	12.112	3.662
b_{11}	-2.9575	12.498	0.237
b_{12}	-108.9854	16.937	6.435
b_{13}	69.7930	13.804	5.056
b_{14}	0.6180	4.144	0.149

$R^2 = 0.9837$ $\overline{R}^2 = 0.8694$

$F(14,2) = 8.6099$ $b_u = 14$ $T = 0.02224$

TABLE - 8.4.

MULTIPLE REGRESSION EQUATION WITH THIRTEEN EXPLANATORY VARIABLES.

Parameter	Estimated Co-efficient.	Std.error of coefficient	t-value(calculated)
b_0	857.4954	143.781	5.964
b_1	-33.5842	6.433	5.221
b_2	-16.7674	6.505	2.578
b_3	69.4451	8.156	8.514
b_4	29.0431	17.193	1.689
b_5	64.8592	9.948	6.520
b_6	-47.1222	14.603	3.227
b_7	-65.0304	14.732	4.414
b_8	-43.0244	8.347	5.154
b_9	103.6104	14.653	7.071
b_{10}	43.3822	8.365	5.186
b_{11}	-2.0467	9.803	0.246
b_{12}	-108.0116	12.830	8.418
b_{13}	68.7763	9.855	6.979

R^2 = 0.9835 $\quad\bar{R}^2$ = 0.912

F (13,3) = 13.7529

b_u = 11 $\quad$ T = 0.06027

TABLE - 8.5.

MULTIPLE REGRESSION EQUATION (FITTED MODEL)

Parameter	Estimate co-efficient.	Std.error of coefficient	t-value(calculated)
b_0	846.9494	120.019	7.057
b_1	-33.3476	5.563	5.995 **
b_2	-15.9109	4.802	3.313 *
b_3	69.2490	7.100	9.754 ***
b_4	27.2104	13.547	2.009 (NS)
b_5	65.2518	8.588	7.598 **
b_6	-45.7636	11.820	3.872 *
b_7	-64.7982	12.859	5.039 **
b_8	-43.1232	7.293	5.913 **
b_9	103.7087	12.812	8.095 **
b_{10}	42.6943	6.894	6.193 **
b_{12}	-109.2711	10.286	10.623 ***
b_{13}	67.1473	6.374	10.535 ***

R^2 = 0.9832 $\bar{R}^2$ = 0.9327

F(12,4) = 19.4674** D-W Stat = 2.042

---

* - Significant at 5 percent.

** - Significant at 1 percent.

*** - Significant at 0.1 percent

NS - Non-significant.