A PROJECT REPORT

ON

WHEAT PRODUCTION IN UTTAR PRADESH :

A REGRESSION MODELLING APPROACH

UNDER THE SUPERVISION

Prof. Madhulika Dube

Professor,

Department of Statistics,

B.B.A.U Lucknow

SUBMITTED BY

Aniket Tiwari

IVRNo.20200009727

Scholarship for Higher Education (SHE) Component under INSPIRE Format for Project Completion Certificate

(To be filled by the Supervisor of Research Project)

1. IVR Number or Application Ref. No. : 202000009727
2. Name of INSPIRE Scholar: Aniket Tiwari
3. Name of College where Scholar is studying: University of Lucknow , Main Campus
4. Name of University to which above College is affiliated: University of Lucknow
5. Bank Account Number of Scholar: 39761421072
6. IFSC Code: SBIN0008314
7. Mobile No. and Email Address of Scholar: 8318138873 , aniket9454578232@gmail.com
8. Title of Project: Wheat Production in Uttar Pradesh : A Regression Modelling Approach
9. Broad Subject Area of Research Project: Study of effect of several factors in Wheat Production over a period of time.
10. Project Duration (in weeks): 12 weeks
    (Minimum 6 to 8 weeks)
11. Date of Start of Research Project: 06/06/2022
12. Date of Completion of Research Project: 01/09/2022
13. Supervisor Name: Dr. Madhulika Dube
14. Designation of Supervisor: Professor , Department of Statistics
15. Affiliation of Supervisor
    (a) Department - Department of Statistics
    (b) College :Babasaheb Bhimrao Ambedkar University, Main Campus

(c) University : Babasaheb Bhimrao Ambedkar(A Central University)

(d) Address : Babasaheb Bhimrao Ambedkar University,(A Central University) , Vidya Vihar, Raebareli Road, Lucknow – 226025

(e) Mobile no. and Email Address : 1800-180-5789 , mdube13@gmail.com

16. Major Specialization area of Supervisor: Econometrics , Regression Modelling , Biostatistics

21/11/22

डॉ० मधुलिका दुबे
प्राचार्य
सांख्यिकी विभाग
बाबासाहेब भीमराव अम्बेडकर विश्वविद्यालय
लखनऊ-226025 (केंद्रीय विश्वविद्यालय)

Date: 21/11/22

(Seal & Signature of Supervisor)

Acknowledgement

At the very outset of this project, I would like to extend my sincere and heartfelt gratitude towards all the people who have helped me in this endeavor. Without the active guidance, help, cooperation and encouragement, I would not have made headway in the project.

First of all, I would like to thank the supreme power the **Almighty God** who has always guided me to work on the right path of life, providing us with everything needed in completion of this project. As the completion of the project gave me much pleasure, I would like to express my deepest gratitude to our mentor, philosopher and guide **Prof. Madhulika Dube**, (Professor, Department of Statistics, B.B.A.U). Her dedication, support, keen interest and above all her benevolent attitude to help her students helped in the completion of project without hindrance.

We are very grateful to our teachers **Prof. Masood Siddiqui**, (Head Department of Statistics, University of Lucknow), **Prof. Rajeev Pandey**, **Prof. Sheela Misra**, **Dr. Ashok Kumar**, **Dr. Shambhavi Mishra** who helped me to reach the required research institute other than our parent institute to begin with the endeavors.

I'm also thankful to my parents for making me available the essentials used in the project and our friends who helped me with their valuable suggestions in completing this project.

Aniket Tiwari

TITLE OF RESEARCH PROJECT

Wheat Production In Uttar Pradesh : A Regression Modelling Approach

AIMS/OBJECTIVES

Among several others, the amount of rain, humidity, temperature, land area under cultivation etc. are well recognized to be the major determinants of any crop yield. The crop of wheat being the primary crop in Uttar Pradesh, the present investigation explores the effect of various predictors on the production of yield of wheat in the state of Uttar Pradesh. However, for past couple of decades extreme weathers, changing raining patterns and extensive climatic changes have started affecting the crop growth pattern and yield of various crops in various agro-climatic zones. Under such scenario, the prior forecasting of yield of field crops such as wheat via modelling techniques can help in simplifying the crop production management system.

The crop of wheat being the primary crop in Uttar Pradesh, the present investigation explores the effect of various predictors on the production of yield in the state of Uttar Pradesh. The study is therefore directed towards developing a regression model for the wheat yield of the state of Uttar Pradesh for a period of 14 years. For this, the data on wheat yield over the years has been taken along with various climatic conditions in the region. Thus study can be taken account for development and implementation of high yield measures at district, state and national levels.

CONTENTS

1 INTRODUCTION 7
1.1 UTTAR PRADESH : THE GEOGRAPHICAL PROFILE 7
1.2 UTTAR PRADESH : THE DEMOGRAPHICAL PROFILE 9
1.3 UTTAR PRADESH : THE AGRICULTURAL PROFILE 9
1.4 WHEAT PRODUCTION IN UTTAR PRADESH 10
2 DEVELOPMENT OF A REGRESSION MODEL 11
2.1 REGRESSION MODEL: AN INTRODUCTION 11
2.2 SIMPLE LINEAR REGRESSION 11
2.3 MULTIPLE LINEAR REGRESSION 11
2.4 CHECKING FIT OF THE LINEAR REGRESSION MODEL 15
2.4.1 COEFFICIENT OF DETERMINATION 15
3 COLLECTION AND ANALYSIS OF DATA 18
3.1 DATA COLLECTION SOURCES 18
3.1.1 DATA SETS 19
3.1.2 PARAMETERS OF MODELLING 20
3.1.3 SCATTER PLOTS 21
3.2 ANALYSIS OF THE DATA 22
3.2.1 REGRESSION COEFFICIENTS TABLE 22
3.2.1 ANALYSIS OF VARIANCE 24
4 RESULTS AND CONCLUSION 25
4.1 INTERPRETATION 25
4.2 CONCLUSION/SUGGESTIONS 25
REFERENCES 26

CHAPTER I

INTRODUCTION

1.1 UTTAR PRADESH : THE GEOGRAPHICAL PROFILE

Uttar Pradesh is situated in Northern India. Nepal is the international boundary of the state. The state is bordered by Rajasthan to West, Haryana and Delhi to northwest, Uttarakhand to north, Nihar to east, Jharkhand to southeast and Madhya Pradesh to southwest. It covers 93,933 miles (243290km). This is most popular state of India. It is fifth largest state of India. It accounts for 6.88% of the total area of the country. The population of the state was about 200 million as per census of 2011, which accounted for 6.49% of the total population of India. Uttarakhand was also a part of Uttar Pradesh till November 2000.

At present the state of Uttar Pradesh has 75 districts, 327 tehsils, 822 blocks and 1,07,452 villages. The state is divided into four divisions, namely Western region consisting of 30 districts, Eastern region comprising 28 districts, Central region comprising of 10 districts and Bundelkhand region comprising 7 districts of the state. Lucknow is the capital of Uttar Pradesh, Kanpur is the economic and industrial capital of the state. Varanasi is famous for 'Banarasi Sari' and also for handicraft goods. Allahabad is most famous place for Hindu religion due to confluence of Ganga, Yamuna and Saraswati. Agra is also internationally well recognized for Taj Mahal. Apart from these cities Shravasti, Khushinagar, Chitrakoot, Jhansi, Meerut, Mathura etc. are also important cities of the state. Ghaziabad and Noida are also emerging cities of the state.

The state has more than 32 large and 6 small rivers, out of them the Ganges, Yamuna, Saraswati, Saryu, Betwa and Ghagra are larger rivers of the state. The state is also divided into 9 agro-climatic zones that differ with respect to the climatic conditions, agriculture produce, density of population among several other factors. These zones are:

1.Terai Region

2.Western Plain Region

3.Central Western Region

4.South Western Region

5.Central Plain Region

6.Bundelkhand Region

7.North Eastern Plain Region

8.Eastern Plain Region

9.Vindhyachal Region

Uttar Pradesh can be divided into three distinct physio-geographical regions namely Bhabar and Terai Belt, the Gangetic Plain and the Plateau Region of Sand. The transitional belt running along the Sub-Himalayan Terai region is called the Terai and Bhabar belt. The Bhabar region is the Northern most part of Uttar Pradesh. The land of this region is very rugged and consists of large boulders and pebbles. The Bhabar tract gives place to Terai area. The terai region runs parallel to the bhabar in a thin strip. The land in terai region is very fertile to cultivate rich crops. Next to terai belt lies the Gangetic plain and includes the Ganges, Yamuna, Doab and the Ghaghra plains. This region lies between Bhabar-Terai region in the North and the plateau region in the South. Also known as the Doab and Ganga-Yamuna plain, this region is very important in terms of economic point of view. The smaller Vindhya range and plateau region is in the south of the state. While the northern part of plateau region is surrounded by the rivers Ganga and the Yamuna, the Vindhya range encompass the region in the South.

The entire alluvial plain is divided into three sub regions i.e. the eastern tract consisting of 14 districts. The flood and drought are common phenomena of this tract. The highest density of population is also found in this tract. On account of highest density of population, the per capita availability of land is very low in comparison to other tracts of the state. The other two regions i.e. the central and western are comparatively much better and well developed in comparison to the eastern and Bundelkhand regions. The irrigation facilities are also well developed in western and central regions. The cropping intensity, production and productivity of different crops of these two regions are also found much better in comparison to other regions of Uttar Pradesh.

Having an agrarian economy, nearly two-thirds of the state's workforce rely on agriculture for their livelihood. Being the largest producer of the food grain' in India, Uttar Pradesh has a variety of agro-climatic conditions that are favorable for agricultural development.While wheat is the state's primary crop, sugarcane is the state's primary commercial

crop, widely cultivated in the western and central belts of Uttar Pradesh.

1.2 UTTAR PRADESH : THE DEMOGRAPHICAL PROFILE

As per census 2011, the population of Uttar Pradesh was 199, 812, 314 of which 77.73 percent lived in rural areas followed by 22.27 percent in urban areas. The percentage of rural population of U.P. was higher than national figure of 68.84 percent. Of the total rural population of 155317 thousand male population accounted for 52.24 percent while 47.76 percent were female population in 2011 in U.P. The male population of total urban population was 52.96 percent against 47.04 percent of female population in 2011 in U.P. Of the total population of 166198 thousand in 2001 in U.P., total workers accounted for 23.67 percent. The total population of workers was 39338 thousand in 2001 in U.P. of which cultivators accounted for 46.98 percent followed by 15.14 percent, 5.32 percent, and 32.56 percent of agricultural labours, workers of industries and workers engaged in other services respectively. The density of population was 828 people per square kilometre. The sex ratio was 912 women per 1000 men in 2011. About 59 million people of the state was found below poverty line in 2004-05. The Literacy Rate of the State according to 2011 census was 70 percent which was below the national average of 74 percent. The literacy rate for men was 79 percent against 59 percent for women. Hindi is the official language of the state.

1.3 UTTAR PRADESH : THE AGRICULTURAL PROFILE

The agriculture sector continues to predominate and contributes a large share of the state output. Agriculture is main source of livelihood to majority of the population of U.P. More than 70 percent of population U.P. directly or indirectly depend on agriculture and allied sectors. The contribution of agriculture to total SDP was 24.11 percent at constant (1999-2000) prices for 2009-10. The GSDP from agriculture and allied sector at constant (1999-2000) price was 602608 million in 1999-2000 which has gone upto Rs. 748134 million in 2009-10, thereby showing 24.15 percent increase over the period. The GDP at current prices has been estimated at Rs. 862746 crores during 2013-14. The per capita income was estimated at Rs. 19233 at constant price (2004-05) and Rs. 36250 at current price. The NSDP was Rs. 403509 crores at 2004-05 price against Rs. 760542 crores at current price.

1.4 WHEAT PRODUCTION IN UTTAR PRADESH

Wheat (Triticum Aestivum) the World's largest cereal crop belongs to Graminae (Poaceae) family of the genus Triticum. It has been described as the "King of cereals" because of the acreage it occupies, high productivity and the prominent position in the international food grain trade. Wheatis consumed in a variety of ways such as bread, chapatti, porridge, flour, suji etc. The term "Wheat" is derived from many different locations, specifically from English, German and Welsh language. Wheat has good nutrition profile with 12.1 per cent protein, 1.8 per cent lipids, 1.8 per cent ash, 2.0 per cent reducing sugars, 6.7 per cent pentosans, 59.2 per cent starch, 70 per cent total carbohydrates and provides 314Kcal/100g of food. It is also good source of minerals and vitamins. For Wheat production target has been fixed in India for 2019 is 102.19 million tonnes against 100 million tonnes last year. Uttar Pradesh is the largest state with maximum contribution towards national production (35.03 per cent) from a large area (35.12 per cent), but with productivity on a lower side of 2.7 tonnes/ha. The wheat production is distributed in three agro-climatic zones, viz. western Uttar Pradesh (3.29 million ha), eastern UP (5.24 million ha) and central Uttar Pradesh (0.68 million ha). The area is 9.2 million ha, with a production of 24.5 million tons and productivity of 2.7 tonnes/ha. The trend during the last five years has shown a marginal decline in production and productivity from nearly stable area of cultivation. Per cent of gross cropped area in 2013- 2014 was 40.55. Wheat crop needs clay loam or loam texture and moderate water holding capacity soil and these features are found in Eastern Uttar Pradesh so this region is suitable for wheat production.

CHAPTER II

DEVELOPMENT OF A REGRESSION MODEL

2.1 REGRESSION MODEL: AN INTRODUCTION

The term "Regression" was coined by Francis Galton in nineteenth century to explain the relation between heights of parents and off-springs. He asserted that the descendants of tall ancestors tend to regress down towards a normal average. For Galton, regression had only this biological meaning but his work was later extended by Udny Yule and Karl Pearson to a more general statistical context. Today, the regression modelling is one of the most sought after area of active research in almost all field of human endeavors. It includes many techniques for modeling and analyzing several variables when the focus is to establish the relationship between a variable and one or more other independent variables. Its broad appeal and usefulness result from the conceptually logical process of using an equation to express the relationship between a variable of interest and a set of related predictor variables. Regression analysis is also interesting theoretically because of elegant underlying mathematics and well- developed statistical theory.

Two types of variables are primarily involved in regression model, the variable that we are trying to understand or forecast generally referred to as the variable under study or dependent variable or the target variable and the other is the set of one or more explanatory or predictor variables. The predictor variables are the factors that influence the target variable and provide us with information regarding the relationship of the variables with the target variable. Owing to its simplicity and widespread applications, the linear regression has captured perhaps the largest attention in regression modelling. The linear regression is an approach for modelling the relationship between a scalar response or dependent variable and one or more explanatory or predictor variables through a regression function. Depending upon the number of predictors used, the regression model may be classified as the Simple Linear or the Multiple Linear Regression Model described in the next sections.

2.2 SIMPLE LINEAR REGRESSION

Suppose y is the variable under study or the response and let x be the regressor variable which are related through the following regression function

(2.1) $$y = f(x)$$

If this function is linear, the relationship between i^{th} response y_i and the i^{th} predictor x_i may be written as

(2.2) $$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ ; \quad i = 1,2 \ldots n$$

where β_0 and β_1 are the coefficients in the models which are unknown, ϵ_i is the error or disturbance term or noise capturing all other factors which influence the dependent variable y_i other than the regressors x_i.

In order to estimate the unknown coefficient in (1.2), it is assumed that regressors are non-stochastic and that the errors are independently and identically distributed each having mean zero and variance σ^2 i.e.

$$E(\epsilon_i) = 0$$

$$E(\epsilon_i \epsilon_j) = 0 \qquad if \ \forall \ i \neq j$$

$$= \sigma^2 \qquad if \ \forall \ i = j$$

The most popular method of estimation of the unknown coefficients in the model is principle of least square. This method minimizes the sum of squared vertical distances between the observed response in the data set and the responses predicted by the linear approximation. In order to use the principle of least squares, suppose $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated values of the regression coefficients β_0 and β_1 in (2.2), so that the estimated value of the study variable y is given

(2.3) $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Estimation of the parameters using principle of least squares consists of minimizing the residual sum of squared where the residual is given by

(2.4) $$e_i = y_i - \hat{y}_i \qquad\qquad i = 1,2, \cdots n$$

Clearly the residual sum of square is

$$\text{RRS} = \sum e_i^2$$

$$= \sum (y_i - \hat{y}_i)^2$$

$$= \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Differentiating with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ which yield the normal equations as

(2.5) $$\frac{\partial}{\partial \hat{\beta}_0}(RRS) = 0 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

(2.6) $$\frac{\partial}{\partial \hat{\beta}_1}(RRS) = 0 = \sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Solving these equations for $\hat{\beta}_0$ and $\hat{\beta}_1$ we get the estimated value of β_0 and β_1 in the simple linear regression model (1.2) which are

(2.7) $$\hat{\beta}_o = \bar{y} - \hat{\beta}_1 \bar{x}$$

(2.8) $$\hat{\beta}_1 = \frac{Cov(x_i, y_i)}{Var(x_i)}$$

$$= \frac{1}{S_{xx}} \sum (x_i - \bar{x})(y_i - \bar{y})$$

where $S_{xx} = \sum_{i=1}^{n}(x_1 - \bar{x})^2$

2.3 MULTIPLE LINEAR REGRESSION

In case when the variables under study or the dependent variable is influenced by a number of regressor variables, the regression equation may be written as

(2.9) $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + --+\beta_{p-1} x_{i(p-1)} --+ \beta_n x_n + \epsilon_i \; ; i = 1,2, \ldots n$ In matrix notations above set of equations may be written as

(2.10) $$y = X\beta + u$$

where y is $nx1$ vector of n observation on dependent variable.

X is an nxp matrix of n observation on p independent variable.

β is $px1$ coefficient vector which is to be determined.

u is $nx1$ vector of disturbance or errors.

In order to estimate the unknown coefficient vector β using principle of least square, certain assumptions are made for the matrix of predictor variables and the disturbances in the model. These assumptions that are made in linear regression model relate to homosedastic and independent Gaussian errors and non stochastic and linearly independent regressor variables. Apart from this, independence of regressors is another very important assumption that is made in the linear regression model. However, it is often observed that among the above mentioned assumptions, one of the most frequently encountered violation of assumption relates to the correlated nature of regressors as the data in regression modeling is generally. This problem of

collinearity in regression models is referred to as the problem of multicollinearity. In case of exact linear dependence of regressors, the matrix of explanatory variables is not full column rank and hence it is not possible to find the unique estimates of unknown regression coefficients. Therefore, it is assumed that the matrix of explanatory variables capital X is non-stochastic and is of full column rank,

i.e., $$Rank(X) = p$$

Firstly, regarding the matrix of independent variable it is assumed that the nxp matrix X is is non-stochastic and is assumed to be of full column rank.

It is also assumed that the elements of disturbance vector are identically and independently distributed each having mean zero and constant variance σ^2 so that

$$E(\epsilon) = o$$

$$Var(\in) = \sigma^2 I_n$$

The consequence of non-stochastic behavior of matrix X and zero mean of disturbance leads to another interesting fact that the error are independent of the explanatory variables i.e.

$$E(X' \in) = X'E(\in) = o$$

Generalizing the method adopted in simple linear regression. Let $\hat{y}$ be estimated value of the dependent variable given by

(2.11) $$\hat{y} = X\hat{\beta}$$

then, the residuals are given by

(2.12) $$e = y - \hat{y}$$

$$= y - X\hat{\beta}$$

and the i^{th} residual is given by

(2.13) $$e_i = y_i - \hat{y}_i \quad ; \qquad i = 1,2 \cdots n$$

Using (2.12) and (2.13) it is easy to see that the residuals sum of squares is given by

(2.14) $$e'e = \left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

For minimizing the residual sum of squares,(2.7) has to be differentiated with respect to the unknown vector $\hat{\beta}$ and equated to zero so that

Differentiating with respect to the unknown vector $\hat{\beta}$, we get

$$\frac{\partial}{\partial \hat{\beta}} e'e = 0 = X'(y - X\hat{\beta})$$

Solving this equation we get the ordinary least square estimator of β which is given by

(2.15) $$\hat{\beta} = (X'X)^{-1}X'y$$

2.4 CHECKING FIT OF THE LINEAR REGRESSION MODEL

A good model always requires the correct set of regressors to be included in the model. When unnecessary explanatory variables are used in any model, not only it results in overfitting the model but also invalidates the inferences drawn for the model and may lead to erroneous results. Overfitting means that our algorithm works well on the training set but is unable to perform better on the test sets. However, when we use less number of explanatory variables than required, it may lead to underfitting of the model, again resulting in wrong inferences. Therefore, once a model is fitted to the available set of observations, there is a need to assess its fit at the observed data values. In linear regression, the commonly used checks for fitting of data include computing of **Coefficient of Determination**, **Adjusted Coefficient of Determination**, the **Root Mean Square Error** and the analyses of the pattern of residuals plotted against predictor or estimated responses in a scatter plot. These measures not only help in assessing the goodness of fit of the assumed regression model, but also help in selection of optimum subset of regressors that provides the best regression equation. These measures are dependent upon two quantities- the total variation in the response from its mean and variation of the responses from the model's predicted values. Different combinations of these two values provide different information about how the regression model compares to the mean model.

2.4.1 COEFFICIENT OF DETERMINATION

In order to understand the concept of coefficient of determination and adjusted coefficient of determination in regression models, let us first understand its computational aspect. From equation (2.13), we can write the expression of i^{th} observation in terms of its fitted or estimated value as :

(2.16) $$y_i = \hat{y}_i + e_i \quad ; \quad i = 1,2 \ldots n$$

so that the total variation in the response variable from its mean may be expressed as

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} {e_i}^2$$

or $$TSS = SSR + SSE$$

where

(2.17) $$TSS = nV(y_i) = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

is the total sum of squares of deviation in the response variable from average responses, SSR is the sum of squares of deviations explained by regressors and is given by

(2.18) $$TSS = nV(y_i) = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Lastly, $SSE = \sum_{i=1}^{n} {e_i}^2$, is the sum of squares due to residuals or errors which is the unexplained variation present in the regression.

It may be noted that the higher systematic variation due to regressors in comparison to random variation, confirms its influence on the given data set. Therefore, the coefficient of determination R^2 of a regression model defined as the proportion of variation in the values of response variable that is explained by the least-squares regression, to the total variation and is given by

(2.19) $$R^2 = \frac{SSR}{TSS}$$

The value of coefficient of determination lies between 0 and 1. Higher values of coefficient of determination indicate a good fit of the data indicating that the estimated values of the responses are close to actual values. On the other hand, if the data is not a good fit, i.e., the estimated values of the responses are unrelated to the actual values then the value of the coefficient of determination is close to zero.

One drawback of R^2 is that it increases as predictor or regressor variables are added to the regression model. In fact, this increase is artificial when predictors are not actually improving the model's fit. In view of this, another measure known as the adjusted R square(R^2_{adj}) is defined:

Definition : The adjusted R square(R^2_{adj}) is a good measure of goodness of fit of regression model that adjusts for the degrees of freedom, and is given by

(2.20) $$R^2_{adj} = 1 - \frac{(n-1)}{(n-p)}\frac{SSE}{TSS} = 1 - \frac{(n-1)}{(n-p)}R^2$$

The Adjusted R^2 measures the proportion of variation explained by only those independent variables that really affect the dependent variable and penalizes for adding independent variable that do not affect the dependent variable. R^2_{adj} does not increase necessarily when an additional regressor is added to the model. However, in general it increases then stabilizes andeventually begins to decrease as number of variables in the model increase.

CHAPTER III

COLLECTION AND ANALYSIS OF DATA

3.1 DATA COLLECTION SOURCES

The data has been collected using Internet and websites used are mentioned below:-

a)-Wheat Yield- Table 52 : STATE WISE PRODUCTION OF FOODGRAINS - WHEAT by RBI.

Website : *https://rbidocs.rbi.org.in*

b)-Land Under Cultivated, Total Area Irrigated and Fertilizers Distributed from the STATISTICAL DIARY UTTAR PRADESH- Economics and Statistics Division State Planning Institute, Planning Division, UTTAR PRADESH.

Website : *http://updes.up.nic.in*

c)-Temperature (Minimum and Maximum) and Relative Humidity from the

Website : *www.climate.nasa.gov.com*

d)-Seasonal Rainfall from RAINFALL STATISTICS OF INDIA by INDIAN METEOROLOGICAL DEPARTMENT

3.1.1 DATA SETS

Table 3.1 : Observed Data for dependent variables and predictors

YEARS	YIELD	L_{CUL}	L_{IRRI}	RAIN	$TEMP_{MIN}$	$TEMP_{MAX}$	FERTLIZERS	RH
2003-04	27.9	9443	9203	40.56	8.87	31.5	3295	53.305
2004-05	25	9374	9136	32.5	10.21	32.43	3310	46.135
2005-06	25.86	9316	9096	23.6	9.95	33.2	3464	41.743
2006-07	27.72	9390	9178	41.2	9.01	31.82	3565	44.121
2007-08	27.99	9399	9190	51.3	8.95	31.6	3756	38.65
2008-09	29.97	9669	9485	36.7	8.96	30.93	3973	60.727
2009-10	28.54	9732	9547	55.6	9.48	32.06	4261	56.098
2010-11	31.11	9801	9622	68.9	8.65	30.83	5088	54.163
2011-12	32.83	9793	9628	63.3	6.78	30.25	4258	56.709
2012-13	32.17	9785	9630	27.4	8.36	31.62	4651	52.343
2013-14	31.1	9768	9617	32.5	9.36	27.84	3842	56.928
2014-15	27.86	9645	9269	39.5	9.4	33.11	4230	44.89
2015-16	28.54	9669	9269	51.3	8.96	31.62	3973	31.495
2016-17	30.38	9885	9113	68.4	8.4	35.53	4261	41.943

3.1.2 PARAMETERS OF MODELLING

The model used is developed in relation to the prevailing climatic conditions of the state of Uttar Pradesh and some additional factors recorded during the Crop growth period. The parameters or factors of modelling included YIELD as dependent parameter and factors as the independent variables.

$Y = \text{YIELD}$ - Wheat Yield per year (Qtls Per Hecatare)

$X_1 = \text{L}_{\text{CUL}}$ - Total Land Area under wheat Production ('000 ha)

$X_2 = \text{L}_{\text{IRRI}}$ - Total Land Area Irrigated under production('000ha)

$\text{X}_3 = \text{RAINFALL}$ -Average Rainfall received during months from October-March

$X_4 = \text{TEMP}_{\text{MIN}}$ - Minimum Temperature during from Oct to March (ºC)

$X_5 = \text{TEMP}_{\text{MAX}}$ - Minimum Temperature during from Oct to March (ºC)

$\text{X}_6 = \text{RH}$ - Relative Humidity upto 2 m height (%)

$\text{X}_7 = \text{FERTILIZER}$ - Fertilizers Distributed in the state ('000 M tonnes)

Crop of wheat is sown in the month of October/November and harvested in the month of March/April. For the same the rainfall(mm) data mentioned in Table-3.1 is collected for the October, November and December for the year of lower limit of class of years and the January February and March of Upper Limit of class of Years in Table-3.1. Similarly for the Minimum and Maximum Temperature these six months are considered. Fertilizers data is the total amount of fertilizers distributed in respective years.

3.1.3 SCATTER PLOTS

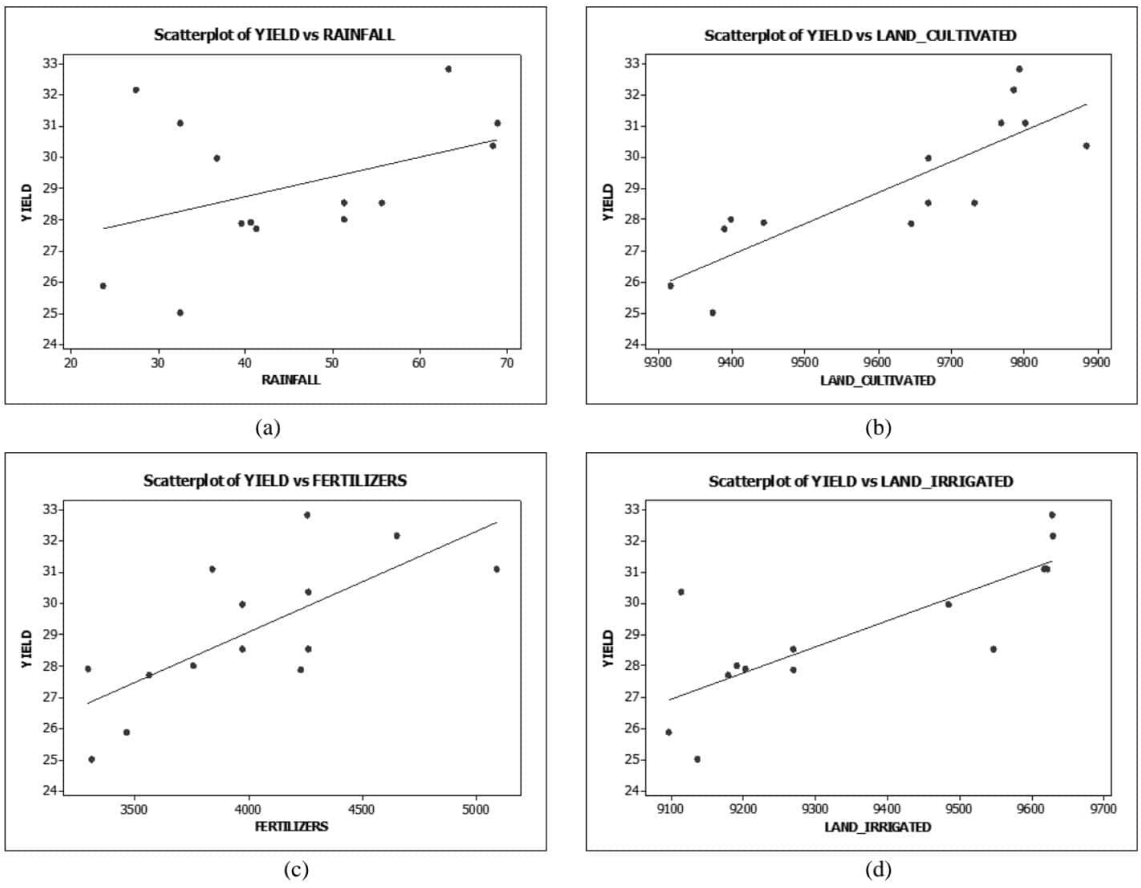

Figure 3.1 Scatter Plots between dependent variable(YIELD) and predictors.

(a) YIELD(Qtls per Hectare) vs Rainfall(mm)

(3.1) $$\text{YIELD} = 26.20 + 0.06353 \times \text{RAINFALL}$$

$$R^2 = 71.1\ \%$$

(b) YIELD(Qtls per Hectare Qtls per Hectare) vs L_{CUL}(‘000ha)

(3.2) $$\text{YIELD} = 66.74 + 0.009960 \times \text{LAND}_{\text{CULTIVATED}}$$

$$R^2 = 71.1\ \%$$

(c) YIELD(Qtls per Hectare) vs FERTILIZER(‘000M tonnes)

(3.3) $$\text{YIELD} = 16.13 + 0.003239 \times \text{FERTILIZERS}$$

$$R^2 = 52.6\ \%$$

(d) YIELD(Qtls per Hectare) vs L_{IRRI}(‘000ha)

(3.4) $$\text{YIELD} = 48.98 + 0.08342 \times \text{LAND}_{\text{IRRIGATED}}$$

$$R^2 = 63.3\ \%$$

With the help of above Figure 3.1 it is evident that the there exists a **linear relationship** between wheat yield and various factors mentioned in Figure 3.1 (a),(b),(c) and (d) that are Rainfall, Total area under Cultivation, Total Land Area Irrigated and amount of Fertilizers distributed in the state during the growth period.

Thus, proceeding with the this data we move towards the analysis part of the data mentioned below.

3.2 ANALYSIS OF THE DATA

After entering the data in Minitab(Version 16) and operating General Regression, the **regression equation** thus obtained is given below:-

(3.5) $Y = 34.3066 + 0.00312995 \times X_1 - 0.00237868 \times X_2 + 0.00688248 \times X_3 - 1.45618 \times X_4 - 0.248548 \times X_5 + 0.0012685 2 \times X_6 + 0.0452588 \times X_7$

3.2.1 REGRESSION COEFFICIENTS TABLE

Table-3.2 : Table for the coefficients determined using the Principle of Least Square Method

TERM	Coefficients	SE Coefficient	T	P	VIF
CONSTANT	34.3066	25.091	1.36725	0.221	NA
L_{CUL}	0.0032	0.0013	2.43404	0.041	4.5382
L_{IRRI}	-0.0024	0.0025	-0.96888	0.370	20.8950
RAINFALL	0.0069	0.0018	3.72974	0.010	2.0911
$TEMP_{MIN}$	-1.4562	0.1825	-7.98038	0.000	1.6355
$TEMP_{MAX}$	-0.2485	0.1693	-1.46803	0.192	6.1475
FERTILIZERS	0.0013	0.0006	1.99071	0.034	7.7927
RH	0.0453	0.0263	1.72169	0.136	3.6233

3.2.1 ANALYSIS OF VARIANCE

Table-3.3 : ANOVA Table

SOURCE	DF	SEQ SS	AJD SS	ADJ MS	F	P
Regression	7	66.7464	66.7464	9.5352	53.84353	0.000054
L_{CUL}	1	48.2130	1.0491	1.0491	5.9245	0.040882
L_{IRRI}	1	5.9890	0.1662	0.1662	0.9387	0.370222
RAINFALL	1	0.5360	2.4634	2.4634	13.9109	0.009740
$TEMP_{MIN}$	1	11.2218	11.2779	11.2779	63.6865	0.000206
$TEMP_{MAX}$	1	0.0079	0.3816	0.3816	2.1551	0.192473
FERTILIZERS	1	0.2538	0.7018	0.7018	3.9629	0.033624
RH	1	0.5249	0.5249	0.5249	2.9642	0.135907
Error	6	1.0625	1.0625	1.0625	-	-
Total	13	67.8089	-	-	-	-

CHAPTER IV

RESULTS AND CONCLUSION

4.1 INTERPRETATION

According to the data entered there are no unusual observations observed. The Standard Deviation of the data values from the Regression Line is found to be ($S = 0.420815$). The model has ($R^2 = 98.43\%$) which means that 98.43% of the variation in the datacan be explained by it. The ($R^2\ (pred) = 91.53\%$) which simply means that if the model is used for prediction it shows 91.53% variation in the Wheat Yield when the climatic data is given for upcoming years.

According to the Table-3.3, the association between L_{CUL} (Land Area Under cultivation), RAINFALL, $TEMP_{MIN}$ (minimum temperature) FERTILIZERS and YIELD comes out to be highly significant since P-value is less than assumed Level of Significance (α=0.05).

In Table-3.2,it is clearly visible that Coefficient of $TEMP_{MAX}$, $TEMP_{MIN}$ and L_{IRRI} comes out to be negative which signifies that with a unit increase in the values of these three independent variables subtracts some value (=coefficient) respectively i.e they have negative effect on the Response Variable i.e YIELD.

4.2 CONCLUSION/SUGGESTIONS

The Developed model has a capability to perform as powerful tool to analyse the effect of several factors impacting the Crop Yield in a region. It can serve good for forecasting the future yields with the available climatic data of the Study Area. It can help in simplifying the crop production management system starting from farmer's level to the policy maker.

REFERENCES

- Montgomery, D.C., Peck, Elizabeth A., Vinning, G.G., *Introduction to Linear Regression Analysis* , 3rd Edition, Wiley(India).
- Gupta, S.C, Kapoor, S.C, *Fundamentals of Applied Statistics*, 4th edition, Sultan Chand & Sons.
- Shankar, U. and Gupta, B.R.D. (1988*) Forecasting paddy yield in Bihar and Orissa states in India based on weather parameters and multiple regression technique, Tropical Agriculture, U.K, 65(3), 265-267.*
- Parthasarathy, B., Mount, A. and Kothawala, D.(1988) , *Regression model for estimation of Indian food grain production for summer monsoon rainfall, Agricultural and Forest Meteorology, 42, 167-182.*
- Singh, M.C., Pal, V., Singh, S. and Satpute, S. (2021) *Wheat yield prediction in relation to climatic parameters using statistical model for Ludhiana district of central Punjab, Agrometeorology, 23 (1), 122-126*
- Web sites – *www.updes.up.nic.in*

  www.agriculture.up.nic.in

  www.wikipidea.org

  www.statista.com

DECLARATION

Declaration by the Scholar : I **Aniket Tiwari** (full name) hereby declare that the details/facts mentioned above are true to the best of my knowledge and I solely be held responsible in case of any discrepancies found in the details mentioned above.

Aniket

(Signature of Scholar)

Date : 21/11/22
Place : Lucknow