International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(3): 88-92

https://www.mathsjournal.com
Received: 09-03-2023
Accepted: 14-04-2023

Sujata Baruah
Department of Agricultural Statistics, Assam Agricultural University, Jorhat, Assam, India

Dr. RP Paswan
Department of Agricultural Statistics, Assam Agricultural University, Jorhat, Assam, India

Borsha Neog
Department of Agricultural Statistics, Assam Agricultural University, Jorhat, Assam, India

Corresponding Author:
Sujata Baruah
Department of Agricultural Statistics, Assam Agricultural University, Jorhat, Assam, India

Factors affecting rapeseed and mustard production in Jorhat District of Assam

Sujata Baruah, Dr. RP Paswan and Borsha Neog

DOI: https://doi.org/10.22271/maths.2023.v8.i3Sb.1002

Abstract
The purpose of the study is to determine the factors affecting the rapeseed and mustard crop production in Jorhat district of Assam. The secondary data collected for the study pertained to the variables area, production, maximum temperature, minimum temperature, total rainfall, bright sunshine hours, and wind speed for the period 1988-89 to 2014-2015. The method of multiple linear regression has been employed for the study. There is a positive relationship between production of rapeseed and mustard crop and area, production and maximum temperature, as well as production and total rainfall. There is a negative relationship between production and minimum temperature, production and bright sunshine hours, as well as production and wind speed. The independent variable area is significant in the model. It can be concluded that the overall model is statistically significant for the rapeseed and mustard crop in Jorhat district of Assam.

Keywords: Factors effect, rapeseed and mustard production, multiple linear regression, Jorhat, Assam

1. Introduction

Jorhat, one of the districts of Assam, has an area of 2,851 sq. km. with a population density of 380 persons per square kilometers (2011 census). It is located in the central part of Brahmaputra valley. Jorhat is bounded by the district Lakhimpur on the north, the state Nagaland on the south, the district Sivasagar on the east and district Golaghat on the west. The river Brahmaputra flows through Jorhat. This river forms the largest riverine island of the world, Majuli in the north of Jorhat. The island Majuli spreads over 924.6 square kilometers. The mean average rainfall of the district is 2029 mm. The climate in Jorhat is warm and temperate. There is a great amount of rainfall in the summers while there is very little in the winters. The average temperature of the location is 24 ºC. The gross cropped area in the district is 152.9 thousand Ha having cropping intensity 149.5%.

Assam shares about 2.69 per cent of total rapeseed and mustard production in India and an area of 4.66 per cent in 2014-15. Rapeseed and mustard is a principal oilseed crop in Assam. Among the oilseed crops, rapeseed and mustard cultivation is done extensively. So, it is considered as a major oilseed crop. Rapeseed and mustard crop in Jorhat cover an area of 3.69 per cent of total area, sharing a production of 4.33 percent in Assam in 2014-15. Rapeseed and mustard production in Jorhat have increased from 7817 tonnes in 2012-13 to 9118 tonnes in 2013-14 and then decreased to 8129 tonnes in 2014-15.

Majuli is a sub-division of Jorhat district in upper Assam. It is a place where rapeseed and mustard is an important crop for the livelihood of the people. As per the report of KVK, Jorhat, Majuli covers an area of 8,500 hectares (Ha) of rapeseed and mustard cultivation with abundant use of the variety M-27 with a productivity of 900 Kg per Ha (Barman *et al*., 2014) [1]. Hence, decrease in the production in the grassroot level can affect the national production as well. Productivity of a crop is influenced by the availability of rainfall in abundance, favorable temperature, well conserved soil, area, production, and optimum sunshine hours.

The aim of this paper is engaged in determining the factor (s) affecting the rapeseed and mustard crop production in Jorhat district of Assam.

2. Methodology

The present work is typically based on the analyses of secondary data collected from various sources. The data on area, and production of rapeseed and mustard in Jorhat district of Assam

Over the period 1988-1989 to 2014-2015 were collected from Directorate of Economics and Statistics, Assam. The meteorological data on maximum temperature, minimum temperature, total rainfall, bright sunshine hours, and wind speed in Jorhat district of Assam over the period 1988-1989 to 2014-2015 were collected from the Department of Agrometeorology, Assam Agricultural University, Jorhat. The analytical work has been performed with the help of the statistical packages, namely, MS Excel and SPSS.

To identify the important factors affecting the rapeseed and mustard production in the Jorhat district of Assam, multiple linear regression method has been used.

In this study, six independent variables and one dependent variable have been used. Thus, the multiple linear regression is taken as,

$$Y=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

where

Y = Production of rapeseed and mustard (Tonnes).

X_1 = Area (Hectares).

X_2 = Maximum temperature (°C).

X_3 = Minimum temperature (°C).

X_4 = Total rainfall (mm).

X_5 = Bright sunshine hours (Hours).

X_6 = Wind speed (Kilometer per hour).

β_0 = Intercept.

$\beta_1, \beta_2 \ldots, \beta_6$ = Coefficients of X_1, X_2, X_6 respectively

ε = Random error term

2.1 Assumptions and diagnosis of Regression

2.1.1 Linearity

There should be a linear relationship between the dependent and explanatory variables.

2.1.2 Homoscedasticity

Error terms should have constant variance. Points in the graph should be dispersed randomly having no trend. If it is satisfied, this indicates that the assumption of homoscedasticity holds.

2.1.3 Normality of error term

The error terms should be normally distributed with the mean zero and variance σ^2. It can be tested by plotting the residual against the cumulative probability.

2.1.4 Autocorrelation

It refers to the degree of correlation between the values of the same variables across different observations in the data. A common method of testing for autocorrelation is the Durbin-Watson test. It is often included as a test in software package SPSS. Its value ranges from 0 to 4. Values close to 2 suggest less autocorrelation and values closer to 0 or 4 indicate greater positive or negative autocorrelation respectively. The residuals should be uncorrelated with the X_i's and Y_i's.

2.1.5 Multicollinearity

It refers to the existence of a high linear relationship among regressors (independent variables). It can be tested by a variance inflation factor (VIF). It is given by,

$$VIF = \frac{1}{1-R_i^2}$$

Where R_i^2 is coefficient of determination when the predictor i is regressed against all the other predictors excluding Y. The minimum value of VIF is 1. As a rule of thumb, any value of VIF > 10 may be considered to have strong evidence of i^{th} variable to be related to the other i = 1 predictor. Removing a relevant predictor from a model should be considered seriously, as it may lead to the misspecification of the model.

2.2 Definition of some related terms

2.2.1 R

It is known as the coefficient of multiple correlation. It is an index of how well a dependent variable can be predicted from a linear combination of independent variables. It ranges from 0 (zero multiple correlation) to 1 (perfect multiple correlation).

2.2.2 R^2

It is known as the multiple coefficient of determination. In the multiple variables model, the proportion of the variation in the dependent variable, Y is explained by the independent variables, X_i's (Gujarati, 2005) [3]. The general interpretation of the coefficient of determination is:

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (Y_i - \bar{Y})^2}$$

Where e_i is the error terms, $\bar{Y}$ is the mean of observed data of dependent variable, Y_i's (Source: Wikipedia).

2.2.3 Adjusted R^2

The adjusted R^2 is an effort to explain the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model. The adjusted R^2 can be negative, and its value will always be less than or equal to that of R^2. The interpretation of adjusted R^2 is as:

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

Where, p is the total number of explanatory variables in the model (excluding the constant term), and n is the sample size.

3. Results and discussion

The adequacy of the model and whether assumptions are satisfied or not have been checked prior to using the results and interpretation.

3.1 Testing the assumptions of Multiple Linear Regression

3.1.1 Linearity assumption

There should be a linear relationship among the residuals of the variables.

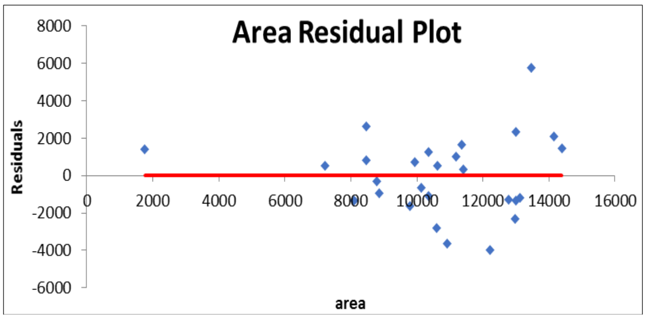

Fig 1: Residual plot of area for rapeseed and mustard crop

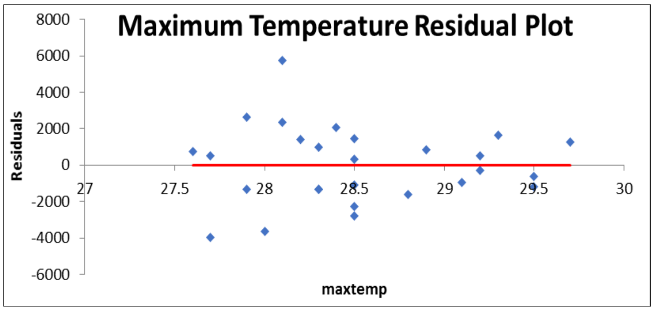

Fig 2: Residual plot of maximum temperature for rapeseed and mustard crop

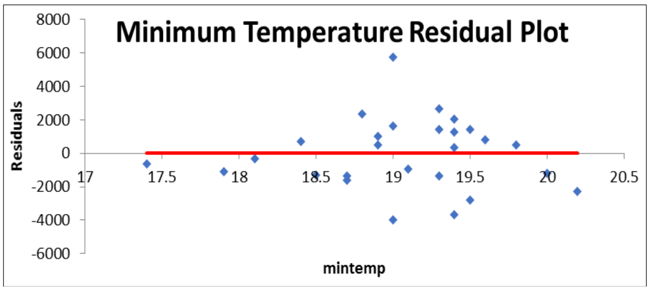

Fig 3: Residual plot of minimum temperature for rapeseed and mustard crop

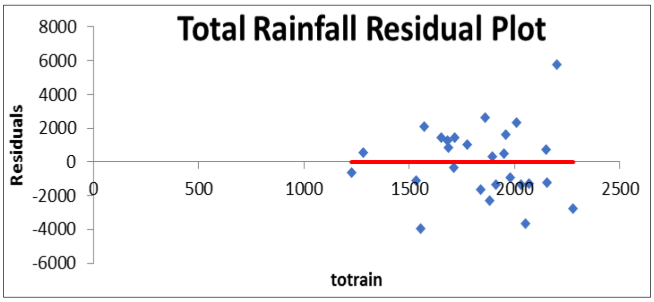

Fig 4: Residual plot of total rainfall for rapeseed and mustard crop

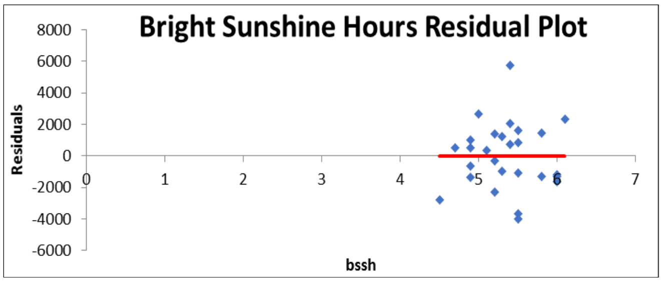

Fig 5: Residual plot of bright sunshine hours for rapeseed and mustard crop

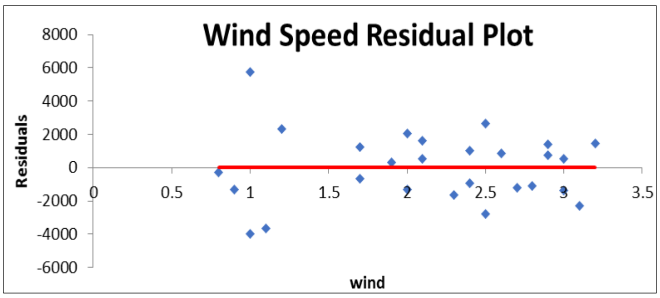

Fig 6: Residual plot of wind speed for rapeseed and mustard crop

It has been observed from figures 1, 2, 3, 4, 5, and 6 that the residual points of the variables, namely, area, maximum temperature, minimum temperature, total rainfall, bright sunshine hours, and wind speed for rapeseed and mustard crop are scattered on both sides of the straight line. Thus, the linearity assumption is fulfilled. Hence, the model is adequate.

3.1.2 Normality assumption

The error terms should be normally distributed with mean zero and variance σ^2. It can be tested by plotting residual against the cumulative probability.

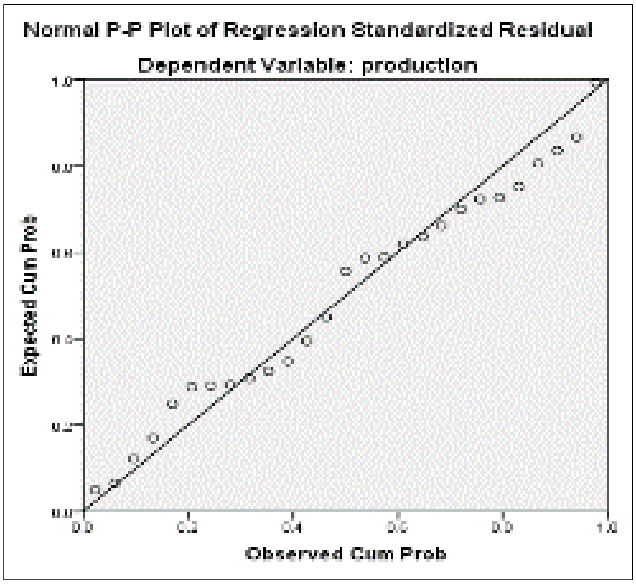

Fig 7: Normal P-P plot of regression standardized residual

Figure 7 is approximately normal. It shows approximately that there exists a normal relationship between the dependent and independent variables which signifies that the model is adequate. Hence, the normality assumption is valid.

3.1.3 Assumption of homoscedasticity

The error terms should have constant variance.

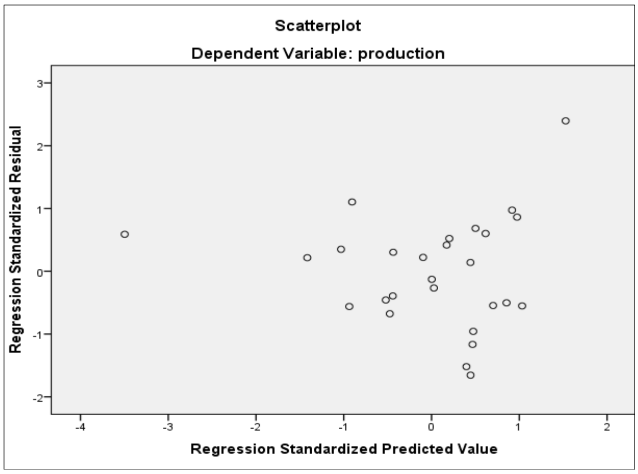

Fig 8: Normal probability plots of residuals

The points in figure 8 are seen to be dispersed in a random manner. Thus, the assumption of homoscedasticity holds true signifying that the model is adequate.

After checking the adequacy of the model, it has been found that the model remains unchanged. So, the model for the present study is:

$$Y=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

Table 1 reports the strength of the relationship between the model and the dependent variable.

Table 1: Model Summary for rapeseed and mustard crop

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.692[a]	.479	.322	2400.43745	1.797

a. Predictors: (Constant), Wind Speed, Maximum Temperature, Bright Sunshine Hours, Minimum Temperature, Total Rainfall, Area
Dependent Variable: Production

According to the given fitted model, from table 1, the coefficient of multiple correlation value (R) of 69.2% indicates that there is a good degree of predictability of the dependent variables from the independent variables and also that the coefficient of determination value (R^2), 47.9% of the variation in the production of rapeseed and mustard crop is explained by the six explanatory variables *viz*., wind speed, maximum temperature, bright sunshine hours, minimum temperature, total rainfall, and area are jointly and linearly fitted. It has also been observed from the table that the Durbin-Watson statistic value of 1.797 which is closer to 2 implies that there is almost no autocorrelation between the values of the same variables of the crop. Hence, the model is adequate.

ANOVA table is a useful test of the ability of a model to explain any variation in the dependent variable; it does not directly address the strength of that relationship.

Table 2: ANOVA for rapeseed and mustard crop

ANOVA[a]					
Model	Sum of Squares	Degrees of freedom	Mean Square	F	Significant
Regression	105798372.904	6	17633062.151	3.060	.027[b]
Residual	115241999.170	20	5762099.958		
Total	221040372.074	26			

a. Dependent Variable: Production
b. Predictors: (Constant), Wind speed, Maximum temperature, Bright sunshine hours, Minimum temperature, Total rainfall, Area

Since the model has a p-value (probability value) of $0.027 < 0.05$, therefore, it is statistically significant at 5% level of significance. It implies that at least one of the coefficients of explanatory variables is different from zero.

3.2 Interpretation of the model coefficient

The results of the regression coefficients required for the model of the rapeseed and mustard crop are presented in table 3.

Table 3: Result of the regression coefficients for the rapeseed and mustard crop

Coefficients[a]						
Model	Unstandardized Coefficients		Standardized Coefficients	T	Significant	Collinearity Statistics
	B	Std. Error	Beta			VIF
(Constant)	-9586.914	28485.852		-.337	.740	
Area	.700	.206	.632	3.395	.003	1.329
Max_Temp	497.321	853.969	.102	.582	.567	1.171
Min_Temp	-171.535	834.987	-.037	-.205	.839	1.235
Tot.Rainfall	1.591	2.039	.143	.780	.444	1.292
Bssh	-711.333	1275.163	-.102	-.558	.583	1.289
Wind_Speed	-669.772	709.277	-.171	-.944	.356	1.261

a. Dependent Variable: Production, Std. error: Standard error

From table 3, it has been observed that Variance Inflation Factor (VIF) for all the variables is less than 10 indicating that multicollinearity is absent, i.e. there is no relationship among the independent variables.
The fitted model showing the relationship between the dependent and explanatory variables is

$$Y = -9586.914 + 0.700 X_1 + 497.321 X_2 - 171.535 X_3 + 1.591 X_4 - 711.333 X_5 - 669.772 X_6$$

where Y, X_1, X_2, X_3, X_4, X_5, and X_6 are production, area, maximum temperature, minimum temperature, total rainfall, bright sunshine hours, and wind speed respectively.

A perusal of the table 3 reveals that the coefficient of area, β_1= 0.700 is positive indicating a direct relationship between production and area of the crop. This implies that when area of the crop is increased by one hectare, the production of the rapeseed and mustard crop is also increased by 0.700 tonnes keeping other explanatory variables constant. The coefficient of maximum temperature, β_2= 497.321 indicates a positive, i.e. a direct relationship between production and maximum temperature of the crop. For a one °C increase of maximum temperature, the production of the crop is increased by 497.321 tonnes keeping the remaining independent variables constant. Similarly, the coefficient of minimum temperature, β_3= -171.535 indicates an inverse relationship between minimum temperature and production of rapeseed and mustard. For a one °C increase of minimum temperature, the production of the crop is decreased by 171.535 tonnes keeping the remaining independent variables constant. The coefficient of total rainfall, β_4= 1.591 indicates the increase of production of rapeseed and mustard by 1.591 tonnes by one mm increase in total rainfall, keeping the effects of remaining variables fixed. The coefficient of bright sunshine hours, β_5= -711.333 implies that the production of the crop is increased by 711.333 tonnes when the bright sunshine is decreased for one hour, keeping the remaining variables constant. The coefficient of wind speed, β_6= -669.772 implies that the production of the crop is increased by 669.772 tonnes if the wind speed is decreased for one kmph, when the effects of the remaining variables are kept fixed. From the table 3, it has been observed that area (p-value = 0.03<0.05) has high significance at 5% level of significance. So, area has a major effect on the production of rapeseed and mustard crop in Jorhat district of Assam.

4. Conclusion

The present study assessed the factors affecting rapeseed and mustard crop production in Jorhat district of Assam. Based on the data and the results obtained from using the multiple linear regression method, the following points were concluded. The econometrical analysis reveals that there is a positive (direct) relationship between production of rapeseed and mustard crop and area, production and maximum temperature, as well as production and total rainfall. On the contrary, there is an inverse (negative) relationship between production and minimum temperature, production and bright sunshine hours, as well as production and wind speed. The independent variable area is significant in the model. Area has been identified as the key factor affecting the rapeseed and mustard production. It can be concluded that the overall model for the crop is statistically significant.

5. References

1. Barman S, Goswami J, Borgohain R. Evaluation of Front-Line Demonstration Programme on Rapeseed variety TS-38 in River Island Majuli, Assam. Journal of Academia and Research (JAIR). 2014;2:12.
2. Desalegn, T. Statistical Analysis of Factors affecting the Production of Wheat crops in West Shoa Zone: A case study at Ambo Woreda. Retrieved from https://www.academia.edu. 2014.
3. Gujarati DN. Basic Econometrics. Tata McGraw-Hill, New York; c2005.
4. https://www.jorhat.gov.in
5. 5.https://en.climate-data.org/asia/india/assam/Jorhat-764433/#:~:text=Jorhat%20climate%20summary,%C2%B0C%20%7C%2075.1%20%C2%B0F.
6. http://agricoop.nic.in/sites/default/files/ASSAM21-Jorhat-30.10.12.pdf
7. https://www.statisticssolutions.com/autocorrelation/
8. https:/www.oxfordreference.com
9. Rao AL, Ketema H. Statistical Analysis of Factors Affecting Wheat Production A Case Study at Walmara Woreda. International Journal of Engineering and Management Research. 2016;6(5):43-53.
10. stats.stackexchange.com
11. Zeru M. Assessment of Factors that affect the performance of Agricultural Production in the case of Amhara region, Ethiopia. Agricultural Sciences. 2018;9:1058-1069.