www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(4): 519-523

www.thepharmajournal.com
Received: 27-02-2023
Accepted: 30-03-2023

Sandip Kumar
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

SN Singh
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

Ravi Ranjan Kumar
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

K Kumari
Department of Entomology, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

Subrat Keshori Behera
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

Corresponding Author:
Sandip Kumar
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

Development of regression model for pre-harvest prediction of rice yield in Banka district of Bihar

Sandip Kumar, SN Singh, Ravi Ranjan Kumar, K Kumari and Subrat Keshori Behera

Abstract
For policy maker prediction of rice yield before harvest play an important role in planning, storage, marketing, price fixation, export-import decision and distribution. However, crop yields depend on various factors like weather parameters, plant conditions, yield attributing factors and different applied inputs. This study was carried out to develop a suitable regression model for pre-harvest prediction of rice yield in the year 2019-20 in Banka district of Bihar. Different nine variables for biometrical characters and one variable as a farmers' appraisal (crop condition) were considered for the present study. For this altogether sixty four samples were collected from farmer's field by using multistage stratified random sampling. Out of these ten used variables, five independent variables such as X_1 (Average plant height), X_3 (Average effective number of tillers), X_6 (Applied potassium), X_7 (Irrigation level), X_9 (Average plant condition) played important role in the development of Model-5($\hat{Y} = 31.897 - 0.709X_1 + 0.700X_3 + 0.233X_6 + 1.382X_7 + 18.696X_9$) which had minimum of coefficient of variation, RMSE, and MAE which were 11.519, 5.110, and 3.524 respectively. Adjusted R^2 (0.197) was observed to be the most suitable variables in the model. After model validation test, the value of percentage errors of this model had less than 12.16 and also from this average value 5.41. The average value of MAPE was close for selected model-5 which was best fitted for forecasting. Thus the estimated yield of rice in Banka district is about 41.25 q/ ha for the year 2019-20.

Keywords: Yield prediction, biometrical characters, farmers' appraisal, regression analysis

Introduction
Rice (*Oryza sativa* L.) is a staple food for billions of people across the world (Fraidual Islam *et al*. 2012) [1]. It contributes about 90% of global rice production and consumption in Asia (Bibi *et al*. 2014). Rice is the most important cereal food crop of India, which occupies about 24% of the gross cropped area of the country. In India, rice occupies 44.50 million hectares area with a production of 172.58 metric tons with productivity of 3.87 tons/hectare (FAO STAT-2018) [3]. In Bihar, production and productivity of rice crop are very low in comparison to other states. It is grown on an area of 3.34 million hectares with a production of 6.23 metric tons with the productivity of 2.46 tons/hectares (Directorate of economics and statistics, govt. of Bihar 2018-19). However, with the help of modern technology and proper planning, productivity should be increased. Forecast of crop yield is of immense utility to the government and planners in formulation and implementation of various policies relating to food procurement, storage, distribution, price, import-export etc. The prediction of crop yields before harvest is considered mainly as an aid to conjecture the final production and therefore, sufficient attention needs to be paid towards their improvement.
Different organizations are involved in developing methodologies before its harvest by using various approaches such as whether variables, morphological characters, agricultural inputs etc. The plant morphological characters like the number of plant population, the number of an effective tiller, length of panicle, plant height etc. may affect the yield of the crop. The agricultural inputs help in growth and its developments of the crop. The growth and development of rice crop are affected by the incidence of pest and disease infestations. Plant characters can be measured through plant use by the other approach. It can be easily assumed that plant characters are integrated effects of all the factors affecting production.
Kumar *et al.* (2019) [6] studied on Yield Estimation of Rice Crop at Pre-Harvest Stage Using Regression Based Statistical Model for Arwal District. Nath *et al.* (2018) [7] worked on pre-harvest forecasting for rice yield through Bayesian approach. Kumar *et al.* (2017) [5] worked on yield estimation of rice crop by using of biometrical characters along with farmer's appraisal

and develop forecasting model. Pandey *et al*. (2013) [8] suggested models for forecasting rice yield in eastern U.P based on weather variables and weather indices (1989-90 to 2009-10) and screened out the weather variables and estimated the model parameters through multiple regression approach.

Materials and Methods

The present investigation was carried out based on following procedure.

Sampling procedure

Multi stage random sampling methods were used select different villages of blocks. AT the First stages blocks were selected purposively, then at the second stage panchayats were selected randomly. In Third stage villages were selected and last in fourth stage two plots of each farmer were selected by SRSWOR (simple random sampling without replacement). Total Sixty four samples were selected in Banka district.

Identification of relevant variables

The plant characters such as average plant height, average plant population, average effective number of tillers, chemical fertilizer, disease and pest infestation etc. were collected for the rice yield prediction of Banka district in Bihar.

By keeping views in mind, following ten parameters were considered for the yield prediction of rice crop in Banka district of Bihar.

Table 1: List of measurable and non-measurable characters

S.N.	Variables	Code of variables	Unit of measurement	Types of characters
1.	Yield	Y	q/ha	Measurable
2.	Average plant height	X_1	cm	Measurable
3.	Average plant population	X_2	number	Measurable
4.	Average effective number of tillers	X_3	number	Measurable
5.	Applied nitrogen	X_4	Kg/ha	Measurable
6.	Applied phosphorous	X_5	Kg/ha	Measurable
7.	Applied potassium	X_6	Kg/ha	Measurable
8.	Irrigation level	X_7	number	Measurable
9.	Disease and pest infestation	X_8	percent	Measurable
10.	Average plant condition	X_9	Eye estimate	Non-measurable

Collection of data and models development

All the primary data such as Plant population, Plant height, Effective number of tillers, Irrigation level, applied nitrogen, phosphorus, potassium, and disease and pest infestation were recorded by self-observations and by personal interviews. By the self-observations, data were recorded from the farmer's field in the area of 1 square meter.

All models were developed by using stepwise regression technique which was based on forward selection of independent variables. These procedures were totally computer intensive procedure and executed using software R (version 3.6.3). Best five models were selected on the basis of minimum AIC values which were used for regression analysis.

Selection of appropriate subset for regression

With the help of software R (version 3.6.3), regression analysis was carried out of selected best five model. On the basis of R^2, Adj.R^2, RMSE, C.V and MAE criteria best sub model was selected.

Model validation through suitable statistical tools

Prevailing assumptions to validation of model were considered *viz*.

1. Relationship between the dependent variable(Y) and independent variables (X's) should be linear.
2. The error should be uncorrelated.
3. Errors should be normally distributed with mean zero and constant variance, σ^2

Results and Discussion

All observed variables were used for development of different models. With the help of software R (version 3.6.3), best five models were selected based on minimum AIC value which were presented in the Table 2.

Table 2: Best five models selected for regression analysis based on minimum AIC value

S.N.	Model	Number	R Square	Adjusted R^2	RMSE	AIC
1.	X9	1	0.877	0.875	6.261	381.3409
2.	X9, X7	2	0.898	0.895	5.748	372.3805
3.	X9, X7, X1	3	0.911	0.906	5.428	366.6756
4.	X9, X7, X1, X3	4	0.920	0.913	5.211	362.8596
5.	X9, X7, X1, X3, X6	5	0.924	0.917	5.110	361.4748

Table 3: Residual results of selected models

S.N.	R. Square	Adj.R^2	RMSE	Coefficient of Variation	MAE
1.	0.877	0.875	6.261	14.114	4.725
2.	0.898	0.895	5.748	12.958	4.289
3.	0.911	0.906	5.428	12.237	3.892
4.	0.920	0.913	5.211	11.748	3.562
5.	0.924	0.917	5.110	11.519	3.524

All possible statistical analysis carried out to compute for 58 observations using software R (version 3.6.3). From the table 2 were it was reflected that if the number of explanatory variables were increased in the model, the value of R^2 and Adj. R^2 were also increased for that particular model and the value of RMSE and AIC were decreased for that particular model. Among all the models, the model-5 was found with highest value of coefficients of determination (R^2) and Adjusted R^2 and lowest value of RMSE, C.V, MAE table 3. After step-down regression selection method it was concluded that the model-5 ($\hat{Y} = 31.897 - 0.709X_1 + 0.700X_3 + 0.233X_6 + 1.382X_7 + 18.696X_9$) was found to be the best fit for the prediction of rice yield of Banka district in Bihar. The model-5 with five regression subset viz. X1, X3, X6, X7 and X9 whose most of the parameters were highly significant at 1% level of significance along with intercept. The value of coefficient of determination (R^2) was 0.924 indicated that 92.40% variation was explained by independent variables which were included in the model and the adjusted R^2 value was 0.917 which was higher than the other models which reflected that the best fitted model for prediction of yield. The value of RMSE and coefficient of variation (C.V) were 5.110 and 11.519 respectively which were within considerable range which were lower than of the other rest models. In ANOVA F value was also highly significant for this model. In Graph for the model-5 (Fig-1) it was reflected that residual variance was

more or less homogeneous in nature which was better fitted as compared to other models. Normal Q-Q plot showed that residuals were present in normal condition (Fig-2). The fitted residual which were green in color which showed the smaller distance and red color showed the far distance which indicated that error was lesser which were in green color.

Regression Model – 5

The fitted equation for the regression model -5 is mentioned below:

$$\hat{Y} = 31.897 - 0.709X_1 + 0.700X_3 + 0.233X_6 + 1.382X_7 + 18.696X_9$$

Table 4: Parameter estimates of 5th model after regression analysis

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	31.897	36.734	0.868	0.389
X9	1	18.696	0.862	21.701	0.000**
X7	1	1.382	0.459	3.013	0.004**
X1	1	-0.709	0.310	-2.286	0.026*
X3	1	0.700	0.319	2.193	0.033*
X6	1	0.233	0.132	1.768	0.083

Anova					
Source	**DF**	**Sum of Squares**	**Mean Sum of Square**	**F Value**	**Pr > F**
Model	5	16531.832	3306.366	126.639	0.000**
Error	52	1357.650	26.109		
Corrected Total	57	17889.482			

Note: ** (1% level of significance)

Root MSE	5.110	R. Square	0.924
MAE	3.524	Adj. R. Square	0.917
Coeff. Var.	11.519		

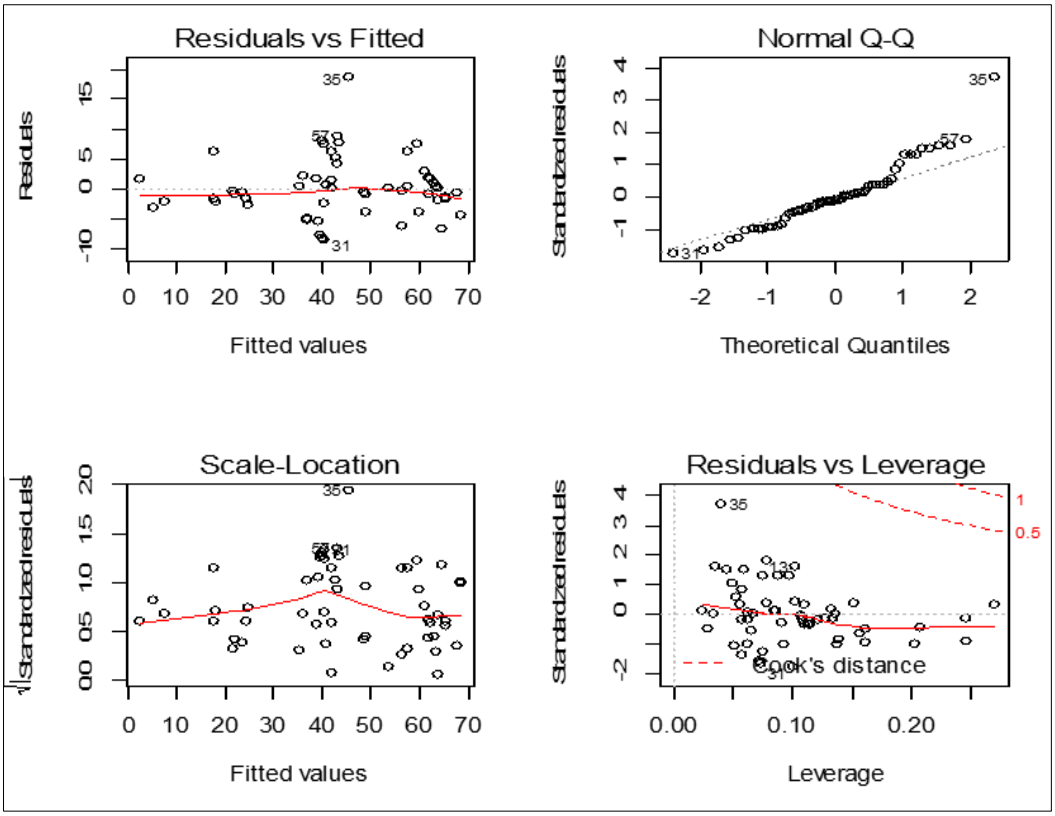

Fig 1: Residuals analysis for Model-5

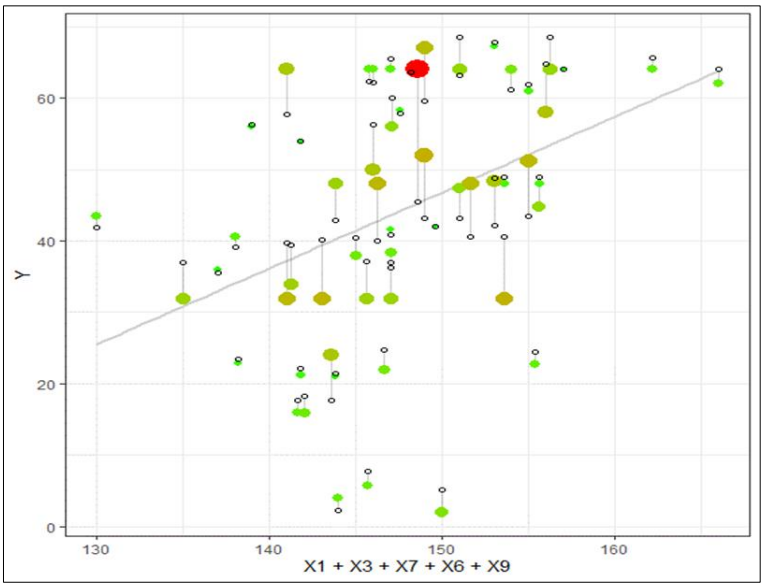

Fig 2: Fitted vs. Residuals for Model-5

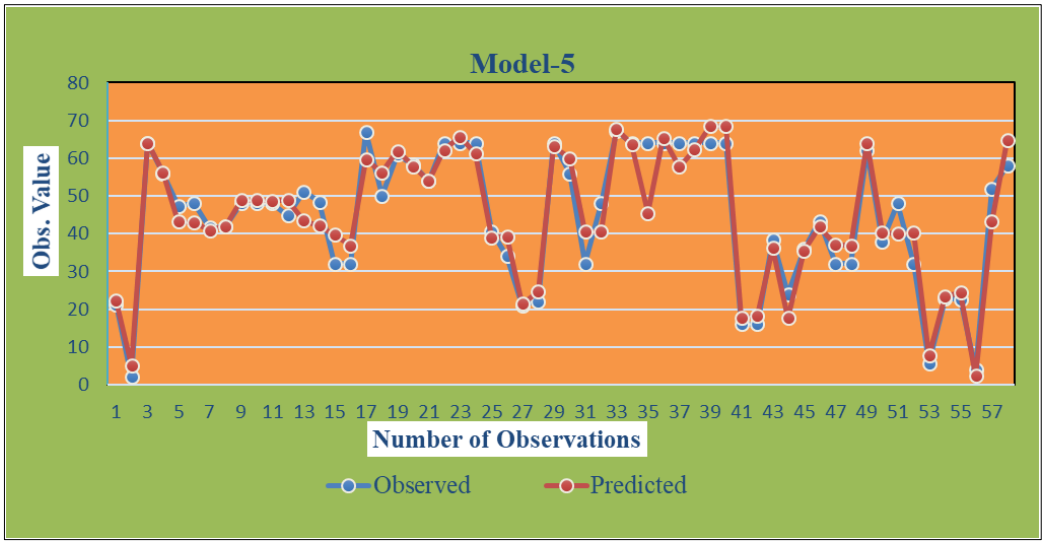

Fig 3: Observed vs. Predicated for model-5

Table 5: Estimating error for the six sets of observations which are not included in model building (5th Model)

S.N.	X1	X3	X6	X7	X9	Y	$\hat{Y}$	$\hat{e}_i = Y - \hat{Y}$	$MAPE = \frac{\lvert\hat{e}_i\rvert}{\hat{Y}} \times 100$
1	118	16	0	5	3	20	22.432	-2.432	12.16
2	117	12	24	5	3	24	25.933	-1.933	8.05
3	118	15	18	4	3	25	24.545	0.455	1.82
4	116	13	19.2	5	4	42	44.920	-2.92	6.95
5	117	17	14.4	6	5	64	65.971	-1.971	3.07
6	118	15	13.71	6	5	64	63.701	0.299	0.46
7	Average=5.41								

The set of six observations mentioned in Table 5, were not used for model formulation. For each set of observation, the estimated deviation and Mean Absolute Percentage Error of prediction presented in this table. After model validation test, it was found that the value of percentage error of this model was less than 12.16 and 5.41 from the average value. The average value of MAPE was close to selected model therefore, we can say that the model was best fitted for forecasting. After using the model-5, the predicted yield of rice was found to be 41.25 q/ha for the year 2019-20 of Banka district in Bihar.

References

1. Islam ASMF, Ali MR, Gregorio GB, Islam MR. Genetic diversity analysis of stress tolerant rice (*Oryza sativa* L.). African Journal of Biotechnology. 2012;11:15123-9.
2. Bibi A, Naheed S, Raja I, Khan N. Past trends and future prospects of rice area and production in Khyber Pakhtum Khaw. International Journal of Agronomy and Agricultural Research. 2014;4(1):40-46.
3. Anonymous. Statistical data on area and production of paddy crop in India. F.A.O statistics; c2018. http://fao.org/faostat/en/#data/QC

4. Anonymous. Statistical data on area and production of paddy crop during season 2018. Directorate of Economics and Statistics, Government of Bihar; c2017.
5. Kumar M, Singh MM, Kumar S. Pre-harvest forecasting of rice yield using biometrical characters along with farmer's appraisal in Muzaffarpur district of Bihar. International Journal of Pure and Applied Bioscience. 2017;5(5):1553-1558.
6. Kumar RR, Singh SN, Kumari K, Nath B. Yield estimation of Rice crop at pre harvest stage using regression based statistical model for Arwal District, Bihar, India. International Journal of Current Microbiology and Applied Sciences. 2019;8(8):2491-2500.
7. Nath B, Singh SN, Rai GK. Pre harvest forecast of rice yield for Bhagalpur district in Bihar. Journal of Pharmacognosy and Phytochemistry. 2018;7(6):2342-2345.
8. Pandey KK, Rai VN, Sisodia BVS, Bharti AK, Gairola KC. Pre-harvest forecast models based on weather variables and weather indices for eastern U.P. Advance in Bioresearch. 2013;4(2):118-122.