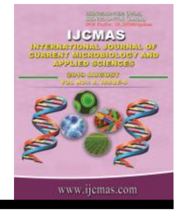
Int.J.Curr.Microbiol.App.Sci (2019) *8*(8): 2491-2500

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 8 Number 08 (2019)**
Journal homepage: http://www.ijcmas.com



Original Research Article https://doi.org/10.20546/ijcmas.2019.808.290

Yield Estimation of Rice Crop at Pre-Harvest Stage Using Regression Based Statistical Model for Arwal District, Bihar, India

Ravi Ranjan Kumar*, S.N. Singh, Kiran Kumari and Bhola Nath

Department of Statistics, Mathematics and Computer Application
Bihar Agricultural University, Sabour, Bhagalpur, Bihar - 813210, India

**Corresponding author:*

Keywords

Yield estimation, Bio-metrical, Characters of rice, Farmer's appraisal, Regression technique

Article Info

Accepted:
22 July 2019
Available Online:
10 August 2019

A B S T R A C T

The estimation of crop yield before harvest helps in different policy making in an order for storage, distribution, marketing, pricing, import-export etc. Crop productions depend on several factors such as weather factors, plant characters and agricultural inputs. The present study was carried out to develop the appropriate statistical model for estimation of rice yield before harvest in the year 2018-19. This research was done on plant biometrical characters along with farmer's appraisal. Sample survey was done on farmer's field through multistage stratified random sampling method and recorded fourteen parameters such as X_1 (Number of irrigation), X_2 (Average plant population), X_3 (Average plant height), X_4 (Average number of effective tillers), X_5 (Average length of panicle), X_6 (Average length of flag leaf), X_7(Average width of flag leaf), X_8 (Average number of filled grain), X_9 (Damage due to pest and disease infestations), X_{10} (Applied nitrogen), X_{11} (Applied phosphorus), X_{12} (Applied potassium), X_{13} (Average plant condition) and Y (Yield). By the help of step-wise regression technique to select thirteen models on the basis of minimum BIC value and then after best models were selected on the basis of minimum AIC value. After regression analysis, one best fitted model was selected on the basis of some important statistics such as RMSE, R^2, Adj.R^2, C.V, Residual and Cook's D statistic. However, 10 % observations were kept for model validation test purpose. Model -2($\bar{Y}$= 27.07355-1.69966X_1 + 0.25058X_2 + 0.24110X_4 + 1.28741X_5-0.45193X_6 + 1.17152X_{13}) had minimum value of coefficient of variation, residual, and student residual which were 6.36430, 0.0000, and -0.0756 respectively. Value of Adj.R^2 (0.8197) which indicated the better to fit of variables in the model. After model validation test, the lowest value of MAPE (1.18 – 5.48) were indicated the good precision for model-2. Thus the estimated rice yield in Arwal district is about 33.28 q/ ha for the year 2018-19.

Introduction

Rice (*Oryza sativa*) is one of the most important cereal crops in India. It is the staple food for millions in the world and feeds more than half of humanity on a daily basis and provides a major and most stable source of income. It is cultivated on 42.96 million hectares of land and producing 158.75 metric tons rice with productivity of 3.95 tons/hectare (F.A.O STAT, 2016). Bihar is also an important rice growing state in the country. Rice is grown on 3.34 million hectares of land and producing 8.24 metric tons with productivity of 2.46 tons/hectare (Directorate of Economics and Statistics, GoB., 2016-17). However, after using the available technology and proper demonstration, it is possible to increase the productivity. The estimation of crop yield before harvest helps in different policy making in an order for storage, distribution, marketing, pricing, import-export etc. (Vogel and Bange, 1999). The estimation of crop yields before harvest are considered mainly as an aid to conjecture the final production and therefore, sufficient attention needs to be paid towards their improvement. That is not only deals with developing model but also considered the accuracy of the model.Thus reliable and timely forecasting of crop yields before harvest are very important.

Different kinds of organisation are involved in developing methodologies before harvest by using various approaches such as plant biometrical characteristics, weather variables, agricultural inputs etc. These approaches can be used individually or in combination. The plant morphological characters like number of plant per plot, number of tillers per plant, numbers of grain per panicle etc. may affect directly and other characters like plant height, leaf area, panicle length etc. may affect indirectly the yield of crop. Chemical fertilizers are helps in growth and development of the crop and incidence of disease and pest infestations are also affected the growth, development and the crop yield.

Nath *et al.,* (2018) worked on pre-harvest forecasting for rice yield through Bayesian approach. Deep *et al.,* (2018), Kumar *et al.,* (2017) worked yield estimation of rice crop by using of biometrical characters along with farmer's appraisal and develop forecasting model. Pandey *et al.,* (2013) suggested models for forecasting rice yield in eastern U.P based on weather variables and weather indices (1989-90 to 2009-10). They used stepwise regression to screen out the weather variables and estimated the model parameters through multiple regression approach.

Materials and Methods

The present investigation was carried out by following steps.

Sampling technique:

By using multi stage stratified random sampling method, samples were selected in different villages of blocks. In First stages blocks were selected purposively, then in 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 simple random system.Total Sixty samples were selected in Arwal district.

Recognition of measurable as well as non-measurable characters

The characters like number of plant per plot, number of tillers per plant, numbers of grain per panicle, plant height, leaf area, panicle length, chemical fertilizers, disease incidence and pests etc. were taken for the yield estimation of rice crop in Arwal district of Bihar.

Data collection and development of regression model

The primary data such as plant population, plant height, number of effective tillers, length of panicle, length of flag leaf, width of flag leaf, number of filled grain per panicle, level of irrigation, applied nitrogen, phosphorus, potassium, 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 one square meter.

Identification of appropriate subset for regression study

With the help of SAS v 9.3, regression analysis was carried out of selected best five model. On the basis of R^2, Adj.R^2, RMSE, Residual analysis and Cook's D criteria best sub model has to be chosen.

Application of statistical tools to test the validity of regression models

For validity of regression models, following major assumptions was considered:

The relationship between the dependent variable(Y) and independent variables (X's) should be linear in nature

The error terms which are assumed to be normally and independently distributed will zero mean and constant variance.

Results and Discussion

All the parameters were used for the development of different models. By using software SAS JMP v 13.0, eight thousand one hundred ninety-two different combinations of regression models were developed. On the basis of minimum BIC value, thirteen best models were highlighted for each term. Out of these thirteen highlighted models, five best models were selected based on the least AIC value which were given in the Table 2.

The all possible statistical analysis was carried out to compute for 54 observations through software SAS v 9.3. From the table 2. The model-1 had four explanatory variables and model-3 had five explanatory variables. For 3rd model the value of R^2 was higher than from the 1th model. That was increment of 0.0074, which was less than 0.01. The value of Adj.R^2 for 3rd model there was increment of 0.0042 which was also very less which showed that there was no need of extra X_4 regressor was for the model- 3. From the model–2, which had six explanatory variables whose value of R^2 was 0.8401. In which there was increment of 0.0084 from the 3rd model and increment of 0.0158 from the 1st model. The value of Adj.R^2 for the 2nd model was 0.8187 that was more than 0.0056 and 0.0098 from the 3rd and 4th model respectively which had higher increment in value as compare to other models. So extra X_2 and X_4 regressors were sufficient for the model-2. The model-4 had seven sub set regression model, 0.8449R^2 values that was increment of 0.0048 from the 2nd model. It was not sufficient in the 4th model. The value of Adj.R^2 in the 4th model was 0.8212 which had 0.0025 increments as compare the 2st model that was very less value, so extra X_{12} variable was not significant for the model-4. From the model-5, which had eight regressors and its value R^2 was 0.8495 and Adj.R^2 value was 0.8228. Both the values had very less precision of results as compare to the model-3 and 4. Hence there was no need to include regressors X_3 and X_9 in the model-5.We may concluded that the Model-2($\bar{Y}$= 27.07355-1.69966X_1 + 0.25058X_2 + 0.24110X_4 + 1.28741X_5-0.45193X_6 + 1.17152X_{13}) was best to fit for the estimation of rice yield in Arwal district of Bihar. It had six regressors viz. X1, X2, X4, X5, X6 and X13 whose most parameters were significant at 1% level of significance along with intercept. The increment of Adj.R^2 value was higher as compared to other models. All observations of residuals were lesser than other models showed that the best fitted model for the predicting yield. The value for coefficient of variation, residual, and student residual for model-2 were 6.36430, 0.0000, and -0.0756 respectively. Which were lower than other model. The analysis of variance (ANOVA) for this model showed that the F value was highly significant at 1% level of significance. Graph of the model-2 (fig-3) showed that low value of residual for most of the observations showed the good accuracy for the model. Variance of inflation were less than two which showed that there was no any sign of multi-collinearity for the parameters.

The set of six observations which were given in Table 4, that corresponds to the variables have been included in the model. These observations were not used in model building. For each set of observation, the estimated deviation and mean absolute percentage error of prediction has been presented.After model validation, it was found that the value of percentage error as this model had less than 5.48 and 2.5600 average value. That indicated that model was used with good accuracy to estimate rice yield. So it was used for estimation of rice yield in Arwal district of Bihar for the year 2018-19. After using the model-2, the estimated yield of rice was found be about 33.28 q/ ha for the year 2018-19. This is totally based on biometrical characters and farmer's appraisal.

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.	Number of irrigation	X_1	numbers	Measurable
3.	Average plant population	X_2	per m^2	Measurable
4.	Average plant height	X_3	cm	Measurable
5.	Average number of effective tillers	X_4	per m^2	Measurable
6.	Average length of panicle	X_5	cm	Measurable
7.	Average length of flag leaf	X_6	cm	Measurable
8.	Average width of flag leaf	X_7	cm	Measurable
9.	Average number of filled grain	X_8	numbers	Measurable
10.	Damage due to pest and disease infestations	X_9	percent	Measurable
11.	Applied nitrogen	X_{10}	kg/ha	Measurable
12.	Applied phosphorus	X_{11}	kg/ha	Measurable
13.	Applied potassium	X_{12}	kg/ha	Measurable
14.	Average plant condition	X_{13}	eye estimate	Non-measurable

Table.2 Five best models for regression analysis

S.N.	Model	Number	R.Square	Adj.R^2	RMSE	AIC
1	X_1,X_5,X_6,X_{13}	4	0.8243	0.8099	2.6091	265.359
2	$X_1,X_2,X_4,X_5,X_6,X_{13}$	6	0.8401	0.8197	2.5413	265.677
3	X_1,X_4,X_5,X_6,X_{13}	5	0.8317	0.8141	2.5802	265.691
4	$X_1,X_2,X_4,X_5,X_6,X_{10},X_{13}$	7	0.8449	0.8212	2.5303	266.94
5	$X_1,X_2,X_3,X_4,X_5,X_6,X_9,X_{13}$	8	0.8495	0.8228	2.5195	268.314

Table.3 Parameter estimates of 2nd model after regression analysis

Variable	D.F	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	27.07355	8.49181	3.19	0.00**	0
X_1	1	-1.69966	0.13203	-12.87	0.00**	1.11663
X_2	1	0.25058	0.15903	1.58	0.12	1.04655
X_4	1	0.24110	0.15082	1.60	0.11	1.01870
X_5	1	1.28741	0.29453	4.37	0.00**	1.07605
X_6	1	-0.45193	0.13613	-3.32	0.00**	1.14886
X_{13}	1	1.17152	0.42964	2.73	0.00**	1.06339
ANOVA						
Source		**DF**	**Sum of Squares**	**Mean Sum of Square**	**F Value**	**Pr > F**
Model		6	1594.83501	265.80583	41.16	0.00**
Error		47	303.53318	6.45815		
Corrected Total		53	1898.36819			

Note:- ** (1% level of significance)

Table.4 Residual analysis of 54 observations used in 2nd model

Obs.	Dependent Variable	Predicted Value	Std Error of Mean Predicted	Residual	Std Error Residual	Student Residual	Cook's D
1	42.0000	43.4394	1.1264	-1.4394	2.278	-0.632	0.014
2	40.0000	40.4822	0.8362	-0.4822	2.400	-0.201	0.001
3	44.5000	46.5284	0.7599	-2.0284	2.425	-0.836	0.010
4	42.2500	40.8277	0.9412	1.4223	2.361	0.603	0.008
5	45.7500	49.2314	0.9217	-3.4814	2.368	-1.470	0.047
6	43.6600	46.6864	0.8635	-3.0264	2.390	-1.266	0.030
7	44.6400	43.2961	0.7291	1.3439	2.434	0.552	0.004
8	45.5000	43.3489	1.1980	2.1511	2.241	0.960	0.038
9	44.7000	44.9995	0.5815	-0.2995	2.474	-0.121	0.000
10	46.4000	48.4227	0.8432	-2.0227	2.397	-0.844	0.013
11	42.0000	41.3669	0.7971	0.6331	2.413	0.262	0.001
12	43.0000	41.1758	0.7986	1.8242	2.413	0.756	0.009
13	46.7500	40.3434	0.9242	6.4066	2.367	2.706	0.159
14	41.5100	40.3956	0.9945	1.1144	2.339	0.477	0.006
15	34.4000	30.4485	1.0935	3.9515	2.294	1.723	0.096
16	31.3500	34.8525	0.8971	-3.5025	2.378	-1.473	0.044
17	47.0200	46.3339	0.8205	0.6861	2.405	0.285	0.001
18	44.0000	45.2479	0.8686	-1.2479	2.388	-0.523	0.005
19	34.0000	33.2203	0.8005	0.7797	2.412	0.323	0.002
20	35.2000	32.8777	0.7018	2.3223	2.442	0.951	0.011
21	44.0000	45.1228	0.6051	-1.1228	2.468	-0.455	0.002
22	46.2400	45.2190	1.0910	1.0210	2.295	0.445	0.006
23	39.4000	40.4533	1.6157	-1.0533	1.962	-0.537	0.028
24	29.6000	31.0823	0.7525	-1.4823	2.427	-0.611	0.005
25	43.7000	42.2871	1.1441	1.4129	2.269	0.623	0.014
26	40.0000	41.1482	0.6896	-1.1482	2.446	-0.469	0.003

27	46.0000	46.4654	1.5684	-0.4654	2.000	-0.233	0.005
28	48.3300	44.8749	0.6973	3.4551	2.444	1.414	0.023
29	45.0000	39.3926	0.4502	5.6074	2.501	2.242	0.023
30	48.1400	45.5058	0.9597	2.6342	2.353	1.119	0.030
31	44.6600	43.4414	1.0155	1.2186	2.330	0.523	0.007
32	40.3300	40.3102	0.9939	0.0198	2.339	0.00846	0.000
33	45.9400	47.7767	0.9833	-1.8367	2.343	-0.784	0.015
34	44.8200	42.9986	0.6952	1.8214	2.444	0.745	0.006
35	41.2500	42.7543	0.9436	-1.5043	2.360	-0.638	0.009
36	44.0000	41.3742	0.6626	2.6258	2.453	1.070	0.012
37	34.0000	36.3670	0.9061	-2.3670	2.374	-0.997	0.021
38	29.1600	33.5334	0.6828	-4.3734	2.448	-1.787	0.035
39	31.6600	36.9424	0.8480	-5.2824	2.396	-2.205	0.087
40	30.0000	31.6787	0.9290	-1.6787	2.365	-0.710	0.011
41	42.0000	43.2402	0.6935	-1.2402	2.445	-0.507	0.003
42	43.0000	43.4860	1.1801	-0.4860	2.251	-0.216	0.002
43	29.0000	29.9153	1.1924	-0.9153	2.244	-0.408	0.007
44	25.0000	25.9059	1.0531	-0.9059	2.313	-0.392	0.005
45	30.0000	31.8661	0.6840	-1.8661	2.447	-0.762	0.006
46	36.0000	34.2859	0.8792	1.7141	2.384	0.719	0.010
47	37.3300	37.1080	0.9918	0.2220	2.340	0.0949	0.000
48	43.2800	41.0747	0.6733	2.2053	2.450	0.900	0.009
49	37.3300	36.5335	1.2133	0.7965	2.233	0.357	0.005
50	37.0000	32.6508	0.8588	4.3492	2.392	1.818	0.061
51	33.3300	36.2254	0.9486	-2.8954	2.358	-1.228	0.035
52	39.6600	39.6944	0.6335	-0.0344	2.461	-0.0140	0.000
53	35.1200	36.9975	0.6107	-1.8775	2.467	-0.761	0.005
54	33.3300	35.0028	0.5149	-1.6728	2.489	-0.672	0.003

Table.4 Estimating error for the six set of observations which are not included in model building (2^{nd} model)

S.N.	X_1	X_2	X_4	X_5	X_6	X_{13}	Y	$\hat{Y}$	$\hat{e}_i = Y - \hat{Y}$	$MAPE = \frac{\lvert\hat{e}_i\rvert}{\hat{Y}}$
1.	12	19	16	23.6	41.2	4	30	31.74	-1.74	5.48
2.	12	22	13	20.2	39	3	28	27.21	0.79	2.90
3.	13	27	15	22.8	38.5	4	31.5	31.99	-0.49	1.53
4.	13	26	17	23.8	41.2	5	32.5	33.47	-0.97	2.89
5.	10	26	15	23.4	38.5	5	39.25	38.79	0.46	1.18
6.	10	23	17	22.9	39	4	37	36.48	0.52	1.42

Fig.1 Diagnostic fit for dependent variable (Y)

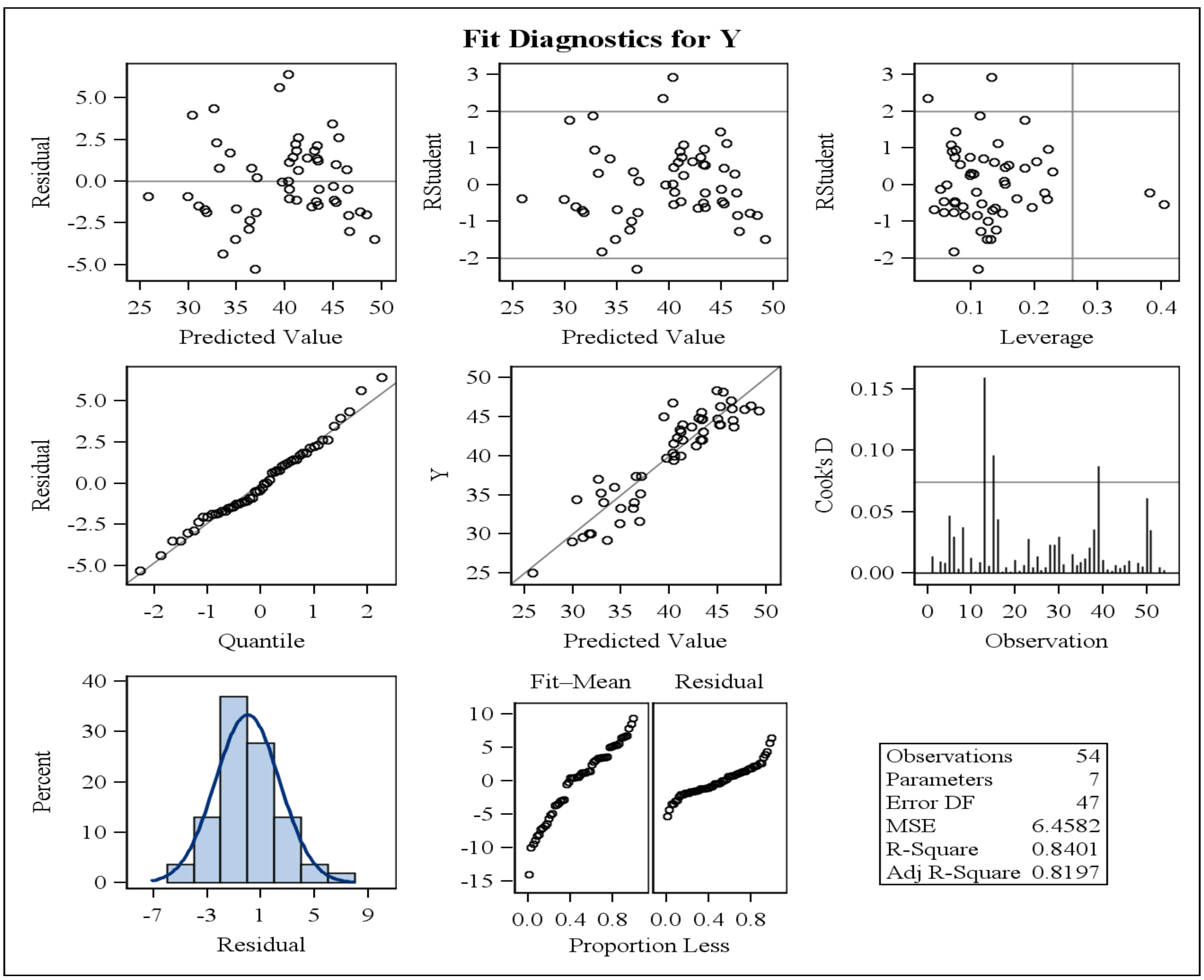
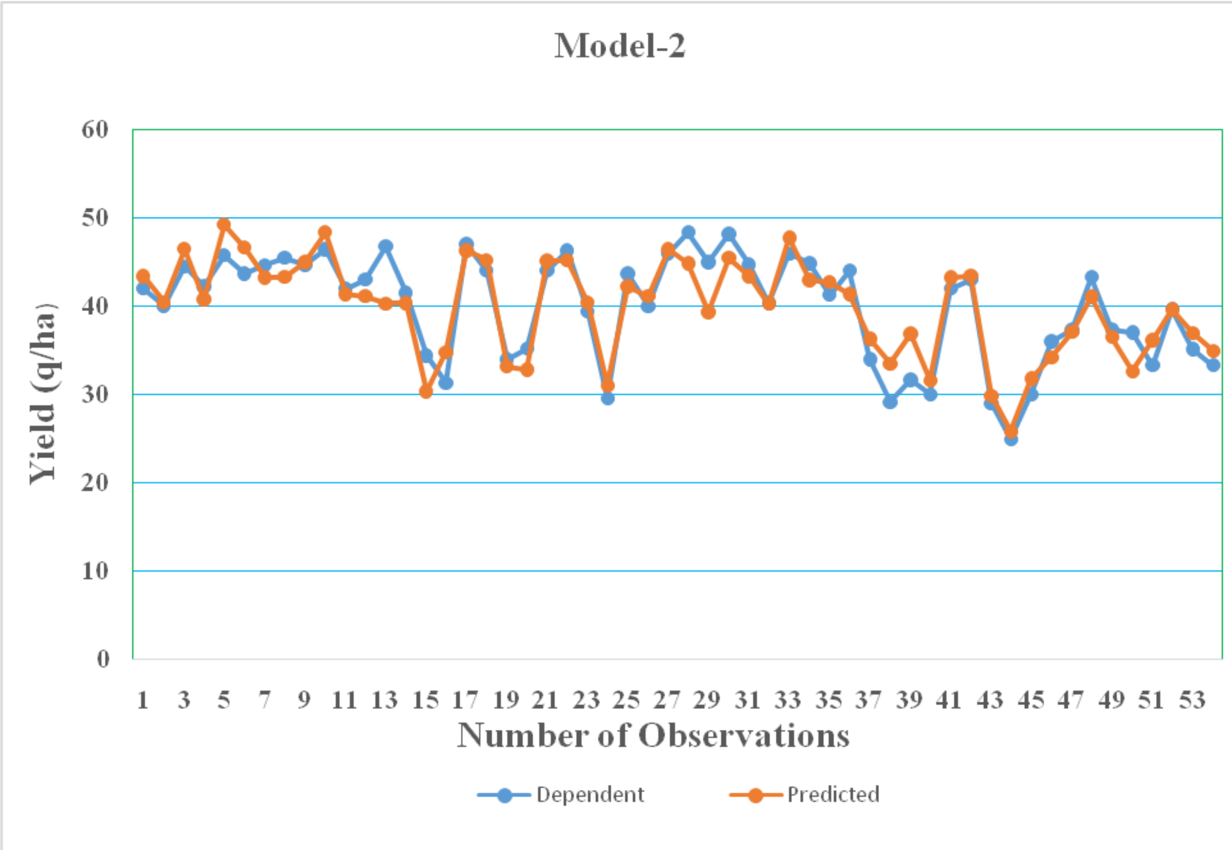


Fig.3 Graph shows the plotting between actual yield and predicted yield



References

Anonymous (2018). Statistical data on area and production of paddy crop in India. F.A.O STAT, *http://fao.org/faostat/en/#data/QC*

Anonymous (2018). Statistical data on area and production of paddy crop during season 2016-17. Directorate of Economics and Statistics, Government of Bihar.

Deep, C. K. Kumar, M. and Kumar, S. (2018). Yield estimation of rice *(Oryzasativa L.*) in Katihar district of Bihar. Advance in Bioresearch, 9 (2), 55-60.

Draper, N. R. and Smith, H. (1966). Application of regression analysis. John Wiley and Sons, New York, 3rd edition, 327-347.

Kumar, M. Singh, M. M. Kumar, S. (2017). Pre-harvest forecasting of rice yield using biometrical characters along with farmer's appraisal in Muzaffarpur district of Bihar. International Journal of Pure & Applied Bioscience, 5 (5), 1553-155.

Nath, B., Singh, S.N. and Rai, G. (2018).Pre-harvest forecast of rice yield for Bhagalpur district in Bihar. Journal of Pharmacognosy and Phytochemistry, 7 (6), 2342-2345.

Pandey, K. K. Rai, V. N. Sisodia. B. V. S. Bharti, A. K. Gairola, K. C. (2013). Pre-harvest forecast models based on weather variables and weather indices for eastern U.P. Advance in Bioresearch, 4 (2), 118-122.

Vogel, F. Bange, G. (1999). Understanding crop statistics. Retrivewed from https: // www.usda.gov/nassinfo/pub 1554.htm.

How to cite this article:

Ravi Ranjan Kumar, S.N. Singh, Kiran Kumari and Bhola Nath. 2019. Yield Estimation of Rice Crop at Pre-Harvest Stage Using Regression Based Statistical Model for Arwal District, Bihar, India. *Int.J.Curr.Microbiol.App.Sci.* 8(08): 2491-2500.
doi: https://doi.org/10.20546/ijcmas.2019.808.290