Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2018; 7(6): 2342-2345
Received: 19-09-2018
Accepted: 21-10-2018

Bhola Nath
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

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

Correspondence
Bhola Nath
Department of Statistics, Mathematics and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India

Pre-harvest forecast of rice yield for Bhagalpur district in Bihar

Bhola Nath, SN Singh and Gaurav Kumar Rai

Abstract
Reliable and timely forecast of crop production are required for various policy decisions related to storage, distribution, pricing, marketing, import-export etc. Pre-harvest forecast of rice yield has a great importance in Bihar as there is much more area and production of paddy in this state [Kumar *et al.* (2013)]. To establish the relationship among Yield (Y), Average Plant Population (X_1), Average Plant Height (X_2), Average number of effective Tillers (X_3), Average Length of Panicle (X_4), Nitrogen (X_5), Phosphorous (X_6), Potash (X_7), Number of Irrigations (X_8), Pest and Disease Infestation (X_9) and Average Plant Condition (X_{10}), a questionnaire based on biometrical character and farmer's appraisal for rice crop was developed. The variable Y was used as dependent and all other X's were as independent. Further, five regression models were selected for minimum RMSE. Regression analysis was performed for each model. All five models were highly significant. Out of five selected regression models, Model-V *i.e.* $\hat{Y} = 16.74955 + 1.55753X_3 - 0.03582X_5 + 0.10947X_6 + 1.22290X_8$ had the minimum Standard Error of Mean Predicted (1.08670) value. Its residuals value was zero and RMSE value was 3.36722. From these analyses it was reflected that Model-V is the best which was used for pre-harvest forecast of rice yield. By using this model pre-harvest forecast of rice yield in Bhagalpur is about 43.80967 (q/ha) for the year 2016-17 based on biometrical characters and farmer's appraisal.

Keywords: Farmers' appraisal, RMSE (root mean square error), Bayesian approach, yield forecast, biometrical characters

Introduction
Rice is the most important cereal crop in India. It occupied 43.855 mha with production of 157.2 Mt and productivity 3.58 t/ha during 2014 (F.A.O statistics, http://faostat3.fao.org/download/Q/QC/E). In Bihar, rice is the most leading food crop which is grown in 3.230 mha with production of 6.802 mt with productivity 2.105 tonnes/ha during the season 2015-16 (Directorate of Economics and Statistics, G.O.B. 2016-17). However, it is easily possible to increase the productivity of rice up to 25.50 q/ha after using available technology and proper demonstration. A study has been carried out to study the feasibility of using farmers' appraisal in the forecast model for sugarcane [Jain *et al.* (1992)] [7]. The results revealed that a reliable forecast could be obtained using plant population and farmers' appraisal.
Another methodology based on farmers' appraisal data has been developed using Bayesian approach. The study has been carried out for wheat in Muzaffarnagar district. Expert opinion data were collected in a number of rounds in a year by interviewing the selected fanners regarding their assessment about the likely crop production and chance of occurrences in various yield classes. From these responses average prior probabilities were computed. Actual harvest yield and farmers' appraisal data on yield for previous year(s) were taken into account to obtain posterior probabilities which were then used for obtaining Bayesian forecast of crop yield for current year (Chandrahas *et al.* 2001) [3]. Forecasting the rice yield for 24 Paragana district of west Bengal has been studied through multiple linear regression analysis (Khan and Chaterjee 1987) [6]. The other approach uses plant characters measured through plant. It can be easily assumed that plants characters are integrated effects of all the factors affecting production.

Materials and Methods
For pre-harvest forecast of rice yield based on biometrical characters and farmers' appraisal for Bhagalpur, district in Bihar, the relationship between crop yield and measurement of plant characteristics were considered as the component of yield, which was recorded much in advance of crop harvest. These relationships were used in making forecast for the next year. This approach had the merits that forecast obtained was objective and a measure of the degree of its production while the traditional approach based on eye observation could not provide. For establishment of the relationship between yield and biometrical characters these steps were undertaken;

(i) Sampling Procedure: In this study 50 samples were collected from five blocks of Bhagalpur district in Bihar. Multistage sampling was adopted and at each stage samples were selected randomly [Chaudhuri and Arnab (1982)] [4]. Data collected from these sampling procedures, 1024 regression models were developed. Multistage sampling method was used to obtain the desire number of samples. At first stage the blocks were selected purposively, at second stage panchayat was selected randomly, at next stage village was also selected randomly. At last stage two plots of each farmer were selected at random by simple random sampling. The purpose of the sampling was to obtain actual data from the farmers' field which was necessary for the establishing the model.

(ii) Identification of qualitative and quantitative characters: The crop yield depends on several plant characters. Some of these characters affect crop yield directly, such as number of plants per plot, number of ear-heads per plant, number of kernels per ear-head and weight per kernel. Other characters like plant height, number of green leaves, Leaf Area Index etc. affect indirectly the yield of crop. The plant characters on which data to be collected for use as explanatory variables in forecast models would be usually searched in consultation with crop physiologists, agronomists and soil scientists for influence of such characters on rice yield. The qualitative and quantitative characters *viz.*, average plant population, average plant height, average number of effective tillers, average length of ear-head, number of green leaves, Nitrogen (N), Phosphorous (P_2O_5), Potash (K_2O), number of irrigations, damages due to pest & disease infestation and average plant condition (Eye estimation) were taken. The list of qualitative and quantitative characters along with their unit of measurement is given in the Table- 1.

Table 1: List of qualitative and quantitative characters

S. N.	Variables	Codes	Unit of Measurement	Type of character
1.	Yield	Y	q/ha	Quantitative
2.	Average Plant Population	X_1	per m^2	”
3.	Average Plant Height	X_2	cm	”
4.	Average No. of effective Tillers	X_3	per m^2	”
5.	Average Length of Panicle	X_4	cm	”
6.	Nitrogen (N)	X_5	kg/ha	”
7.	Phosphorous (P_2O_5)	X_6	kg/ha	”
8.	Potash (K_2O)	X_7	kg/ha	”
9.	No. of Irrigations	X_8	Numbers	”
10.	Pest and Disease Infestation	X_9	Percentage	”
11.	Average Plant Condition	X_{10}	Eye Estimation*	Qualitative
Note: Eye Estimation* denotes the likely condition of the crop judged by engaged farmer.				

(iii) Collection of data and development of regression models: A questionnaire was developed for recording the observations on the farmer's field. The questionnaire intended the informations about name of the district, name of the block, name of the farmer, number of plots, area(ha), variety, number of irrigations, average plant population, average plant height (cm), average number of effective tillers, average length of panicle(cm), damages due to pest & disease infestation (%), average plant condition (Eye Estimation), yield (q/ha) and fertilizers applications (kg/ha) *viz.*, Urea, D.A.P., M.O.P. and Zinc Sulphate ($ZnSO_4$). Development of regression models was carried out based on all possible regression through Stepwise Regression.

(iv) Testing the validity of model using suitable statistical tools: For validity of regression model following assumptions will be required: Relationship between Y and X_i's is linear: there should be a linear relationship between the dependent variable and independent variable. Errors should be distributed with zero mean and constant variance. The error should be uncorrelated: error terms should not be correlated with each other; this type of correlation is called as autocorrelation in notations it can be expressed as $corr(e_i, e_j) = 0$ where, $i \neq j$. Errors are normally distributed: generally the error term associated with the model follows $N \sim (0, \sigma^2)$.

Results and discussion

It has been stated that out of five selected regression models the Model-V having minimum residuals values, CV 9.08366 (%) as well as RMSE 3.36722 and maximum value of R^2 value 0.7024, adj. R^2 0.6726 with least regressors and least forecasting error valued was used for the forecasting the yield of rice during the season 2016-17 for Bhagalpur district in Bihar. The summary of the selected five regression models are as follows:

Table 2: Five better regression models based on minimum RMSE:

Number	Model	R^2	Adj. R^2	C.V. (%)	RMSE
1.	X_3, X_5, X_6, X_8, X_{10}	0.7194	0.6834	8.9330	3.3114
2.	X_3, X_4, X_5, X_6, X_8, X_{10}	0.7246	0.6811	8. 9649	3.3232
3.	X_1, X_3, X_4, X_5, X_6, X_8, X_{10}	0.7283	0.6770	9.0237	3.3450
4.	X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_{10}	0.7339	0.6748	9.0540	3.3563
5.	X_3, X_5, X_6, X_8	0.7024	0.6726	9.0836	3.3672

The regression equation from Model-V for the forecasting of rice yield is given as $\hat{Y} = 16.74955 + 1.55753X_3 - 0.03582X_5 + 0.10947X_6 + 1.22290X_8$. The model has four regressors X_3, X_5, X_6 and X_8 having RMSE value 3.36722, R^2 value 0.7024, adj. R^2 0.6726 and CV 9.08366.

Table 3: Parameter Estimates for Model-V

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	16.74955	3.20292	5.23	0.00**	0
X3	1	1.55753	0.27991	5.56	0.00**	1.23373
X5	1	-0.03582	0.01378	-2.60	0.01**	5.69184
X6	1	0.10947	0.03192	3.43	0.00**	5.80615
X8	1	1.22290	0.43018	2.84	0.00**	1.19545

Table 4: Analysis of variance for Model-V

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1070.44323	267.61081	23.60	0.00**
Error	40	453.52735	11.33818		
Corrected Total	44	1523.97058			

Table 5: The estimates of the dependent values for 45 observations used in the Model-V

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D
1	29.8897	28.8058	1.2378	1.0839	3.131	0.346	0.004
2	33.6200	37.5237	1.5507	-3.9037	2.989	-1.306	0.092
3	30.8757	27.2266	1.2944	3.6491	3.109	1.174	0.048
4	36.7988	30.0222	0.9734	6.7766	3.223	2.102	0.081
5	34.0263	30.6053	0.9057	3.4210	3.243	1.055	0.017
6	38.4371	33.2793	1.3142	5.1578	3.100	1.664	0.099
7	27.7547	32.9877	1.3127	-5.2330	3.101	-1.688	0.102
8	28.2800	30.8418	0.8450	-2.5618	3.259	-0.786	0.008
9	34.0263	33.5226	0.8601	0.5037	3.256	0.155	0.000
10	37.1769	35.5235	0.8466	1.6534	3.259	0.507	0.003
11	39.6974	41.0067	0.9186	-1.3093	3.239	-0.404	0.003
12	40.3275	38.4462	0.9296	1.8813	3.236	0.581	0.006
13	46.0186	41.4751	1.2304	4.5435	3.134	1.450	0.065
14	40.3451	40.1580	0.8905	0.1871	3.247	0.0576	0.000
15	36.5468	35.6712	0.7608	0.8756	3.280	0.267	0.001
16	29.6155	33.9626	0.8603	-4.3471	3.255	-1.335	0.025
17	44.7383	41.3085	0.8773	3.4298	3.251	1.055	0.016
18	46.6287	44.9298	1.2378	1.6989	3.131	0.543	0.009
19	38.4271	36.6526	1.0160	1.7745	3.210	0.553	0.006
20	45.9865	42.2275	1.1032	3.7590	3.181	1.182	0.034
21	39.5368	43.7985	1.0004	-4.2617	3.215	-1.326	0.034
22	37.1867	39.0560	0.9618	-1.8694	3.227	-0.579	0.006
23	39.7078	36.9018	0.6690	2.8060	3.300	0.850	0.006
24	37.1867	36.3970	0.7837	0.7897	3.275	0.241	0.001
25	28.8600	35.1741	0.9116	-6.3141	3.241	-1.948	0.060
26	29.3700	32.2586	0.9595	-2.8886	3.228	-0.895	0.014
27	29.8700	33.1333	0.9234	-3.2633	3.238	-1.008	0.017
28	27.8520	30.9539	0.9711	-3.1019	3.224	-0.962	0.017
29	30.1100	32.1202	0.8807	-2.0102	3.250	-0.619	0.006
30	25.5100	28.9845	1.8315	-3.4745	2.826	-1.230	0.127
31	35.2940	34.9291	1.6640	0.3649	2.927	0.125	0.001
32	34.6600	33.4268	0.9067	1.2332	3.243	0.380	0.002
33	38.4460	34.8846	0.7854	3.5614	3.274	1.088	0.014
34	38.4460	42.0570	1.1871	-3.6110	3.151	-1.146	0.037
35	45.9900	43.1667	1.3408	2.8233	3.089	0.914	0.031
36	44.7380	41.0676	0.9271	3.6704	3.237	1.134	0.021
37	46.6200	46.4291	1.1298	0.1909	3.172	0.0602	0.000
38	44.1082	44.1770	1.0710	-0.0688	3.192	-0.0215	0.000
39	42.2178	44.1204	1.1374	-1.9026	3.169	-0.600	0.009
40	39.6974	40.5217	0.7157	-0.8243	3.290	-0.251	0.001
41	37.2061	39.4686	1.2788	-2.2625	3.115	-0.726	0.018
42	34.6836	39.8825	1.1598	-5.1989	3.161	-1.645	0.073
43	37.7858	37.8415	1.2483	-0.0557	3.127	-0.0178	0.000
44	40.3049	42.2773	1.7313	-1.9724	2.888	-0.683	0.034
45	43.5000	38.9002	1.7606	4.5998	2.870	1.603	0.193

Based on above tabulated values Model-V *i.e.* ($\hat{Y} = 16.74955 + 1.55753X_3 - 0.03582X_5 + 0.10947X_6 + 1.22290X_8$) was found to be the best model and used for forecasting pre-harvest Rice yield. Thus, Model-V will be used for the forecasting the preharvest yield of the paddy for the Bhagalpur district of Bihar. Forecasting errors for the observations which were not included in the model development for Bhagalpur district.

Model Validation

In the proposed Model-V *i.e.* ($\hat{Y} = 16.74955 + 1.55753X_3 - 0.03582X_5 + 0.10947X_6 + 1.22290X_8$), for forecasting the yield there were four repressors *viz.*, X_3 was Average number of tillers (per m^2), X_5 Nitrogen (Kg/ha), X_6 Phosphorous (Kg/ha), X_8 was number of irrigations. There were ten per cent observations kept for model validation. Using this set of data residual analysis was performed and the result was as follow:

Table 6: Residual Analysis with kept ten per cent observations

S. N.	X_3	X_5	X_6	X_8	Y	$\hat{Y}$	$\hat{e}_i = Y - \hat{Y}$	MAPE = $\frac{\hat{e}_i}{\hat{Y}} \times 100$
1.	8.80	213.22	129.91	2	40.98	39.48508	1.49	3.786033
2.	11.20	89.55	54.56	2	47.26	39.40474	7.85	19.93178
3.	14.40	89.55	54.56	2	47.89	44.38883	3.50	7.885019
4.	15.00	81.97	54.56	4	46.01	48.04059	-2.03	4.225574
5.	14.80	81.97	54.56	4	47.90	47.72908	0.17	0.361028

Where, Y was Yield (q/ha), e_i was the i^{th} residual, MAPE denotes Mean Absolute Percentage Error and the value was 7.2378 for the Model-V.

For Model-V, the constant value (intercept) was 16.74955, that showed when only seed was applied in the field fitting all the values of four regressors set to zero, the minimum yield would be 16.74955 q/ha. The coefficients of three regressors; X_3, X_6 and X_8 were positive, that meant as the average number of tillers (X_3) increased by one unit the dependent variable (Y) would also be increased by 1.55753 per cent. Phosphorous (X_6) had positive relation with the yield, which meant that when the dose of phosphorus was increased, the increment in yield would be by 0.10947 per cent. The model elaborated, when the number of irrigations was increased by one unit, the yield was definitely increased by 0.10947 per cent. The regressor X_5 (Nitrogen) had negative relationship with the yield of paddy, when Nitrogen level was changed by one unit the yield was decreased by 0.03582 per cent which indicated the over use of nitrogenous fertilizer. Farmers of that localities are generally used the excess dose of nitrogenous fertilizers that affects their economic conditions.

Model-V *i.e.* $\hat{Y} = 16.74955 + 1.55753X_3 - 0.03582X_5 + 0.10947X_6 + 1.22290X_8$ had the minimum Standard Error of Mean Predicted (1.08670) value. Its residuals value was zero and RMSE value was 3.36722. From these analyses it was reflected that Model-V was the best and was used for pre-harvest forecast of rice yield. By using this model pre-harvest forecast of rice yield in Bhagalpur is about 43.80967 (q/ha) for the year 2016-17 based on biometrical characters and farmer's appraisal.

References

1. Anonymous. Statistical data on area and production of different crops in India. F.A.O statistics, 2017. http://faostat3.fao.org/download/Q/QC/E.
2. Anonymous. Statistical data on area and production of paddy crop during the season 2015-16. Directorate of Statistics and Evaluation; Government of Bihar, 2017.
3. Chandrahas, Rai T. Pilot study for developing Bayesian probability forecast model based on farmer's appraisal data on wheat crop. Project report IASRI, New Delhi, 2001.
4. Chaudhuri Arijit, Arnab Raghunath. On unbiased variance-estimation with various multi-stage sampling strategies. The Indian journal of statistics. 1982; 42(1):92-101.
5. Kumar Mahesh, Singh MM, Singh SP, Bharat RC, Sanjeev Kumar, Reddy Ravindra B. Pre-harvest Forecasting of Rice Yield using Biometrical Characters along with Farmers Appraisal in Samastipur District of Bihar. R.A.U. Journal of Research. 2013; 23(1-2):41-45.
6. Khan SA, Chatterjee BN. Crop weather analysis for forecasting the yield of autumn rice in West-Bengal. Indian journal of agricultural sciences. 1987; 57(11):791-794.
7. Jain RC, Agrawal R, Singh KN. A within year growth model for crop yield forecasting. Biometrical Journal. 1992; 34(7):789-799.