

Statistical Modeling and Significance of Inputs Studies for the Rice Crop Productivity in Pakistan

Muhammad Islam^{*1} and Rabia Siddiqui³

¹Department of Statistics, The Islamia University of Bahawalpur, Pakistan

²Crop Reporting Service Agriculture Departments Punjab

*Corresponding Author: mislam6667@gmail.com

Abstract

Lower production of rice crop and increasing population trend is a tangible question for Pakistan. The current study based cross-sectional data collected from crop reporting service Punjab. The multiple linear regression (MLR) model is applied to investigate the significant factor for rice crop yield enhancement. The inputs variable i.e. owned land, seed rate, DAP, Urea, no. of water, no. of ploughs, no. of levels, crop life periods days, other fertilizers, variety super yes or no, seed type yes or no, spray no or yes and disease attack yes or no are studied in MLR model for rice productivity. All the factors found to be statistical significant except land, level, crop life period and super variety. Adj R^2 is found to be 0.422 and it is good fit for cross-sectional data.

Keywords: Factors Significant, Multiple Linear Regression Model, Rice Crop.

Introduction

Wheat, rice, maize, jowar, bajra and barley are the food crops of Pakistan (PBS, ESP 2006-07). Table 1 depicts that the yield of rice is low about -52.50, -32.76, -16.31, -50.76, -51.96, -13.98, -59.57, -7.75 as compared to competitive countries like Korea, Indonesia, Bangladesh, China, Japan, Sri Lanka, USA and Malaysia (Badar, 2007). The population growth rate in Pakistan is about 2.05% and it is reported high in Asian countries. Low production and increasing population trend is a tangible question for Pakistan (Taye, 2014; Sohail et al. 2021).

Table 1: Comparison of rice crop yield with other countries

Countries	Avg yield kg/hect of rice crop (2002-03 to 2006-07)	% dec
Korea	6438	-52.50
Indonesia	4548	-32.76
Bangladesh	3654	-16.31
China	6211	-50.76
Japan	6366	-51.96
Sri Lanka	3555	-13.98
USA	7563	-59.57
Pakistan	3058	0.00

Malaysia	3315	-7.75
----------	------	-------

Shah et al. (2014) statistically analyzed the agricultural crops by two stage systematic sampling scheme for district Bahawalpur on the data taken from CRS. They reported that lack of govt. cost and labor policies have decreased the yield of rice crop in district Bahawalpur. Singh et al. (2014) stated that Chhattisgarh region of India and applied different regression equations i.e. linear, logarithmic, inverse, quadratic etc. on paddy crop yield and reported that linear and logarithmic regression model found to be better fitted model. Kaloo et al. (2014) used simple linear regression to study rice crop and reported coefficient of determination 34.88% and concluded that production of rice depends on time.

Problem Statement

In Pakistan average yield rice crop is low as compared to other countries while the population is increasing across the year. The government is taking several steps to cope with the population explosion. It is reported that the population of Pakistan is 188.02 million in 2014 (ESP 2013-14) and it will touch 363 million in 2050 (world population data sheet 2013, ESP 2013-14). Crop production is a function of different physical and non-physical factors (Taye, 2014, Sohail et al., 2017). Physical factors means the factors which are controlled by nature and non-physical factors means factors which are controlled by human beings such as various inputs of fertilizers, irrigation, varieties, seed rate, seed treatment and pesticides etc.

Objectives of the Study

It is necessary to statistically analyze and build the regression model to investigate the significant of the different inputs to get the better production of rice crop to meet the challenge of food availability in Pakistan.

Material and Methods

Data collection

Crop reporting service (CRS) is govt. organization and is a wing of agriculture department of Punjab. The domain of the department is forecasting of area, yield and production of all major and minor crops in Punjab. Secondary data is collected from CRS district Bahawalpur for 2007-08 to 2014-15. About 423 fields and 10998 entries for rice has been utilized and investigated in SPSS.

Regression Model Analysis

Multiple linear regression (MLR) are also applied to investigate the significant on rice crop productivity

$$Y_i = \beta_0 + \beta_1 L + \beta_2 SQ + \beta_3 DAP + \beta_4 Urea + \beta_5 W + \beta_6 P + \beta_7 LV + \beta_8 CLP + \beta_9 OFR + \beta_{10} VARI + \beta_{11} ST + \beta_{12} SP + \beta_{13} DES + e_i \quad (1)$$

where β_0 is intercept, β_i are the slope of regression and e_i is error term.

Y_i = Avg. yield in mds/acre rice crop, L = land in acres, SQ = Seed quantity, DAP = DAP kg/acre, $Urea$ = Urea kg/acre, W = no. of water, P = No. of plough, LV = No. of levels, CLP =

Crop life periods, OFR= other fertilizers, VARI=Variety super, ST= Seed type, SP= Spray, DES= Disease attack.

Diagnosis Regression

Normality and Linearity is checked graphically by P-P plots and histogram with normal curve. Autocorrelation tested by Durban Watson test and multicollinearity tested by the variation inflation factor (VIF), (Montgomery, 2003).

$$VIF = \frac{1}{1 - R^2_i} \quad (2)$$

VIF < 10 indicates absence of multicollinearity. F-test: is used to test the overall significance of the model.

$$F = \frac{SSR_{eg.} / k - 1}{SSR_{es} / n - k} \quad (3)$$

where ‘k’ is no of independent variables and ‘n’ is no of observations. t-test is used to check the significance of coefficients

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \quad (4)$$

The R^2 is used to check the explained variation in model

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad (5)$$

Data Analysis

For most of the statistical analysis normality in the data is pre-requisite for (Qayyum et al. 2012). Figures 1 and 2 shows the normality by histogram and P-P plots of district Bahawalpur and shows complete linear and normal behavior.

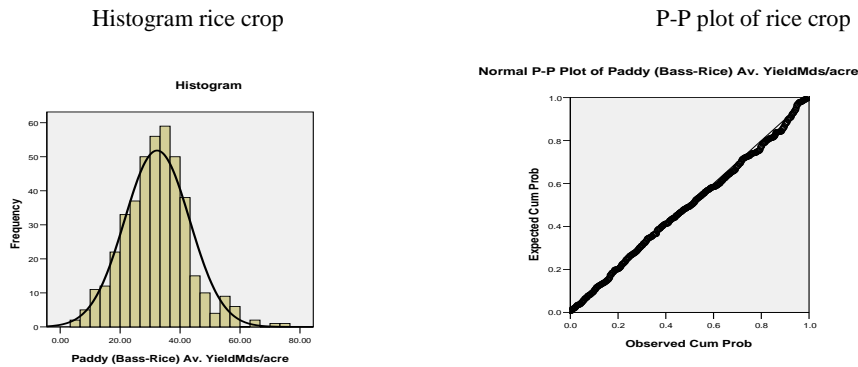


Figure 1: Histogram and P-plot of rice crpe

Multiple Linear Regression Model for Rice Productivity

Effects of all the factors investigated by the multiple linear regression. Table 2, value of R^2 and Adj R^2 shows that model is good fit in view of cross-sectional datasets and 42.2% variation is explained by the model. Mean square error (MSE) reported as 68.059. The Value of the Durbin-Watson test is 1.096 indicate there is no autocorrelation in model.

Table 2: Overall model summary for rice crop

Model	Sum of Squares	df	MSE	F	Sig	R^2	Adj R^2	Durbin Watson
Regression	21857.12	13	1681.317	24.70	.000	0.440	0.422	1.096
Residual	27836.17	409	68.059	-----	-----	-----	-----	-----
Total	49693.3	422	-----	-----	-----	-----	-----	-----

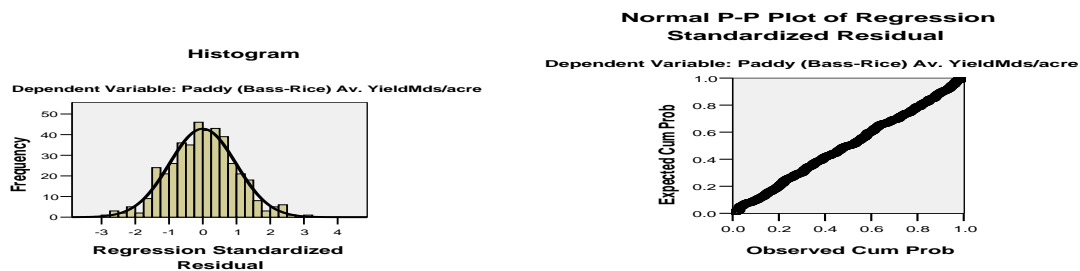
Significance of Regression Coefficients

Table 3, determined that the own land coefficient is -0.021 which implies that a unit increase in the owned land in acre will decrease the rice production 0.021% and is statistically insignificant. The reason is that rice crop is sown in Bahawalpur district on small scale and mostly rice crop farmers are small farmers. Seed quantity coefficient is -0.653 which shows that a unit increase in recommended seed rate per acre will decrease the production about 0.653 percent and it is significant. Coefficient of the DAP and urea are 0.069 and 0.053 which means increase in DAP and urea will enhance the rice production about 0.069% and 0.053% and these coefficients are statistically significant. The model depict that the coefficient of other fertilizers is 6.86, indicate increases in the other fertilizers will enhance the rice production up to 6.86% and is statistically significant, water for the rice crop are essential for yield maximization and its coefficient is 0.371 means one percent increase in the no. of water will maximize the yield 0.371% and it is statistically significant indicating no. of water is affecting the rice production. The coefficient of Plough is 1.795 which means increase in the no. of plough will increase the rice crop production up to 1.795% and it is found to be statistically significant. Level coefficient value is 0.129 indicate that one more use of no. of level will rise the rice production up to 0.129% and is statistically insignificant. For the crop life time from sowing to harvesting, coefficient is -0.001 indicating that one day late sowing leads to decrease the rice production up to 0.001% and it is statistically insignificant. For basmati super coefficient value is found to be -0.821 which means one percent increase super basmati will decrease rice production 0.821% and is statistically insignificant which indicates to shift the area toward the advanced rice varieties for better productivity. There are two types of seed used home seed or certified and its coefficient is 6.837 indicating that one percent increase in the certified seed will increase the rice production 6.837% and is statistically significant. Spray coefficient value is 5.77 indicating that one percent increases in spray will increase the rice production 5.77% and is statistically significant which shows that it is affecting factor for rice production. Disease attack coefficient value is -7.398 indicating that one percent increase the share of disease crop will decrease the rice production about 7.398% and it is statistically significant which shows that it is affecting factors for rice crop.

Table 3: Regression model analysis for rice crop district Bahawalpur

Model MLR	β_i	T	Sig	Tolerance	VIF
Constant	17.463	3.724	0.000		
Owned Land	-0.021	-0.646	0.519	0.966	1.035
Seed rate	-0.653	-2.705	0.007	0.896	1.116
DAP	0.069	3.987	0.000	0.856	1.168
Urea	0.053	3.369	0.001	0.859	1.164
No. of water	0.371	2.574	0.010	0.814	1.229
No. of ploughs	1.795	4.026	0.000	0.865	1.156
No. of levels	0.129	0.215	0.830	0.97	1.031
Crop life periods days	-0.001	-0.019	0.985	0.88	1.136
Other fertilizers	6.86	5.742	0.000	0.889	1.125
Variety Super yes or no	-0.821	-0.951	0.342	0.946	1.057
Seed Type yes or no	6.837	5.826	0.000	0.898	1.113
Spray no or yes	5.77	5.815	0.000	0.654	1.529
disease attack yes or no	-7.398	-7.203	0.000	0.705	1.419

Figures 2, histogram and P-P plot indicates that error term follows normality and linearity. Durban-Watson test is less than 2 indicate there are no autocorrelation and model is good fit. All values of VIF and tolerance supporting that there is no multicollinearity.



Conclusions

From multiple linear regression model it is found that coefficients of owned land, seed rate, DAP, Urea, no. of water, no. of ploughs, no. of levels, crop life periods days, other fertilizers, variety Super yes or no, seed type yes or no, spray no or yes and disease attack yes or no are reported as -0.021, -0.653, 0.069, 0.053, 0.371, 1.795, 0.129, -0.001, 6.86, -0.821, 6.837, 5.77 and -7.398 respectively. All the factors found to be statistical significant except land, level, crop life period and super variety. Adj R^2 is found to be 0.422 and it is good fit for cross-sectional data.

Acknowledgement

We are thankful to the editor and reviewers for their positive direction which improves our work.

References

- Badar, H., Ghafoor, A and Adil, S.A. (2007). Factors affecting agricultural production of Punjab (Pakistan). Pak. J. Agri. Sci., Vol. 44(3).
- Biswas, R., Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in West Bengal. Journal of Crop and Weed, 9(2):26-31.
- Crops Area and Production (By Districts), 1981-82 to 2008-09, Volume I, Food and Cash Crops. Government of Pakistan, Statistics Division Federal Bureau of Statistics (Economic Wing) Islamabad.
- Economic surveys of Pakistan 2006-07 to 2013-14.
- Kaloo, M.J., Patidar, R and Choure, T. (2014). Production and productivity of rice in Jammu and Kashmir: An Economic Analysis. International Journal of Research (IJR) Vol-1, Issue-4.
- Montgomery D.C., (2003). Introduction to linear regression analysis, 3rd ed. Wiley Publishing, New York.
- Punjab Agricultural Statistics 2003-04 to 2007-08., 2008. Directorate of Agriculture, Crop Reporting Service, Punjab.
- Qayyum, A., 2012. Model based wheat yield estimation in the Punjab, Pakistan. PhD. Thesis Department of Statistics G.C University Lahore.
- Sohail, M. U., Shabbir, J., Sohail, F., and Kadilar, C. (2021). An introduction to sensible constraints for the imputation of missing values. *Journal of Statistics and Management Systems*, 1-29.
- Sohail, M. U., Shabbir, J., and Ahmed, S. (2017). Modified Class of Ratio and Regression Type Estimators For Imputing Scrambling Response. Pak. J. Statist, 33(4), 277-300.
- Shah, M. A. A., Yasmeen., A and Qasim, H., 2014. Statistical analysis of agriculture crop data by two stage systematic sampling. J. Agric. and Environ. Sci., 14 (10): 1026-1029.
- Singh, D.P., Kumar, D., Paikra, M.S and kusro, P.S. (2014). Developing Statistical models to study the growth rates of Paddy crops in different districts of Chhattisgarh. American International Journal of Research in Formal, Applied and Natural Sciences, 5(1), pp. 102-104.
- Taye, D. (2014). Statistical analysis of factors affecting the production of wheat crop in west Shoa Zone; A case study at ambo worda. Thesis for bachelor degree in Statistics. Department of Statistics, Ambo University.