Pakistan Journal of Applied Sciences 3 (4): 236-239, 2003
ISSN 1607-8926

All Possible Regression Study of Wheat Crop

Yousaf Hayat, [1]Syed Muhammad Asim, [1]Qamaruz-Zaman and [2]Naeem Khan
Department of Maths/Stat and Computer Science, NWFP, Agricultural University Peshawar, Pakistan
[1]Department of Statistics University of Peshawar, Pakistan
[2]Department of Weed Science, NWFP, Agricultural University Peshawar, Pakistan

ABSTRACT

The wheat crop is more important than other crops, where the year 1998 was declared as wheat year (Economic Survey, 1997-98). In this paper an efforts has been made to highlight necessary variables for increasing the production (grain yield) of wheat. The best subset model was identified with "all possible regression" approach based on selection criteria R^2-adjusted. For analysis, data was taken from the experiment conducted at Malakandher farm NWFP Agricultural University Peshawar during the Rabi season 2000-01, to assess the contribution of different agronomic variables on grain yield of wheat (Ghaznavi-98). The best subset model selected was the model having spikelets spike^{-1} as predictor because it contributes 82.2% of the variation in the grain yield. The plant heights showed zero% variation in the grain yield.

Key words: All possible regression, R^2- adjusted and F-test

INTRODUCTION

Wheat is used as a major food source all over the world. It is the staple food of Pakistan and meets the major dietary requirements. The cultivation of wheat seed is simple and adaptable to varied soil and climatic conditions. It is also known as the "King of cereals". Besides food, wheat is also used for livestock and poultry feed. A large population of the world consumes wheat in a number of ways. Wheat supplies about 73% of the calories and protein of the average diet (Heyne, 1987).

The term regression is used to establishing the actual relationship between two or more variables. But scientific, social, economic and agricultural phenomena do not confine to two variables. A large number of studies involve two or more than two variables. In these studies we often need to give actual relationship between two or more variables (Agarwal, 1991). For this purpose we choose the method of all possible regressions. This technique requires that investigator fit all the subset regression models involving one predictor variable, two predictor variables and so on. Each subset regression model was then evaluated according to some suitable criterion like R^2, R^2- adjusted and Mallow's Cp statistic and the best subset regression model was selected (Draper and Smith, 1981).

In the given study we used the criteria of R^2-adjusted. It is the rescaling of "R^2" by degrees of freedom so that it involves a ratio of mean squares rather than sum-of-squares and is given by the relation: R^2-adjusted = 1-MSE/MS (total).

MATERIALS AND METHODS

An experiment was conducted at Malakandher farm NWFP Agricultural University Peshawar during the Rabi season 2000-01, to assess the effect of different agronomic variables on grain yield of wheat crop (Ghaznavi-98). The study includes six variables (Weed Density m^{-2}, 1000-grain weight (g), spikelets spike^{-1}, tillers plant^{-1}, plant height (cm) and grain yield (kg ha^{-1}). The average data were collected from 21 plots of equal size of wheat variety on different agronomic variables. The data were then analyzed by fitting all possible subset regression models, to find out the best subset regression model(s) in one variable, two variables, three variables and four variables, which explained most of the variation in response variable. The criterion used was R^2-adjusted to select the best subset regression model(s).

For the purpose of simplicity the variables were presented as:

Y = Grain yield (Kg ha^{-1}), called the response variable
X_1 = Weed density m^{-2}
X_2 = 1000 grain weight (g)
X_3 = Spikelets $spike^{-1}$
X_4 = Tillers $plant^{-1}$
X_5 = Plant height (cm) at maturity
Where X_1, X_2, X_3, X_4 and X_5 were called predictors (independent variables).

RESULTS AND DISCUSSION

First of all, all possible subset regression models having only one predictor were fitted. The results showed that 82.2% of the variation in grain yield (response) was explained by the linear combination of spikelets $spike^{-1}$ and the F- test for regression co-efficient was highly significant (P-value = 0). While plant height showed zero% variation in the response. The F-tests (for regression co-efficient) were highly significant for 1000 grain weight, spikelets $spike^{-1}$ and that for tillers $plant^{-1}$, the F-tests for weed density m^{-2} and plant height were insignificant and showed 7.9 and zero% variation in the grain yield, respectively.

Table 1: All possible subset regression models were having only one predictor

Predictor	Regression coefficient	R^2	R^2-adjusted	F-ratio	P-value
X_1	55.18	0.125	0.079	2.730^{Ns}	0.115
X_2	366.09	0.828	0.819	91.53*	0.000
X_3	413.23	0.831	0.822	93.61*	0.000
X_4	418.87	0.821	0.812	87.11*	0.000
X_5	0.444	0.000	-0.053	0.000^{Ns}	0.998

Ns = Non-significant at 5% level of significance ** = Significant at 5% level of significance

Table 2: All possible subset regression models were having two predictors

	Regression coefficients					
Predictors	1	2	R^2	R^2-adjusted	F-ratio	P-value
X_1,X_2	18.461	353.29	0.840	0.823	47.65*	0.000
X_1,X_3	-37.23	474.59	0.870	0.856	60.26*	0.000
X_1,X_4	11.25	408.20	0.826	0.806	42.61*	0.000
X_1,X_5	60.55	71.981	0.138	0.042	1.44^{Ns}	0.263
X_2,X_3	196.75	226.65	0.901	0.890	81.92*	0.000
X_2,X_4	201.07	216.25	0.879	0.865	65.17*	0.000
X_2,X_5	367.05	-28.77	0.830	0.811	44.02*	0.000
X_3,X_4	230.69	216.57	0.889	0.876	71.73*	0.000
X_3,X_5	413.46	13.72	0.832	0.813	44.50*	0.000
X_4,X_5	442.01	-135.41	0.866	0.851	58.24*	0.000

Ns = Non-significant at 5% level of significance * = Significant at 5% level of significance

Table 3: All possible subset regression models were having three predictors

	Regression Coefficients						
Predictors	1	2	3	R^2	R^2-adjusted	F-ratio	P-value
X_1,X_2,X_3	17.74	164.01	286.93	0.908	0.892	55.85*	0.000
X_1,X_2,X_4	12.27	202.67	203.02	0.884	0.864	43.28*	0.000
X_1,X_2,X_5	17.91	353.89	-6.565	0.841	0.813	30.03*	0.000
X_1,X_3,X_4	-22.83	305.04	173.01	0.901	0.883	51.45*	0.000
X_1,X_3,X_5	-40.60	479.60	-32.12	0.872	0.850	38.75*	0.000
X_1,X_4,X_5	-3.302	446.02	-140.54	0.866	0.843	36.78*	0.000
X_2,X_3,X_4	141.34	177.09	121.15	0.913	0.898	59.76*	0.000
X_2,X_3,X_5	198.04	225.29	-8.083	0.901	0.884	51.67*	0.000
X_2,X_4,X_5	160.64	273.47	-96.40	0.899	0.881	50.57*	0.000
X_3,X_4,X_5	183.00	271.58	-77.15	0.900	0.883	51.17*	0.000

* = Significant at 5% level of significance

Table 4: All possible subset regression models were having four predictors

	Regression coefficient							
Predictors	1	2	3	4	R^2	R^2-adjusted	F-ratio	P-value
X_{1234}	-13.74	121.98	229.18	108.00	0.917	0.897	44.41*	0.000
X_{1235}	-20.79	162.78	292.66	-27.67	0.910	0.887	40.26*	0.000
X_{1345}	-27.99	263.00	231.04	-95.19	0.918	0.898	44.84*	0.000
X_{1245}	2.554	162.86	268.04	-91.88	0.899	0.874	35.78*	0.000
X_{2345}	127.36	146.67	171.78	-57.78	0.920	0.900	45.85*	0.000

* = Significant at 5% level of significance

The estimated multiple linear regression equation including all five variables is:

$$Y = 2379.09 - 19.83X_1 + 95.15X_2 + 212.54X_3 + 168.30X_4 - 75.46X_5$$

Table 5: Analysis of variance table

Source of variation	DF	Sum of squares	Mean squares	F-ratio	P-value
Regression	5	9752008.15	1950401.63	38.35*	0.00
Residuals	15	76284.84	50856.12		
Total	20	10514849.99			

* = Significant at 5% level of significance

Secondly all possible subset regression models having two predictors were fitted and we found that the model having 1000-grain weight and spikelets spike^{-1} as independent variables explained larger variation as compare to other subset regression models, which was 89%. The subset regression model having weed density m^{-2} and plant height as predictors showed 4.2% of the variation in response which was very small as compare to other subset models. Also the F-test for overall significance of regression coefficients of this model was insignificant, while highly significant for other subset regression models, at 5% as well as 1% level of significance.

The subset regression model having three predictors, 1000-grain weight, spikelets spike^{-1} and tillers plant^{-1} contributing 89.9% of the variation in response which is greater as compare to other subset regression models of three predictors. The lowest variation in response was 81.3% that was due to weed density m^{-2}, 1000-grain weight and plant height used as regressors in the subset model. All the F-tests for overall significance of regression coefficients were highly significant at both 5 and 1% level of significance.

At last all possible regression models having four regressors were fitted, in which the best subset regression model choosed, was the model having 1000-grain weight, spikelets spike^{-1}, tillers plant^{-1} and plant height as predictors. Because it explained 90% variation in the grain yield. The lowest variation showed by the subset regression model having weed density m^{-2}, 1000-grain weights, tiller plant^{-1} and plant height as independent variables that was 89.9%. All the F-tests for over all significance of regression coefficients were not accepted at both 5 and 1% level of significance.

The variation in response explained by all the predictors was 92.7% and the F-test was highly significant for over all significance of regression coefficients at both 5 and 1% level of significance.

The following conclusions can be made from the results obtained from the studies.

A wheat variety produced more spikelets spike^{-1} will be used for cultivation, because it is the only variable whose contribution in the grain yield is greater as compare to other variables.

The plant height (cm) explained zero% variation in response (Grain yield). This indicates that grain yield is greatly affected by the plant height, i.e. the plant height and grain yield are inversely proportional to each other. This happens because of lodging of plants as a result the grain yield is reduced.

1000-grain weight, spikelets spike^{-1} and tillers plant^{-1} explained 89.8% variation in the grain yield which was high as compared to other subset regression models. It means that as the number of tillers plant^{-1} increases, increase occurs also in spikelets spike^{-1} as a result the grain yield is maximized.

Weed density m^{-2} has no contribution in the grain yield.

REFERENCES

Agarwal, B.L., 1991. Basic Statistics, Second edition, Willey Eastern Limited, New Delhi.

Ahuja, K.N. and R.K. Sharma, 1988. Multiple Correlation and Regression studies in Pigeonpea under irrigated conditions, Haryana J. Agric., 4: 56-57.

Draper, N. and H. Smith, 1981. Applied Regression Analysis, John Wiley and Sons, New York.

Gomez, K.A. and A.A. Gomez, 1984. Statistical Procedures for Agriculture Research, 2nd Edition.

Government of Pakistan, 1998. Pakistan Economic Survey 1997-98. Finance Division, Islamabad, Pakistan.

Heyne, E.G., 1987. Wheat and Wheat Improvement. Second edition. Madison Wisconsin, USA.

Montgomery, D.C. and E.A. Peek, 1982. Introduction to Linear Regression Analysis John Willey and Sons; Inc. New York.