WALIA journal 31(S3): 33-xx, 2015
Available online at www.Waliaj.com
ISSN 1026-3861

Wheat harvest losses forecast and data analysis using dummy variables model

Amin Reza Jamshidi *, Ali Afrous

Department of Engineering, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran

Abstract: The increase in production and cut the dependence on imports of wheat caused authorities to pay special attention to production systems and strategies to increase production and reduce their losses. According to figures released in Iran, planted area is 2.61 million hectares and the production rate is 12.57 million tones with an average yield of 2.8 tones in irrigated agriculture. A significant portion of these products by the combine at harvest time spread on the ground and die. Apart from setting different parts of combine, combine type in this have significant role and with respect to time and harvested area, yield, farm, life combine, crop varieties and cultivation (irrigated and rain-fed), these values are different in a variety of combines. In this study we examined two types of combines John Deere 955 class and Sahand 68S w. based on the cited factors data from different parts under wheat harvesting of Khozestan was studied. Finally, using dummy variables model a plan for predicting precipitation and data analysis was presented.

Key words: *Dummy variables model; Losses prediction; Harvest losses; Mechanization*

1. Introduction

Our country area is 165 million hectares of which 18 million hectares is in the agricultural production cycle to provide the need for approximately 75 % of the population. Wheat in the middle, as the most important product, in terms of both food and independence policy of agricultural, is an important issue. One of the issues discussed in this field in recent years has been the issue of crop losses from production to consumption and a way to prevent it, especially in the harvesting by grain harvesting machinery (combine). The emphasis on this issue is consistent with the emphasis on increased production of wheat. These losses that were observed at harvest time were even 20% (Bukhari et al., 1983). On the other hand, while some countries have yield 9 tons per hectare, the average yield of irrigated and rainfed wheat in 1372, has been announced 1750. This statistics for the years 1380 and up until then increased and in year 1383 have been reached to average 3800 (Dillon et al., 1993).

Therefore, regarding the strategy of the Ministry of Agriculture and also the centrality of this product and highly subside allocated by the government for this purpose and culture of Wheat respect among people, the need to scientifically following in the field of to these losses reduction is felt.

However, post-harvest loss to the place of consumption itself has its own discussion and only in this paper only harvest loss due to natural mortality has been paid.

Different factors such as combine setting, lack of timely harvest, the combine type, seed type... are effective on the loss of wheat at its harvest time. Losses during harvest are 4 to 5 percent in the industrialized countries (Evatt, 1991).

In recent years, measures have been initiated in this regard. Including the 1372 cereal harvest waste reduction plan by former Ministry of Agriculture of Agriculture Deputy in cooperation with the Agricultural Machinery Development Board has been implemented and now is the most important issues at harvest time (Lockie, 1997).

Also the reduction of wheat losses from production to Consumption in the form of National specific projects of country's Scientific Research Council is running.

Harvest losses forecast and data analysis using dummy variables model was the aim of this article. In this context, precise understanding of the factors influencing mortality in determining the loss, will help a lot to the proposed model.

2. Materials and methods

Dummy variables model is a kind of regression model by which we can assess the impact of qualitative variables on the dependent variable. These models can be only formed of dummy variables (qualitative) in which case it is called the variance analysis model and or can be a combination of dummy and quantitative variables in which case it is called analysis of covariance model.

In this model, all variables by means of suitable quality should be converted to zero and one. To estimate the model in this study a linear function in the form below is used.

* Corresponding Author.

$$y = \alpha + \sum_{i=1}^{n} \beta_i X_i$$

Where Xi represents the independent variables and y shows the dependent variable and the values of α and βi are the coefficients indicate that we are trying to estimate. This model can be easily estimated using OLS. But the ways of interpreting coefficients of the dummy variables will be different

Independent variables are entered in the model to estimate the relationship between the influx of regional varieties, combine model, the culture, the life of the combine harvester, harvesting date and time.

The dependent variable as the percent of loss is the sum of harvesting and product natural loss. Variable region contains four Ahvaz, Shoushtaer, Andimeshk, Dezful in which the city of Ahvaz was selected as the baseline in the region and other regions variables, respectively, Shoushtar, Dezful and Andimeshk were entered into the model. In this case, if the amount of all three areas is equal to zero represents the city of Ahvaz, if Shoushtar variable is equal to one and the other variables are zero indicates Shoushtaer and if the variable Dezful is equal to one and the other variables are zero represents the city dezful and finally, if Andimeshk variable is equal to one and other variables equal to zero indicates that the city will be the Andimeshk city.

In simple terms, for each city a qualitative variable is defined that value one, indicating the removal of the city and zero represents the harvest in another city. As mentioned earlier, if all three variables are equal to zero indicates the based city or Ahvaz.

Similarly for four cultivars of wheat varieties that plateau, bright, Gaskvzhn and SARDARI the shelf varieties selected as basic varieties and cultivars bright, Gaskvzhn and SARDARI by qualitative variables Roshan, Sardary and Gaskojen is entered to the model. Combine models including the John Deere and Sahand S68 is entered to the model by a qualitative variable that is called Model.

A value of one indicates Deere combine and zero shows the S68 combine harvester. Cultures containing both irrigated and rainfed conditions that are specified by the variable Abideym. The value one represents irrigated agriculture and a value of zero indicates rain fed.

Thus the qualitative variables including regional varieties , crop type and combines models entered to the model, other variables including age , date of harvest and harvest time are quantitative variables that are introduced with the name of Age, time, Time.

Age varies on the year, Days variable on the number of days since the first of 84 and time variable on the time interval of 12 hours per night, which is a value between zero and one, are considered. Thus, the overall shape of the model is as follows:

Eq.1

$$y = \alpha + \beta_{Taybad}.Taybad + \beta_{Khaf}.Khaf + \beta_{Torbatjam}.Torbat\,jam + \beta_{Roshan}.Roshan + \beta_{Sradary}.Sardary + \beta_{Gaskojen}.Gaskojen + \beta_{Abideym}.Abideym + \beta_{Model}.Model + \beta_{Omr}.Omr + \beta_{Tarikh}.Tarikh + \beta_{Saat}.Saat$$

The overall significance of the regression using the F-test and significance of coefficients using t-test are investigated.

3. Results

Referred model in former section on 81 observations of the four districts Khuzestan, shoushtar, andimeshk and dezful was studied. The model estimation is done through software Spss11. Table 1 showing the coefficients of model variables and values of the t-statistic that is derived from the output of the software. According to this, the loss forecast function using the table below is derived as follows:

F -statistic is 8.92 times that is significant of the very high confidence level (over 99%), therefore, the overall significance of the regression equation is approved. In the following, to investigate the significance of each of the variables t-test is used. The results show that the coefficients of the variables Andimeshk, Abideym, Age and Shoushtar accordance with Table 1 , respectively, have the highest level (above 90%) and coefficients of variables Time, Days and Amalkard have the lowest of significance level and is not statistically difference from zero . Significance level of other variables can also be easily derived using Table 2 . the amount of R^2 is 0.612 . so 61.2% of the variation in the amount of loss can be explained by the variables presented in Eq.2 .

$$Rizesh = 9.22 - 3.27Taybad - 1.5Torbat\ jam - 1.34Khaf + 0.44Model + 0.075\,Omr + 0.83Roshan - 2.23Gaskojen - 2.07Sardary - 3.55Abideym - 0.015Amalkard - 0.0017Tarikh + 0.39Saat$$

Table 1: Estimation results of Eq.2

Sig	t	value	Coefficient/statistic
0.068795	1.84907	9.223971902	(Constant)
0.000369	-3.7489	-3.272638095	ANDIMESHK
0.110932	-1.61504	-1.509334587	SHOUSHTAR
0.252536	-1.15401	-1.341054888	DEZFUL
0.528008	0.634308	0.44357626	MODEL
0.046426	2.028521	0.075421104	AGE
0.489179	0.695393	0.831158432	ROSHAN
0.021933	-2.34536	-2.235987668	KASKOJEN
0.167269	-1.39596	-2.07715061	SARDARI
0.008177	-2.7246	-3.557281767	ABIDEYM
0.935807	-0.08084	-0.0151634	AMALKARD
0.933042	-0.08433	-0.001763976	DAYS
0.939048	0.07675	0.398745634	TIME
0	-	8.921218452	F
-	-	0.611549815	R Squre

Now we are looking to improve the level of significance of the model variables and overall significance of regression. For this each of the meaningless variables is omitted in each estimation. By eliminating the nonsense variables in the form of trial and error and comparing the results of different models it is characterized that eliminating Time variables with only 0.07 confidence levels is the best way to increase the capability of estimating of the model. The new model consistent with the results of table 2 is presented below:

Eq.3

$$Rizesh = 9.44 - 3.27Taybad - 1.52Torbat\ jam - 1.33Khaf + 0.43Model + 0.075Omr + 0.817Roshan - 2.23Gaskojen - 2.08Sardary - 3.56Abideym - 0.017Amalkard - 0.0019Tarikh$$

Table 2: The results of Eq.3 estimation

Sig	t	amount	Coefficient/statistic
0.023392	2.318498	9.441836	(Constant)
0.000331	-3.77922	-3.27428	ANDIMESHK
0.099418	-1.67016	-1.52265	SHOUSHTAR
0.250111	-1.15984	-1.33502	DEZFUL
0.52799	0.634288	0.437781	MODEL
0.04252	2.066676	0.075756	AGE
0.488316	0.696725	0.817461	ROSHAN
0.020859	-2.36477	-2.23763	KASKOJEN
0.162118	-1.41311	-2.08387	SARDARI
0.007515	-2.75424	-3.56338	ABIDEYM
0.923665	-0.09617	-0.01764	AMALKARD
0.925426	-0.09394	-0.00194	DAYS
0	-	9.874	F
-	-	0.612	R Squre

Like what was said about the equation 2 value of variables coefficients, t-statistics and significance levels can be easily extracted from Table 2.

The value of the F statistic is 9.8 times that confirms the high level of regression significance. Statistic R^2 is 0.612. Thus, 61.2 percent of the loss can be explained by the independent variables. Also by the comparison of tables 1 and 2 can be easily found that the accuracy of Eq.3 in estimating the loss to Eq.2 is more, so just an interpretation of the coefficients of Equation 3 will be discussed.

If the values of all quality variables are equal to zero (the baseline for each variable) loss equation will be obtained in terms of quantitative variables, yield and harvest date on the basis of qualitative variables in the following equation.

Eq.4

$$Rizesh = 9.44 + 0.075Omr - 0.017Amalkard - 0.0019Tarikh$$

Eq.4 shows the loss equation in terms of quantitative variables in the harvest in the city of Ahvaz (baseline of region variable) using combine Sahand S68 (ground state of variable combine models), using a variety plateau (baseline varieties variables) and rainfed (the basic state of culture variables) so this relationship will be used to predict the loss under experimental conditions.

Suppose we want to give more flexibility to the model Eq.4, so that the amount of loss can not only be predicted in the city of Ahvaz , but also can be predicted in all the cities. For this it is necessary to add the four qualitative variables, Andimeshk, Shoushtar and Dezful to the Eq.4:

Eq.5

$$Rizesh = 9.44 - 3.27Taybad - 1.52Torbat\ jam - 1.33Khaf + 0.075Omr - 0.017Amalkard - 0.0019Tarikh$$

So relation 5 is able to predict the amount of loss using combine Sahand S68, cultivating of plateau Varieties and dry irrigation at 4 cities of Ahvaz, Shoushtaer, Dezful and Andimeshk.

Now suppose the following conditions that Eq.4 predicts the loss in using different combine models instead of in different cities. In this case the qualitative variables Model must be added relation 4 to obtain the new model as follows.

Eq.6

$$Rizesh = 9.44 + 0.43Model + 0.075Omr - 0.017Amalkard - 0.0019Tarikh$$

The relation is able to predict the amount of loss of plateau varieties and rain fed irrigation under the condition of using different combine models in the city of Ahvaz. With similar method some relationships can be presented in a manner that can predict the amount of loss in various situations. For example, if we want to forecast the amount of loss in different areas and different varieties by the use of

Sahand S68 combine and rain fed irrigation Eq.7 can be used.

Eq.7

$$Rizesh = 9.44 - 3.27Taybad - 1.52Torbat\ jam - 1.33Khaf + 0.817Roshan - 2.23Gaskojen - 2.08Sardary + 0.017Amalkard - 0.0019Tarikh$$

If we want to provide a link to predict the loss under all Variable factors conditions equation 3 that all variables have been used in it will be used.

Another point is the interpretation of the coefficients of the variables in Equation 3. The coefficient of each qualitative variable in fact indicates the differences with the baseline of that variable. For example, the coefficient -3.27 related to the quality variable Andimeshk shows that the average loss in the city is 3.27% less than baseline. So the average loss of Andimeshk is 3.27 percent less than the city of Ahvaz. In the same way, it can be argued that the average loss occurring in the city of Shoushtaer and Dezful are respectively 1.52% and 1.33%, lower than the city of Ahvaz.

Coefficient 0.43 of variable Model indicates that when the variable is equal to one (using John Deere Combine) compared to baseline (zero) loss increase 0.43 percent. Similarly, it can be said about the variety variable that the amount of light falling is 0.817 % higher than the plateau and the amount of SARDARI and Gaskvzhn loss are 2.23% and 2.08% lower than plateau respectively. Abideym the ground state variable is rain fed. Thus, the loss of irrigated agriculture on average is 3.56% lower than the rain fed.

The remained point is that the information above is true about variables that t-statistic and significant level indicates a high confidence degree of the estimated coefficient.

For example, about the variable Days with the confidence level 0.07 times can be said that the estimated coefficient is not statistically significantly different from zero. This is largely due to the vast differences that exist at the time of harvest, by farmers in each region. So it makes the possible to compare the effects of different harvesting times on the loss practically impossible and hence coefficient related to the harvest time in equation 3 is meaningless.

References

Bukhari, S.B., Baloch ,J.M., Rattar F.M. (1983) Losses in wheat Harvesting and threshing. Agricultural Mechanization in Asia , Africa and Latin America 14(4): 61-67

Dillon, J., L. Hardaker and J. Brian. 1993. Farm management research for small farmer development. FAO Farm systems management series. FAO. No.6:302p.

Evatt Foundation (1991) Australian South Sea Islanders: A Report on the Current Status of South Sea Islanders in Australia, Sydney: Evatt Foundation.

Lockie, S. (1997) 'Is 'subsumption' still relevant? The question of control in Australian broadacre agriculture', Rural Society, 7 (3/4): 27-36.

Mabbett, J. and Carter, I. (1999) 'Contract farming in the New Zealand wine industry: An example of real subsumption', in Burch, D., Goss, J. and Lawrence, G. (eds) Restructuring Global and *Regional Agricultures: Transformations in Australasian agri-food economies and spaces*, Aldershot: Ashgate, 275-288.