A REGRESSION MODEL FOR THE TONS OF PRUNING PER HECTARE

Ana María Martínez-Rodríguez PhD
Francisco Javier Gallego PhD
Julio Terrados PhD
Eulogio Castro PhD
University of Jáen , Spain

Abstract

More than nine million hectares of olive trees are cultivated all over the world, but especially in the Mediterranean countries. As an essential operation, pruning of olive trees produces a huge amount of biomass which is lacking of industrial applications and must be eliminated from fields to prevent propagation of vegetal diseases. This work deals with the development of a regression model to be used in the estimation of the real biomass production from olive tree pruning.

Keywords: Regression model, Olive trees, Biomass production

Introduction

More than nine million hectares of olive trees are cultivated all over the world, but especially in the Mediterranean countries. From the pruning operations carried out in the plantation a huge amount of biomass can be obtained (McKendry, 2002). These residues are usually eliminated by in-field burning or by grinding and scattering in order to prevent propagation of vegetal diseases so there is no an economic benefit (Askew and Holmes, 2001). As an alternative, the use of olive tree pruning biomass as raw material for ethanol and other chemicals production or as a source of energy has been proposed (Cara et al, 2008a, 2008b).

The quantification of the yearly available biomass is an issue that has not yet been properly established. The published reports estimate the production of biomass in a range as wide as between 1 and 5 tons per hectare.

This work deals with the development of a regression model for the number of tons per hectare a year that can be used in the estimation of the real biomass production from olive tree pruning.

The remainder of this work is structured as follows. In Methodology section the regression model development is described. The main regression coefficients are interpreted in Results and finally, some conclusions are drawn in Conclusions.

Methodology

In order to estimate the biomass production from olive tree pruning Jaén has been divided in six homogeneous agroclimatic areas of olives trees. These areas have been treated as strata in a stratified sampling. In fact, the considered areas are the areas of the Phytosanitary Alert and Information Network (RAIF) developed in Andalusia since 1996 in order to know the actual phytosanitary state of the main crops in the region. This RAIF areas where grouped according to the experience of the experts at RAIF network.

The experimental determinations consisted in weighing the actual biomass produced by pruning by means of a dynamometer, the moisture content was taken into account in order to calculate de dry biomass obtained. At the same time, several important factors relative to the pruning location were also recorded, among them: slope, irrigation, tree density, water soil retention, soil capability, altitude, number of stems of the tree and density (number of trees by ha). A regression model was developed to predict the dry biomass obtained from the explicative or independent variables. Regression models find a relationship between the independent variable and the explanatory variables as an equation in which the independent variables have parametric coefficients, which may enable future values of the dependent variable to be predicted and also give an interpretation about the behavior of the dependent variable when there is variation in the value of significant factors. In addition, in the regression model developed, as several qualitative factors have been considered, dummy variables were introduced to evaluate the effect that the presence or absence of different level of the qualitative variables may have in the weighted dry biomass obtained. Table 1 shows de factor and levels analyzed in the regression study. The approach used to fit the regression model has been a stepwise regression by backward elimination, so in an starting point all independent variables were included in the model, in the following steps it was decided the deletion of each variable using p-values as model comparison criterion, deleting the variable (if any) that improved the model the most by being deleted, and repeating this process until no further improvement was possible.

Factor	Number of levels	Levels
Raif area	6	Sierra Morena, Campiña, Mágina Sur, Loma, Sierra de Cazorla, Sierra de Ahillos
Slope	3	Low (< 8%) Moderate (8% < x <15%) High (≥15%)
Soil depth	3	The soil depth present in the sample points varies between 2 and 4
Irrigation	2	No irrigation Irrigated
Soil capability	4	Capability_1: marginal and unproductive land Capability_2: moderate to marginal productive land Capability_3: good to moderate productive land Capability_4: excellent productive land
Density	3	Low (<70) Medium (70<x<140) High (≥140)

Table 1. Factors and levels

In the development of the model using the number of tons by hectare a year as dependent variable several regression models were developed but, as the specification of the model has being tested by the Ramsey regression equation specification error test (RESET) showed p-values under 0.05, the dependent variable was transformed and the log of the number of tons by hectare a year was considered. Also, initial testing, for simplicity, was by a linear model. In order to improve the coefficient of determination non-linear relations formed by product of the independent variables were analyzed, it has to be taken into account that if an interaction where a factor is involved is going to be analyzed it is necessary to consider a product between each combination of the factor levels. Specifically:

Interaction 1 = No irrigation*Number of stems
Interaction 2 = Medium density*NumPies
Interaction 3= High density* Number of stems
Interaction 4= High slope* Number of stems

Interaction 5= Moderate slope* Number of stems
Interaction 6= altitude* Number of stems
Interaction 7= Water oil retention* Number of stems

After the regression model was fitted, the diagnosis of the model was performed showing the presence of heteroscedasticity. For this reason, a model with the coefficients estimated under the presence of heteroscedaticity was fitted. The Table 2 shows the estimated coefficients of the significant variables and interactions. The characteristic parameters of the regression model calculated are shown in Table 3. The p-value in the ANOVA table is smaller than 0.01, so there is a statistically significant relationship among the variables of the model, with a 99% of confidence. The R2 of 0.67 indicates that the model explain approximately 67% of the variability in the log of the quantity of the dry biomass obtained. Also, normality has been tested with several normality tests and it cannot be rejected with a 5% of signification.

Depedent variable: log t/ha

	Coefficient	Standard deviation	t-statistic	p-value	
constant	0.156795	0.36052	0.4349	0.66433	
Loma	-0.255239	0.0499084	-5.1141	<0.00001	***
Campiña	-0.175203	0.0740684	-2.3654	0.01944	**
Mágina Sur	-0.52876	0.0662442	-7.9820	<0.00001	***
Moderate	0.207764	0.0592541	3.5063	0.00062	***
No_irrigation	-0.803963	0.200794	-4.0039	0.00010	***
Soil_depth_2	-0.710603	0.311134	-2.2839	0.02395	**
Soil_depth_3	-1.12647	0.317666	-3.5461	0.00054	***
Interaction1	0.260033	0.0782257	3.3241	0.00114	***
Interaction2	0.122745	0.0251056	4.8892	<0.00001	***
Interaction3	0.25827	0.041314	6.2514	<0.00001	***
Interaction4	0.0809807	0.027494	2.9454	0.00380	***
altitude	0.002006	0.000318249	6.3032	<0.00001	***
Interaction6	-0.00069321	0.000119523	-5.7998	<0.00001	***
Interaction7	0.00204773	0.000321622	6.3669	<0.00001	***

Table 2. Estimated coefficients and p-values

R^2	0.674127	R^2 corrected	0.640080
F(14, 134)	19.80021	p-value (of F)	3.95e-26
Log-likelihood	-270.7967	Akaike	571.5935
Schwarz	616.6527	Hannan-Quinn	589.9003

Table 3. Regression model characteristics

It should be pointed out that related to use of the developed regression model in prediction, as the dependet variable is the logarithm of the amout of pruning per hectare it is necessary make a correction multiplying by $exp\left(\frac{\sigma^2}{2}\right)$, that is, in this case:

$$\hat{Y} = \exp\left(\frac{0.394361^2}{2}\right)\exp\left(\widehat{lnY}\right) = 1.0810\exp(\widehat{lnY})$$

Where $\widehat{lnY}$ is obtained by the regression model.

Results

In this case, as the dependent variable is the logarithmic of the original variable, for the explanatory quantitative variables an increment of a unit in X_j represents an increment in the original variable of $100[e^{\beta_j} - 1]$%. In this case, specifically:

- As the level “Loma” of the factor RAIF has a coefficient -0.255239 that means that in La Loma the amount of pruning per ha. is a 22.53% less than in Sierra de Cazorla (that was the level considered as the base level).
- As the level “Campiña” of the factor RAIF has a coefficient -0.175203 that means that in Campiña the amount of pruning per ha. is a 16.07% less than in Sierra de Cazorla.
- As the level “Mágina Sur” of the factor RAIF has a coefficient -0.52876 that means that in La Loma the amount of pruning per ha. is a 41.1% less than in Mágina Sur.
- As the level “No irrigation” has a coefficient -0.803963 that means that the amount of pruning per ha. is a 55.24% less than in irrigated lands.

The interaction coefficients show that the independent variable behavior in relation of one of the explanatory variables depends on the other variable considered in the interaction. For instance the behavior of the amount of pruning in relation to the altitude depends on the number of the stems of the tree.

Conclusion

The results show that the homogenous areas considered have influence on the quantity of biomass produced by pruning olive trees, in other words, there is a statistically significant difference in the weighted dry biomass obtained in the different areas of Jaén. Other significant factors are the age of the tree, height of the tree and the number of trunks and the tree density. Interesting conclusions can be obtained from the parameter estimates. In relation with the statistical significance of the model it should be pointed out that the p-value in the ANOVA table is smaller than 0.01, so there is statistical significant relationship among the dependent variable and the significant explanatory variables with a 99% level of confidence. Once the regression model has been fitted, the goodness of fit of the model has been checked by the diagnosis of the model showing that there are not violations of the statistical assumptions. Also, the specification of the model has being tested by the Ramsey regression equation specification error test (RESET). The p-values obtained are greater than 0.01 so there are not statistical reason to think that the model suffers from miss-specification. In this sense the regression model fitted is adequate to predict the dependent variable.

Acknowledgements

This work was supported by the excellence project P09-TEP-5254 of the Plan Andaluz de Investigación, Desarrollo e Inovación of the Junta de Andalucía (Spain).

References:

Askew M.F. and Holmes, C.A. The potential for biomass and energy crops in agriculture in Europe, in land use, policy and rural economic terms. International Sugar Journal (2002), 104(1247) 482-591.

Cara, C., Ruiz, E., Oliva, J.M., Sáez, F. and Castro, E. Conversion of olive tree biomass into fermentable sugars by dilute acid pretreatment and enzymatic saccharification. Bioresource Technology, 2008, 99 (6), 1869-1876.

Cara, C., Ruiz, E., Ballesteros, M., Manzanares, P., Negro, M.J. and Castro, E. Production of fuel ethanol from steam-explosion pretreated olive tree pruning. Fuel, 2008, 87 (6), 692-700.

McKendry, P. Energy production from biomass. Overview of Biomass Bioresource Technology, 2002, 83 (1), 37-46.

RAIF (2014), http://www.juntadeandalucia.es/agriculturaypesca/raif/.