academicJournals

Vol. 9(10), pp. 914-920, 6 March, 2014
DOI: 10.5897/AJAR2013.7282
ISSN 1991-637X

http://www.academicjournals.org/AJAR

African Journal of Agricultural Research

Full Length Research Paper

Applied multiple regression for autocorrelated sugarcane data

Gabriele Torino Pedroso[1], Renata Silva-Mann[2], Maria Emília Camargo[3] and Suzana Leitão Russo[4]*

[1]Post Graduation Program in Biothecnology, Universidade Federal de Sergipe, Brazil.
[2]Department of Agronomy, Universidade Federal de Sergipe, Brazil.
[3]Department of Management, Universidade de Caxias do Sul, Brazil.
[4]Department of Statistics and Actuarial Sciences, Universidade Federal de Sergipe, Brazil.

Received 13 July, 2013: Accepted 20 February, 2014

The major concern with the substitution of fossil fuels is the air pollution which contributes to countries like Brazil to adopt new sources of clean and renewable energy as ethanol from sugarcane. It is extremely important to maximize the production of sugarcane mainly focused on the production of ethanol. This work aimed to establish a model of multiple regression that targets sugarcane traits, which could contribute to prediction of alcohol production. The data were obtained at São José dos Pinheiros distillery in Laranjeiras municipality, State of Sergipe. The sugarcane traits assessed were production of alcohol, Brix, percentage of sucrose (Pol) and fiber content. A descriptive analysis was carried out to verify the homogeneity of the traits and multiple regression analysis. To guarantee the consistency of results, the verification of the coefficient of multiple determinations, the significance of the model (ANOVA), the Student t-test and the mean absolute percentage error (MAPE) was performed. The software utilized was the Statistica. Through the regression by the method of ordinary least squares (OLS), the simulation results pointed an autocorrelated residue. Therefore, the application of a correction method, the iterative process Cochrane-Orcutt, to remove the autocorrelation of the variables is required. The new model allows found statistically significant coefficients at the 5% probability. With the new equation, it can be concluded that values of Pol and Fiber content in sugarcane may assist the prediction industrial sugarcane production.

Key words: Multivariable analysis, autocorrelation, *Saccharum* spp.

INTRODUCTION

The sugarcane (*Saccharum* ssp.) is a species from Asia which was brought to Brazil during the colonization in the first decade of the sixteenth century.

The social-economic importance of sugarcane has ancient roots in the national agribusiness due to potential production, which adapts easily in tropical climate with

*Corresponding author. E-mail: suzana.ufs@hotmail.com.

specific rainfall and solar lighting, besides be source for sugar and biomass energy.

Brazil is the world largest producer of sugarcane, with 604,513,600,000 tons in 2009/2010 crop year, in an area of 604,513.6 billion hectares, and with a yield of 81.585 kg/ha (CONAB, 2010). This performance is due taking into account, that the real agricultural productivity in many regions is only a small part of the real genetic potential of the crop (Taiz and Zeiger, 2004).

Interuniversity Network for the development of sucro-energetic sector (RIDESA) is an example of successful breeding networking in Brazil. For the logistic, each university creates clones in their respective states from seeds produced at “Serra do Ouro” Germplasm Bank. Annually, the best clones selected from each university are interchanged, which allow an increasing number of genotypes to be evaluated by each university in their respective edaphic and climatic conditions. The network is constituted by ten federal universities: UFPR, UFSCar, UFV, UFRRJ, UFS, UFAL UFRPE, UFG, UFPI and UFMT.

At the agro-industrial fields, new experimental plots are established for a period of 3 consecutive years. This phase is called Experimental Phase (EP). The promissory material is able to be registered as cultivars.

RIDESA was established in order to incorporate the activities of the extinct Planalsucar Program, and has the development continued for research aiming to improve the productivity of the sector. Its activities’ are based on thedevelopment of experimental research in 31 research stations strategically located in states where the sugarcane crop has higher expression.

In the state of Sergipe are six sugarcane agro-industry as São José dos Pinheiros in Laranjeiras municipality; the Taquari, UTE- Iolando Leite (Old distillery Carvão) and the Junco Novo, located in the municipality of Capela-SE; CBAA-Japoatã in Japoatã and the Campo Lindo, in the municipality of Nossa Senhora das Dores. Of these six distilleries, only São José dos Pinheiros is a RIDESA partnership and its data were used in this work.

Ethyl alcohol or ethanol is a flammable liquid obtained by the distillation of fermented organic products such as sugar, starch and cellulose. Ethanol from sugarcane is obtained from sucrose (sugar) broth. At the end of the distillation and rectification process, the product generated is an ethanol-water mixture that can be used directly as automotive fuel, or can undergo dehydration to obtain the anhydrous alcohol which is used in a mixture with gasoline as addictive.

According to information from National Electric Energy Agency (ANEEL), Brazil currently has 266 distilleries burning bagasse that produce electricity. These companies work with capacity of 3,682 MW, equivalent to 3.5% of Brazil's total generating facilities, or 16% of the energy produced by thermal sources in the country.

In recent decades, Brazil has doubled the area planted with sugarcane, primarily due to the production of alcohol, this expansion occurred mainly in regions with fertile soil and favorable climate (Takeshi, 2008).

For the production of alcohol (ethanol), Brazil is the second major producer, behind only the USA. Both countries account for 70% of all ethanol produced in the planet (Council for Biotechnology Information, 2009). However, the American product is exclusively derived from corn and supply the internal market. Thus, Brazil is also the world's largest ethanol exporter accounting for 54% of the market.

The production of ethanol from sugarcane contributes to environment preservation and sustainability in supply chains. Their systematic use can promote technical progress and the advent of positive changes in ecological impact and working conditions. In addition, the sugarcane is CO_2 sequester from atmosphere which is important to reduce the effect of global warming (Orlando Filho, 2007).

The prediction of alcohol production for sucro-energetic agro-industries, or even for small producer is extremely important aiming strategies to make decision on planting dimension, and reduce the risks of this farming activity. However, it is not just the knowledge that represents the characteristics of sugarcane production; we can estimate models to relate independent production traits aiming prediction of the possible commercial alcohol production. The method of multiple regression permits to create an equation showing the relationship between independent variables which influence on a dependent variable.

The objective of this work was to establish a model by multiple regression analysis, aiming insert variables of sugarcane, which could contribute to the prediction of ethanol production.

MATERIALS AND METHODS

The data of ethanol production were obtained at São José dos Pinheiros distillery and refers to period of 2009/2010. The traits used were Brix (total solid content, sugars and non-sugars), the percentage of sucrose (Pol), the Fiber content expressed in ton per day, and the alcohol yield expressed in liters per day. A descriptive analysis of the data was performed to verify the homogeneity of variables. The software Estatistica was used for all estimative.

Model of multiple regression

The multiple regression analysis according to Tabachnick and Fidell (1996) is defined as a set of statistical techniques that allows the evaluation of the relationship among a dependent variable, and several independent variables. The major objective of this analysis is to identify the equation that describes the relationship between these variables so that we can predict the value of dependent variable attributing values for the independent variables (Ragsdale, 2001; Subramanian et al., 2007).

According to Anderson et al. (2002), this model can be described as:

$$y = \beta_0 + \beta_1 . x_1 + \beta_2 . x_2 + \ldots + \beta_p . x_p + \varepsilon \quad (1)$$

Where $x_1, x_2, \ldots, x_p$ are constants and $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are

parameters called partial regression coefficients and ε the residues.

The following assumptions is assumed for the model:

(1) The error vector $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]$ is random, that is, the components $i = 1,2,\ldots,n$ are random variables;
(2) The hope of each component of ε is zero, that is, $E(\varepsilon) = 0$;
(3) The components of the vector ε are not correlated or better $(\varepsilon_i, \varepsilon_j) = 0$, i≠j and have constant variance, ... Thus, the covariance matrix of ε is the diagonal matrix...In, where In is the identity matrix of order n, $V(\varepsilon) = \cdots In$. The distribution of $\varepsilon_i \; i = 1,2,\ldots,n$ is the Normal.

Given this assumption, we have the model:

$$y/_X \sim Normal\,(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k) \quad (2)$$

To assure the quality of the results, the coefficients of the multiple regression model, the significance test, the analysis of residues and the Mean absolute percentage error (MAPE) were calculated

The coefficient of multiple regression

To guarantee the quality of the results, the coefficients of the multiple regression, the significance test, and the analysis of residues were estimated.

The partial coefficient of regression

This coefficient measures the strength of the relationship between the dependent variable and an independent, when the effect of the other independent, are constant. This value is used to identify the independent variable with the high explanatory power of the dependent variable in addition to other variables in the model.

The beta coefficient

This is the coefficient of standardized, or that is used for a process in which the original data are processed into new variables with mean equal to zero. A standard deviation, and thus the term βo (intercept) assume the value of zero thereby enabling a direct comparison among the coefficients and relative powers of explanation of the dependent variable.

The coefficient of correlation (R)

This coefficient indicates the strength of association between any two metric variables. The value can range from -1 to +1 and the sign indicates the direction of the relationship.

The coefficient of multiple determinations (R^2)

R^2 measures the degree of the regression equation fits to, and it indicates the proportion of variance in the dependent variable according to its mean that is explained by the independent variables (Gujarati, 2000).

The significance test of the model

The significance test of the multiple regression model was performed, the F-test which is the average square divided by the average of the squared residuals (errors). According to Levine et al. (2000), this test is used to test whether there is a significant relation between the dependent variable and the entire set of independent variables.

Considering that there is more than one independent variable, you use the null hypothesis and the alternative hypothesis:

$H_0: \beta_{1=}\beta_{2=} \ldots = \beta_{K=}0$ (There is no linear relation between the dependent variable and the independent variables.)
H_1: At least one non-zero B_j, j=1,2,…,k (There is a linear relationship between the dependent variable and at least one of the independent variables).

The residues

To verify whether a model is appropriate, it is necessary to investigate if the assumptions made for the development of the model are satisfied.

For this purpose, the behavior of the model is studied using the set of observed data, notably the discrepancies between the observed values and the adjusted values by the model and this way proceeded to analyze the waste (Bussab, 2002).

Basically, this analysis provides evidence of possible violations of the model assumptions, such as normality and homoscedasticity, and when necessary it provides signs of lack of adjustment from the proposed model (Charnet et al., 1999; Subramanian et al., 2007). The residue for the observation *i* is the difference between the observed value of y_i and your estimated value:

$$\hat{u}_i = y_i - \hat{y}_i \quad (3)$$

According to Russo et al., (2010), we can evaluate the suitability of the adjusted regression model by plotting the residues on the vertical axis and the corresponding values to the values of X_i of the independent variable on the horizontal axis. If the adjusted model is appropriate for the data, there will be no apparent pattern of waste relative to X_i. However, if the adjusted model is not appropriate, there will be a relation between the values of X_i and the residues ε_i (Levine et al., 2000).

The best model will be the one in which is calculated the lowest MAPE:

$$MAPE = \frac{\sum\left|\frac{Xi-Yi}{X_i}\right|}{n}.100\% \quad (4)$$

Where X_i are the observed and estimated values. Then, it was detected the existence of autocorrelation of the variables included in the model.

The autocorrelation

When the hypothesis of homoscedasticity (constant variance) is not kept, we say that the errors are heteroskedastic.

When $Var\,(u_t/X)$ depends on X, it often depends on the explanatory variables of the model t,x_t . When considering the hypothesis of the absence of serial correlation $Corr\,(u_t, s_t) = 0$ false, we say that the errors suffer serial correlation, or autocorrelation, because they are correlated over time (Wooldridge, 2006).

The autocorrelation function measures the correlation degree from a variable, in a certain instant, with itself, in an instant of time later. It allows analyzing the degree of irregularity of a variable.

The test to detect the autocorrelation was developed by the statisticians J. Durbin and G.S. Watson and it is commonly known

as statistical d of Durbin-Watson and it is defined as the ratio between the sum of squared differences of successive residues and the residuals squared sum (RSS).

The Durbin-Watson statistic evaluates the existence of residuals independence, that is, it tests the null hypothesis that the covariance between the residuals variables is zero. The reference value of Durbin-Watson should be 2.00, so in that way the correlation does not occur.

Based on the results obtained through the regression analysis, it was necessary for the application of the Cochrane-Orcutt corrective method to remove the autocorrelation of the variables.

Cochrane-Orcutt corrective method

In the presence of serial correlation, the estimators of ordinary least squares (OLS) are inefficient and it is essential to seek corrective measures. The solution depends on the knowledge about the nature of interdependency of disturbances; one of the situations (addressed in this study) is when the structure of autocorrelation is unknown (Wooldridge, 2006; Gujarati, 2000).

Suggestion is used for the interactive process of Cochrane-Orcutt, which is a method based upon the statistic of Durbin-Watson that uses estimated residuals $\hat{u}_t$ to obtain information about the unknown (Gujarati, 2000).

Consider the model of two variables:

$$y_t = \beta_0 + \beta_1 . x_1 + \mu_t \quad (5)$$

And suppose that $\hat{u}_t$ is generated by the scheme AR(1), that is,

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t \quad (6)$$

This way, Cochrane and Orcutt recommend the following steps to estimate:

(1) Estimate the model by the usual routine of the OLS and obtain the residues, $\hat{u}_t$.
(2) Using the estimated residuals, run the following regression:

$$\hat{u}_t = \hat{p}\hat{u}_{t-1} + V_t \quad (7)$$

(3) With the result of Equation 5, run the general difference equation:

$$(Y_t - \hat{\rho} Y_{t-1}) = \beta_1 (1 - \hat{\rho}) + \beta_2 X_t - \hat{\rho}\beta_2 X_{t-1} + (u_t - \hat{\rho} u_{t-1}) \quad (8)$$

or

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \hat{e}_t \quad (9)$$

(4) At first, we are not sure whether the result of the Equation 9 is the “best” estimate. Replace the values of $\beta_2^* = \beta_1(1-p)$ and β_2^* obtained from the Equation 9 in the original regression of Equation 5 and obtain the new residuals, shall we say, like this:

$$\hat{u}_t^{**} = Y_t - \hat{\beta}_1^* - \hat{\beta}_2^* X_t \quad (10)$$

which can be easily calculated, because Y_t, X_t, β_1^* and β_2^* are all well-known.

(5) Now estimate this regression:

$$\hat{u}_t^{**} = \hat{\rho}\hat{u}_{t-1}^{**} + w_t \quad (11)$$

which is similar to Equation 5. Thus, $\hat{\rho}$ is the estimate of the second hound of ρ. As we do not know whether this estimate of the second hound of ρ, we may enter in the estimate of the third hound, and so forth. As suggested by the previous steps, the Cochrane-Orcutt method is interactive (repetitive). But how far should it proceed? Generally, to stop with the successive repetitions of ρ when this differs from values lower than 0.01 (Gujarati, 2000).

Then, it was carried out a regression variance analysis.

Regression variance analysis

The idea of the variance analysis is to compare the variation due to the treatments (varieties) with the variation due to the chance (residual), according to Medeiros (1999). In order to calculate the variance analysis, we should first calculate the degrees of freedom (G1), the correction factor (C), so that we calculate the sum of squares, the squared mean and the value of F.

In an experiment with K treatments and r repetitions, the average treatments indicated by $Y_1, Y_2, Y_3, \ldots, Y_k$ and the grand total which is given by the total sum of the treatments being = Σy, we should have:

The degrees of freedom for the treatments = k-1; for the total = n-1, with n = kr; for the residuals = (n-1)-(k-1) = n-k

The Correction factor

$$(C) = (\sum y)2/n \quad (12)$$

The sum of total squares

$$(SQT) = \sum Y^2 - C \quad (13)$$

The sum of squares of treatments

$$(SQTr) = \sum k^2/r - C \quad (14)$$

The sum of squares of residuals

$$(SQR) = SQT - SQTr \quad (15)$$

The mean square of treatments

$$(QMTr) = SQTr/k - 1 \quad (16)$$

The mean square of residual

$$(QMR) = SQR/n - k \quad (17)$$

The value

$$F = QMTr/QMR \quad (18)$$

For theoretical reasons, a variance analysis should only be applied to a set of data when it can admit certain assumptions, such as the independence, normality and homoscedasticity (Vieira and Hoffmann, 1989).

The F-value is linked to the number of freedom degrees of treatments (numerator) and to the number of freedom degrees of residual (denominator) and it is the value which measures the level of significance. When the calculated F is higher than the value of the critical F (tabulated), we can conclude that the regression is significant.

When the F calculated is greater than the F critical value (tabulated), it can be concluded that the regression is significant (LEVIN AND FOX, 2004).

Table 1. Summary of statistical descriptive analysis of the model of multiple regression for ethanol production of data obtained in São José dos Pinheiros distillery, Laranjeiras-SE, 2011.

Variable	Average	Median	Variance	Standard deviation	CV%
Ethanol production	101,746.0	113,619.0	921,080,129	30,349.30	0.30
Pol	17.2	17.1	1	0.84	0.05
Fiber	14.7	14.9	0	0.61	0.04

Table 2. Model of coefficients of multiple regression for ethanol production obtained from data of the São José dos Pinheiros distillery, Laranjeiras-SE, 2011.

Variable	Beta	Standard coefficient Beta	B	Standard error of B	t	p-value
Intercept			58,032.8	66,595.76	0.8714	0.3853
Pol	0.2938	0.0864	13,242.5	3,897.95	3.3972	0.0009
Fiber	-0.2218	0.0864	-15,168.3	5,914.80	-2.5644	0.0116

Dependent Variable: Alcohol production (Palcohol).

RESULTS

Descriptive analysis of the variables

The record of the descriptive statistics of the multiple regression for alcohol production is described in Table 1. The model was used to verify the homogeneity of variances.

Table 1 show the Pol and Fiber variables with coefficients of variation (CV) of 0.05 and 0.04%, respectively for the most homogeneous variables; while Alcohol production (Alcohol yield) with coefficient of variation 0.30% stands out as the most heterogeneous variable.

To create a multiple regression model in which the alcohol production is dependent feature, numerous equations with the sugarcane traits would eventually be adopted by the independent by the model in such a way that their coefficients were significant. The model of multiple regression chosen was that with Pol and Fiber production traits as predictors of alcohol production.

From this analysis, the coefficients of the model of multiple regression by the method of OLS for alcohol production carry the following equation which was estimated:

$$\text{Alcohol yield} = \underset{(66{,}595.76)}{58032.8} + \underset{(3{,}897.95)}{0.2938}\,\text{Pol} - \underset{(5{,}914.80)}{0.2218}\,\text{Fiber} \quad (19)$$

$R = 0.37$, $R^2 = 14.13\%$, R^2 Adjust = 12.63%, Standard of estimate = 24955, DW = 1.18, MAPE = 26.78%.

According to linear correlation coefficient (R) found, the level of linear association between these variables is 37%, while 14.13% of the variability in the dependent measure (Alcohol production) is explained by Pol and Fiber, independent variables.

From the student's t-test and the respective p-values (Table 2), we observe that all coefficients are significant at 5% probability, thus, indicating that both Pol as fiber influence on the dependent variable alcohol yield.

DISCUSSION

Considering all observations (119) and the two independent explanatory variables (Pol and Fiber), the observed value for statistic of Durbin-Watson (which assesses whether independence of error, in other words, which test whether the null hypothesis that the covariance between variables residual is zero) was 1.18.

This value is inferior to the reference value of 2.00, indicating that the correlation between residues is present. Thus, it was necessary to apply the corrective method of Cochrane-Orcutt (CO) to remove residues autocorrelation.

Once performed the corrective measures that method of Cochrane-Orcutt proposes. The new data generated were used to perform another multiple regression analysis, with consequence determination of the coefficients and the explanatory equation, as well as the value of Durbin-Watson was calculated for the new model found with the purpose of verifying absence of autocorrelation between residues.

According to Table 3, from the student's t-test and the respective p-values, we can observe that all coefficients are highly significant at 5% probability, thus, indicating that both Pol and Fiber influence on the Alcohol production, dependent variable.

These coefficients contained in Table 3 are related to regression analysis of the model fixed, according to the following equation:

$$\text{Alcohol yield} = \underset{(81766{,}93)}{214{,}944.7} + \underset{(3{,}000.06)}{0.2491}\,\text{Pol} - \underset{(4{,}167.44)}{0.3619}\,\text{Fiber} \quad (20)$$

Table 3. Model of coefficients of multiple regression after removal of autocorrelation for Ethanol production from data obtained in the São José dos Pinheiros distillery, Laranjeiras-SE, 2011.

Variable	Beta	Standard coefficient Beta	B	Standard error of B	T	p-value
Intercept			214,944.7	81,766.93	2.287	0.00973
Pol	0.2491	0.831	8,995.8	3,000.0	2.9985	0.00332
Fiber	-0.319	0.0831	-18,152.2	4,17.44	-4.3557	0.00002

Dependent variable: Alcohol production.

Table 4. Analysis of variance of the regression for the ethanol production obtained from data of the São José dos Pinheiros distillery, Laranjeiras-SE, 2011.

Sum of squares	GL	Mean	Square mean	F	p-value
Regression	2.1753	2	1.0876	14.5133	0.000002
Residue	8.933	116	7.4943		
Total	1.0868				

R = 0.44, R^2 = 20%, R^2 Adjust = 18.63%, Standard of estimate = 24955, DW = 1.95, MAPE = 19.85%.

Based on the estimation of the new linear correlation coefficient (R) for the prediction model of ethanol production, which was 44%, it can be observed that the level of linear association of these variables increased. The coefficient of multiple determination of the model indicates how much variability in the dependent measure (alcohol production) is explained by the independent variables and Pol Fiber also had an increasing (R2 = 20%).

The Durbin-Watson value calculated 1.95 for the new production model of alcohol proves the elimination of autocorrelation between the residues of the series.

The MAPE of the model of multiple regression for ethanol production was 26.78%, and the model generated after the removal of this error autocorrelation decreased to 19.85%, thereby improving the level of fitness of the model.

The analysis of variance contained in Table 4 provides the value for the F-test to verify the null hypothesis with all coefficients and are statistically equal to zero, against the alternative hypothesis that at least one of the coefficients is different from zero. Since the value of the F-test is 14.5133, and confirmed by statistical p-value, 0.000002, it is confirmed that at least one coefficient is statistically different from zero. As the calculated F is higher than the critical F-value (tabulated), it can be concluded that the regression is significant.

According to this study, the model of multiple regression generated from sugarcane data, which is aimed at predicting the future alcohol production presented a residue of autocorrelated variables except Brix. It is necessary to apply a corrective to the other variables method.

After removal of autocorrelations, the new equation generated by the model of multiple regression expresses how each independent variable is influencing the prediction of the dependent variable, which will facilitate the work of the farmer in planning the sugarcane planting and sizing alcohol production.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

Anderson DR, Sweeney DJ, Williams T (2002). Statistics for business and economics. 3 ed., South-Western.

Bussab WO, Morettin PA (2002). Estatística Básica. Editora Saraiva.

Charnet R, Freire CA, De L, Charnet EMR, Bonvino H (1999). Análise de Modelos de Regressão Linear com Aplicações. Campinas, SP: Unicamp.

CONAB (2010). Companhia Nacional de Abastecimento. Available in: http://www.conab.gov.br/detalhe.php?c=18847&t=2#this. Access in May 10 of 2010.

Gujarati DN (2000). Econometria básica. 3. ed. São Paulo: Makron Books.

Levin J, Fox JA (2004). Estatística para ciências humanas. 9 ed. São Paulo.

Levine DM, Berenson ML, Stephan D (2000). Estatística: Teoria e Aplicações. Rio de Janeiro, LTC.

Medeiros E da S (1999). Matemática e Estatística Aplicada. 1 ed, Editora Atlas.

Orlando Filho J (2007). A produção da cane. Available in: http//www.palazo.pro/cana/archives/net. Access in: June 02 of 2010.

Ragsdale CT (2001). Spreadsheet Modeling and Decision Analysis. 3 ed., South-Western College Publishing, Cincinnati, Ohio.

Russo S.L, Camargo ME, Simon VH (2010). Avaliação de perfis sônicos sintéticos em poços de petróleo perfurados nas unidades geológicas pertencentes a bacia sedimentar Sergipe-Alagoas. Revista Gestão Industrial, 06(01):217-238.

Subramanian A, Coutinho AS, Silva LB, de S (2007). Aplicação de método e técnica multivariados para previsão de variáveis termoambientais e perceptivas. Produção, 17(1):052-070, Jan./Apr.

Tabachnick B, Fidell LS (1996) Using multivariate statistics (3ª ed.). New York: Harper Collins.

Takeshi H (2008) Cana no estado de São Paulo. Available in: http//www. Takeshi.inf.br. Access in June 02 of 2010.
Taiz L, Zeiger E (2004) Fisiologia vegetal. Porto Alegre: Artmed.
Vieira S, Hoffmann R (1989) Estatística Experimental. Editora atlas S.A. São Paulo.
Wooldridge JM (2006) Introdução à econometria: uma abordagem moderna. São Paulo: Pioneira Thomson Learning.