



Isotonic regression analysis of Guzerá cattle growth curves¹

Adriano Rodrigues², Lucas Monteiro Chaves², Fabyano Fonseca Silva^{3*}, Idalmo Pereira Garcia⁴,
Darlene Ana Souza Duarte³, Henrique Torres Ventura⁵

10.1590/0034-737X201865010004

ABSTRACT

The objective of this study was to apply data transformation via isotonic regression in growth curves studies of Guzerá cattle whose data presented disturbances characterized by decreased body weight in certain age groups. Weight-age data were collected on newly weaned Guzerá males according to the methodology of weight gain tests (WGT) defined by the Brazilian Association of Zebu Breeders (ABCZ). The Logistic, Von Bertalanffy and Gompertz models were fitted to weight-age data using the generalized least squares method for non-linear regression models through the Gauss-Newton algorithm. The proposed transformation based on isotonic regression theory proved to be efficient; and the Logistic model was the best to describe the growth of animals, with a high percentage of convergence (100%) and goodness of fit assessed by the mean squared error (MSE) and the coefficient of determination (R^2).

Key words: data transformation; isotonization; non-linear models.

RESUMO

Regressão isotônica para estudos de curvas de crescimento de bovinos Guzerá

Objetivou-se aplicar a transformação de dados via regressão isotônica em estudos de crescimento de animais da raça Guzerá cujos dados apresentam distúrbios caracterizados por decréscimos de pesos em determinadas faixas de idades. Foram tomadas observações de peso-idade de machos da raça Guzerá recém-desmamados obedecendo à metodologia das provas de ganho de peso (PGP) definidas pela Associação Brasileira dos Criadores de Zebu (ABCZ). Os modelos Logístico, Von Bertalanffy e Gompertz foram ajustados aos dados de peso-idade por meio do método dos quadrados mínimos ordinários generalizados para modelos de regressão não linear via algoritmo de Gauss-Newton. A transformação proposta por meio da teoria de regressão isotônica mostrou-se eficiente, e o modelo Logístico foi o mais adequado para descrever o crescimento dos animais por apresentar alta porcentagem de convergência (100%) e qualidade de ajuste satisfatória, medida por meio do quadrado médio do erro (QME) e do coeficiente de determinação (R^2).

Palavras-chave: transformação de dados; isotonização; modelos não lineares.

Submitted on November 30th, 2015 and accepted on November 10th, 2017.

¹ Work funded by FAPEMIG (PPM-00563-16) and CNPq (445539/2014-3 and 309188/2014-8).

² Universidade Federal de Lavras, Departamento de Ciências Exatas, Lavras, Minas Gerais, Brazil. adrianorodrigues155@yahoo.com.br; lucas@dex.ufla.br

³ Universidade Federal de Viçosa, Departamento de Zootecnia, Viçosa, Minas Gerais, Brazil. fabyanofonseca@ufv.br; darlene.duarte@ufv.br

⁴ Universidade Federal de Minas Gerais, Departamento de Zootecnia, Belo Horizonte, Minas Gerais, Brazil. idalmo.garcia@gmail.com

⁵ Associação Brasileira dos Criadores de Zebu (ABCZ), Uberaba, Minas Gerais, Brazil. henrique@abcz.org.br

* Corresponding author: fabyanofonseca@ufv.br

INTRODUCTION

In general, a characteristic growth curve of livestock animals has the sigmoid shape and therefore, the nonlinear models, such as Logistic, Richards, Gompertz, Brody, and Von Bertalanffy, have been widely used in the literature in this type of data in the literature (Silveira *et al.*, 2011). Moreover, these models have parameters with biological significance such as adult weight and maturity rate, and their estimates can be used for the evaluation of the growth efficiency of the herd or as phenotypic observations in breeding programs.

The presence of disturbances in the weight-age data contributes to the non-convergence of the algorithms used to estimate the parameters of the growth models and lead to a low goodness of fit of these models (Silveira *et al.*, 2011). These disturbances are characterized by decreases in weight over time due to environmental effects, including lack of nutrients, diseases, and improper management in general. Different studies regarding growth curves in Animal Science mention these decreases, among them are a study by Mendes *et al.* (2009) that worked with Hereford cattle raised in the south of the country, and a study by Silveira *et al.* (2011), that worked with data from crossbred lambs (Dorper x Rabo Largo) raised in Bahia.

One way to overcome this problem is to use a data transformation method that takes into account the expected trajectory of the growth curve. Ramsay (1998) and Hussian *et al.* (2004) suggested a data transformation based on isotonic regression as a technique feasible for pre-correction of longitudinal observations with such disturbances. Gunn & Dunson (2005) also applied this transformation to progesterone concentration data and were able to remove oscillations that were preventing the fit of normal regression models.

From the foregoing, our objective was to apply a data transformation methodology by using isotonic regression in Guzerá cattle growth curves studies whose data presented disturbances characterized by decreased body weight in certain age groups, aiming to avoid possible problems of convergence and to improve the fit of the nonlinear models tested.

MATERIAL AND METHODS

Isotonic regression refers to the fact that an increase in the independent variable will result in an increase in the response variable. Thus, assume that $X = \{t_1, t_2, \dots, t_k\}$ is a finite set of times in which $t_i < t_j$ for $i < j$. A function $f: X \rightarrow R$ is called isotonic or non-decreasing if $f(t_i) \leq f(t_j)$ for $i < j$. If g is a function in X , a function g^* is the isotonic regression of g with weights ω if, and only if, g^* is a monotonic function that minimizes the following sum of squares: $\sum_{t \in X} [g(t) - f(t)]^2 \omega(t)$. Thus, it is said

that g^* is the non-decreasing function closest to the function g weighted by the weights ω .

A widely used Algorithm for isotonic regression is the Pool-Adjacent-Violators algorithm (PAVA) (Barlow *et al.*, 1972). It starts with the function $g(t)$, and if $g(t)$ is isotonic, then $g(t) = g^*(t)$. Otherwise, there must be some index i such that $g(t_{i-1}) > g(t_i)$, being the value $g(t_{i-1})$ called violator. These two values are replaced by their weighted averages, $Av(i-1, i)$, given by: $Av(i-1, i) = \frac{g(t_{i-1})\omega(t_{i-1}) + g(t_i)\omega(t_i)}{\omega(t_{i-1}) + \omega(t_i)}$. After calculating this weighted average, the weights $\omega(t_{i-1})$ and $\omega(t_i)$ are replaced by the sum $\omega(t_{i-1}) + \omega(t_i)$. If this new set, with values $k-1$ is isotonic, then and $g^*(t_{i-1}) = g^*(t_i) = Av(i-1, i)$ and $g^*(t_j) = g(t_j)$, with $j \neq i, i-1$. Otherwise, there is a violator and the preceding process is repeated until a set with isotonic values is obtained. Note that the values of g^* are the weighted averages of the blocks of values for which a violator existed. In this way, isotonic regression is obtained by a simple procedure, with statistical significance given by the use of local averages for the values that presented violators (disturbances in the expected behavior of the curve).

For the growth curves, consider the observations y_i referring to the weights of an animal taken at the times t_i , $i = 1, 2, \dots, k$, and assume $y = f(t, \theta) + \varepsilon$ is a non-linear regression model where: $y = [y_1, y_2, \dots, y_k]$, $t = [t_1, t_2, \dots, t_k]$, $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k]$, assumed as $\varepsilon_i \sim N(0, \sigma^2)$, and θ is the vector of parameters of the model. Since data y_i are not necessarily in non-decreasing order due to environmental oscillations, data transformation to obtain increasing values can be done via isotonic regression, which is called isotonization. With this procedure, the original values $y = [y_1, y_2, \dots, y_k]$ are transformed into values $y = [y_1^*, y_2^*, \dots, y_k^*]$, which are non-decreasing in relation to the evaluation dates. Therefore, the growth models will be adjusted to the data set $y = [y_1^*, y_2^*, \dots, y_k^*]$.

Rodrigues *et al.* (2010) argued that the isotonic regression theory is sufficiently robust to the effects of outliers when the weights ($\omega(t_i)$) are given by the inverse of the difference between the value of the initial datum y_i and the corresponding value after the isotonization $y = y_1^*$, $\omega(t_i) = 1/|y_1 - y_i^*|$, if $y_i \neq y_1^*$. With these weights, the original data can be again isotonized, obtaining new values $y^{**} = [y_1^{**}, y_2^{**}, \dots, y_k^{**}]$. The concept for the use of such weights is that if y_i^* is distant from y_1^* , that is, y_i is a discrepant value in the data set, the weight $\omega(t_i)$ will be relatively small and will decrease its influence in obtaining the isotonized data.

To achieve a transformation even more robust to the influence of the outliers, the above described process can be successively applied to obtain an iterative process as follows. Step 1: Transform the original data $y = [y_1, y_2, \dots, y_k]$ via isotonic regression with equal weights to give $y^{**} =$

$[y_1^{*(1)}, y_2^{*(1)}, \dots, y_k^{*(1)}]$; Step 2: Transform the original data $y = [y_1, y_2, \dots, y_k]$ via isotonic regression with different weights $\omega(t_i) = I / |y_1 - y_i^*|$, if $y_i \neq y_i^*$, to give $y^{*(2)} = [y_1^{*(2)}, y_2^{*(2)}, \dots, y_k^{*(2)}]$; Step m: Transform the original data $y = [y_1, y_2, \dots, y_k]$ via isotonic regression with different weights $\omega(t_i) = I / |y_1 - y_i^{*(m-1)}|$, if $y_i \neq y_i^{*(m-1)}$, to give $y^{*(m)} = [y_1^{*(m)}, y_2^{*(m)}, \dots, y_k^{*(m)}]$. Hence, the process is applied until there are no significant differences between the values isotonized in the $(m-1)^{\text{th}}$ step and the m^{th} step.

The data used in the present study refer to the weight-age observations of Guzerá cattle provided by the Brazilian Association of Zebu Breeders (ABCZ) located in Curvelo-MG. These data were collected from a weight gain test (WGT) performed at Meleiro Farm, which is located in the same municipality. The animal evaluations were performed according to the WGT methodology defined by ABCZ. A total of 45 newly weaned Guzerá males were evaluated, with mean initial weight of 219.9 kg and standard deviations of 38.05 kg and mean age of 325.8 days with standard deviation of 28.0 days. WGT was carried out with the animals fed on *Brachiaria brizantha* pastures and multiple supplementation. In addition to receiving the appropriate sanitary treatment, with the newly weaned animals kept under uniform conditions of management and feeding for 294 days. During the dry and wet seasons, the animals were fed protein salt and underwent usual vermifuge treatment. The animals were weighed at the beginning of WGT and, after an adaptation period of 70 days, they were weighed again to start the test itself. The evaluations were carried out every 56 days during the 224 testing days. Thus, each animal was weighed six times: one initial and five others for the weight gain test.

The growth models Gompertz ($y_i = A \exp(-b \exp(-kti)) + ei$), Logistic ($y_i = A / (b \exp(-kti) + ei)$), and Von Bertalanffy ($y_i = A(1 - b \exp(-kti)) + ei$) were adjusted to the weight-age data using the *nls* function of R (R Development Core Team, 2015), using the generalized ordinary least squares method for nonlinear regression with the iterative Gauss-

Newton algorithm. The weighted and unweighted isotonic regressions were also implemented in R according to the functions developed by Rodrigues *et al.* (2010).

RESULTS AND DISCUSSION

The Table 1 shows the goodness of fit of the models tested, whose Mean Squared Errors (MSE) and R^2 values are represented by the means of values obtained from individual adjustments.

Although the data used in this study presented disturbances in the growth curve shape, no improvements were observed in the percentage of convergence using the isotonization methods when adjusting the Logistic and Gompertz models, once that 100% convergence (Table 1) was found for the three methodologies used (original data, corrected data by isotonic regression with equal weights and corrected data by isotonic regression with different weights). A small improvement in percentage of convergence was found for the Von Bertalanffy model when corrected via isotonic regression with different weights (84%) in relation to the original data (80%).

The Logistic model showed the lowest MSEs (Table 1) regardless of the methodology used. In addition, the isotonic transformations had the expected effect on the fit of this model, so the original data, data transformed by isotonic regression with equal weights, and data transformed by isotonic regression with different weights presented the values of 924.7541, 864.0190 and 630.1000, respectively. Although the differences between the models are small, the Logistic model also showed the best R^2 , especially when using the isotonic transformation with different weights. Thus, based on MSE and R^2 , it is possible to infer that the logistic model via isotonization of the data with different weights best described the weight of Guzerá animals in the conditions of this study. In addition, it also provided 100% convergence, indicating that besides the quality of fit, this model also has benefits in terms of ease of convergence.

Table 1: Goodness of fit (Mean Squared Error, MSE; and Determination Coefficient, R^2) and percentage of convergence (% C) of Logistic, Von Bertalanffy, and Gompertz models fitted for growth curves of Guzerá cattle using different transformation methods (isotonization) for weight-age data

Models	Methodology	% C	MSE	R^2
Logistic	Original (y)	100	924.7541	0.9697324
	Simple Isotonic Regression (y*)	100	864.0190	0.9622322
	Weighted Isotonic Regression (y**)	100	630.1000	0.9818794
Von Bertalanffy	Original (y)	80	125021.6	0.9550946
	Simple Isotonic Regression (y*)	82	121601.8	0.9596846
	Weighted Isotonic Regression (y**)	84	132956.8	0.9694715
Gompertz	Original (y)	100	6204.444	0.9507605
	Simple Isotonic Regression (y*)	100	6169.423	0.9530842
	Weighted Isotonic Regression (y**)	100	6482.893	0.9616000

Table 2: Mean, standard deviations and coefficient of variation (as %) for individual estimates of parameters of the logistic model fitted for growth curves of Guzerá cattle using different transformation methods (isotonization) for weight-age data

Methodology (data transformation)	Parameters		
	A	b	K
Original (y)	384.7524	7.4107	0.0087
	(59.4221)	(1.1151)	(0.0014)
	15.4%	15.0%	16.1%
Simple Isotonic Regression (y*)	385.3174	7.4631	0.0087
	(59.3915)	(1.1079)	(0.0014)
	15.4%	14.8%	16.1%
Weighted Isotonic Regression (y**)	424.8557	7.9059	0.0084
	(59.8319)	(1.2186)	(0.0012)
	15.2%	15.4%	14.3%

The Table 2 shows the means, standard deviations, and coefficients of variation for the estimates of parameters of the Logistic growth model, which presented the best results as shown in Table 1.

Comparing with the estimates of adult weight obtained by Perotto *et al.* (1997) (457 kg) and Tedeschi *et al.* (2000) (466 kg) in growth curve studies of Guzerá cattle using the same models of the present study, we verify that our estimates (Table 2) for the parameter A (adult weight) tend to approximate these values by using the proposed isotonization methods. Thus, we may infer that the isotonization methods contributed to more realistic growth curves, since the disturbances related to decreased weight in certain age ranges led to an underestimation of the adult weight.

CONCLUSIONS

The Mean Squared Error (MSE) and the R^2 values in this study show that the Logistic model via isotonization of the data with different weights was the most suitable to describe growth curves of Guzerá cattle.

Overall, data transformation via isotonic regression analysis increased the convergence percentage and the goodness of fit of the models and provided estimates of adult weight more consistent with the breed standard.

REFERENCES

- Barlow RE, Bartholomew DJ, Bremner JM & Brunk HD (1972) Statistical inference under order restrictions: the theory and application of isotonic regression. London, John Wiley. 388p.
- Gunn LH & Dunson DB (2005) A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics*, 3:434-449.
- Hussian M, Grimvall A, Burdakov O & Sysoev O (2004) Monotonic regression for assessment of trends in environmental quality data. In: European Congress on Computational Methods in Applied Sciences and Engineering, Jyväskylä. Proceedings, Eccomas. p. 24-28.

Mendes PN, Muniz JA, Silva FF, Mazzini ARA & Silva NAM (2009) Análise da curva de crescimento difásica de fêmeas Hereford por meio da função não-linear de Gompertz. *Ciência Animal Brasileira*, 10:454-461.

Perotto D, Castanho MJPC, Rocha JL & Pinto JM (1997) Descrição das curvas de crescimento de fêmeas bovinas guzerá, Gir, holandês x guzerá e holandês x gir. *Revista Brasileira de Zootecnia*, 2:283-288.

R development core team (2015) R: a language and environment for statistical computing. Viena, R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>. Acessado em: 25 de junho de 2010.

Ramsay JO (1998) Estimating smooth monotone functions. *Royal Statistical Society*, 2:365-375.

Rodrigues A, Chaves LM, Silva FF & Zeviani WM (2010) Utilização da regressão isotônica em estudos de curvas de crescimento. *Revista Brasileira de Biometria*, 28:85-101.

Silveira FG, Silva FF, Carneiro PLS, Malhado CHM & Muniz JA (2011) Análise de agrupamento na seleção de modelos de regressão não-lineares para curvas de crescimento de ovinos cruzados. *Ciência Rural*, 41:692-698.

Tedeschi LO, Boin C, Nardon RF & Leme PR (2000) Estudo da curva de crescimento de animais da raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação: avaliação dos parâmetros da curva de crescimento. *Revista Brasileira de Zootecnia*, 5:1578-1587.