



ADRIANO RODRIGUES

**UTILIZAÇÃO DA REGRESSÃO ISOTÔNICA
EM ESTUDOS DE CURVAS DE CRESCIMENTO**

**LAVRAS – MG
2010**

ADRIANO RODRIGUES

**UTILIZAÇÃO DA REGRESSÃO ISOTÔNICA EM ESTUDOS DE
CURVAS DE CRESCIMENTO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. Lucas Monteiro Chaves

Co-orientador

Dr. Fabyano Fonseca e Silva

**LAVRAS- MG
2010**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Rodrigues, Adriano.

Utilização da regressão isotônica em estudos de curvas de crescimento / Adriano Rodrigues. – Lavras : UFLA, 2010.
92 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2010.
Orientador: Lucas Monteiro Chaves.
Bibliografia.

1. Análise de regressão. 2. Modelo não linear. 3. Transformação de dados. I. Universidade Federal de Lavras. II. Título.

CDD – 519.536

ADRIANO RODRIGUES

**UTILIZAÇÃO DA REGRESSÃO ISOTÔNICA EM ESTUDOS DE
CURVAS DE CRESCIMENTO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 23 de julho de 2010.

Dr. Augusto Ramalho de Moraes	UFLA
Dr. Fabyano Fonseca e Silva	UFV
Dr. Joel Augusto Muniz	UFLA
Dr. Idalmo Garcia Pereira	UFVJM

Dr. Lucas Monteiro Chaves
Orientador

**LAVRAS- MG
2010**

À Jesus de Nazaré, único e verdadeiro doutor,
Aos meus pais, Sebastião e Olinda,
À Angela, Tales e Lucas,
que sempre acreditaram em mim,

DEDICO

AGRADECIMENTOS

Primeiramente a Deus por ter me dado a vida e forças ao longo dessa jornada.

Aos meus pais Sebastião e Olinda, pelo exemplo de determinação e pela confiança em mim depositada.

A minha esposa Ângela pela companhia sempre presente em toda a minha caminhada. Agradeço ainda a compreensão pelos momentos em que estive ausente e pela força que sempre me deu em meus projetos.

A meus filhos, Tales e Lucas, que enchem de alegria minha vida e que depositam em mim confiança plena. Agradeço o sorriso de vocês, que me motivou a caminhar até aqui.

A Universidade Federal de Lavras, por ter me oportunizado esta capacitação.

Ao professor Lucas, meu orientador, que caminhou comigo ao longo desses anos, ensinando-me muito com seu exemplo de caráter e dedicação.

Ao professor e amigo, Fabyano, que acreditou em mim mesmo antes de eu entrar neste curso. Obrigado por não medir esforços para ajudar-me, co-orientando e me ensinando muito, com seu exemplo de pesquisador nato.

Aos colegas Renata e Edcarlos, irmãos da isotonização. Agradeço ainda pelo incentivo que sempre me deram, nas discussões e estudos.

Aos professores do programa de Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA, pelo conhecimento transmitido durante o curso.

Ao professor Idalmo, por mediar a concessão dos dados utilizados na pesquisa junto à Associação Mineira de Criadores de Zebu (AMCZ).

Aos amigos do doutorado pelo companheirismo, pelos momentos de estudo e descontração, principalmente Verônica, Ademária, Anderson, Carla, Devanil e Ana Paula.

Ao colega Walmes, que muito auxiliou-me no processo de simulação com suas habilidades.

A todos que, de alguma forma, contribuíram para a realização deste trabalho.

RESUMO

Modelos de regressão não-linear têm se mostrado adequados para descrever curvas de crescimento de animais domésticos de interesse Zootécnico, pois apresentam parâmetros que podem ser interpretados biologicamente. Estes modelos são ajustados a dados de peso-idade por meio de algoritmos iterativos, como o de Gauss-Newton. Um problema frequentemente relatado é não convergência deste algoritmo na presença de oscilações na trajetória esperada da curva, caracterizadas muitas vezes por uma perda de peso abrupta dos animais decorrente da influencia de efeitos ambientais como pouca disponibilidade de forragens e/ou presença de enfermidades. Assim, faz-se necessário o desenvolvimento de procedimentos de estimação que contemplem tal fato, e que de alguma forma considerem a naturalidade da resposta esperada nos experimentos em questão. Em alguns estudos envolvendo ajuste de modelos a dados de dose-resposta, a técnica de regressão isotônica foi usada com sucesso, pois se caracterizou como uma forma apropriada de transformação de dados em situações nas quais as curvas observadas não correspondem as esperadas. O objetivo do presente trabalho foi propor uma metodologia de transformação de dados, via análise de regressão isotônica, para estudos de curvas de crescimento cujos dados apresentam distúrbios caracterizados por decréscimos de pesos em determinadas faixas de idades. Além de investigar a eficiência da metodologia baseada em regressão isotônica em relação ao aumento da convergência e da qualidade do ajuste do modelo, objetivou-se também propor um procedimento iterativo de isotonização cujo intuito foi obter uma transformação ótima para os dados. Todas as metodologias mencionadas foram avaliadas por meio de um estudo de simulação Monte Carlo e aplicadas a dados reais de curvas de crescimento de bovinos da raça Guzerá. Por meio do estudo de simulação de dados foi possível verificar que as metodologias de isotonização adotados resultaram em maiores porcentagens de convergência e menores erros quadráticos médios (EQM) para os parâmetros dos modelos Logísticos, Von Bertalanffy e Gompertz. A aplicação destas metodologias a dados reais de peso-idade de bovinos da raça Guzerá também propiciou bons resultados, sendo estes mais evidentes ao se considerar o ajuste do modelo Logístico, o qual foi o mais adequado para descrever o crescimento dos animais por apresentar alta porcentagem de convergência (100%) e boa qualidade de ajuste, medida por meio do quadrado médio do erro (QME) e do coeficiente de determinação (R^2).

Palavras-chave: Regressão isotônica. Modelo não-linear. Curva de crescimento.

ABSTRACT

Models of non-linear regression have proved adequate to describe growth curves of domestic animals of zootechnical interest, since they have parameters that can be interpreted biologically. These models are fitted to weight-age data by means of iterative algorithms such as Gauss-Newton. A problem often reported is no convergence of this algorithm in the presence of oscillations in the expected path of the curve, widely characterized by an abrupt weight loss of animals due to the influence of environmental effects such as lack nutrients and / or presence of disease. Thus, it is necessary to develop estimation procedures that address this fact, and that somehow consider the naturalness of the expected response in the experiments in question. In some studies involving model adjustments to the dose-response data, the isotonic regression technique was used successfully, as has been characterized as an appropriate way of processing data in situations where the observed curves do not match those expected. The aim of this study was to propose a methodology for processing data via isotonic regression analysis, for studies of growth curves whose data have disorders characterized by decreased body weight at certain ages. Besides investigating the efficacy of the method based on isotonic regression in relation to increased convergence and the quality of the fit of the model aimed to also propose an iterative procedure isotonization whose aim was to obtain an optimal transformation for the data. All mentioned methods were evaluated using a Monte Carlo simulation study and applied to real data of growth curves of bovine Guzera. Through the simulation study data was verified that the methodologies adopted isotonization resulted in higher percentages of convergence and smaller mean squared errors (MSE) for the parameters of the models Logistic, Gompertz and Von Bertalanffy. The application of these methodologies to real data of weight-age Guzera cattle also gave good results, which are more evident when considering the setting Logistic model, which was suitable to describe the growth of animals because of the high percentage Convergence (100%) and good quality of fit, measured by mean square error (MSE) and coefficient of determination (R^2).

Keywords: Isotonic regression. Nonlinear model. Growth curve.

LISTA DE FIGURAS

Figura 1	Gráficos CSD e GCM para o conjunto de dados ilustrativos apresentados por Barlow et al. (1972).....	41
Figura 2	Gráfico da Regressão Isotônica para o conjunto de dados ilustrativos apresentados por Barlow et al. (1972).....	42
Figura 3	Diagrama de somas acumuladas e seus maiores minorantes convexos.....	43
Figura 4	Representação gráfica dos processos de isotonização efetuados.....	50
Figura 5	Representação gráfica do processo iterativo baseado em regressão isotônica com pesos diferentes.....	52
Figura 6	Esquema ilustrativo da inserção da função de distúrbio na curva de crescimento original.....	55
Figura 7	Curvas de crescimento de seis animais geradas via modelo Logístico.....	60
Figura 8	Curvas de crescimento de seis animais geradas via modelo Von Bertalanffy.....	61
Figura 9	Curvas de crescimento de seis animais geradas via modelo Gompertz.....	62
Figura 10	Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Logístico ao longo de 10 iterações consideradas pela metodologia proposta.....	67
Figura 11	Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Von Bertalanffy ao longo de 10 iterações consideradas pela metodologia proposta.....	68
Figura 12	Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Gompertz ao longo de 10 iterações	

	consideradas pela metodologia proposta.....	69
Figura 13	Curvas de crescimento de seis animais da raça Guzerá sob prova de ganho de peso a pasto.....	72
Figura 14	Comportamento do QME para os modelos Logístico, Von Bertalanffy e Gompertz ao longo de 10 iterações consideradas pela metodologia proposta.....	77
Figura 15	Comportamento do R^2 para os modelos Logístico, Von Bertalanffy e Gompertz ao longo de 10 iterações consideradas pela metodologia proposta.....	78

LISTA DE TABELAS

Tabela 1	Propriedades dos modelos não-lineares (y): Brody (1), Gompertz (2), Logístico (3), Richards (4) e Von Bertalanffy (5). Taxa de crescimento instantânea ($\partial y / \partial t$), taxa de maturidade absoluta ($A^{-1}[\partial y / \partial t]$), taxa de crescimento instantânea relativa ($(\partial y / \partial t) / y$) e ponto de inflexão (y_i, t_i).....	21
Tabela 2	Resultado do processo de simulação para o modelo Logístico.....	63
Tabela 3	Resultado do processo de simulação para o modelo Von Bertalanffy.....	63
Tabela 4	Resultado do processo de simulação para o modelo Gompertz.....	64
Tabela 5	Resultado da aplicação das metodologias propostas ao se considerar o ajuste do modelo Logístico a dados de peso-idade de bovinos da raça Guzerá.....	73
Tabela 6	Resultado da aplicação das metodologias propostas ao se considerar o ajuste do modelo Von Bertalanffy a dados de peso-idade de bovinos da raça Guzerá.....	74
Tabela 7	Resultado da aplicação das metodologias propostas ao se considerar o ajuste do modelo Gompertz a dados de peso-idade de bovinos da raça Guzerá.....	74
Tabela 8	Médias e desvios-padrão para as estimativas individuais dos parâmetros do modelo Logístico ajustado a dados de peso-idade de bovinos da raça Guzerá.....	80
Tabela 9	Médias e desvios-padrão para as estimativas individuais	

	dos parâmetros do modelo Von Bertalanffy ajustado a dados de peso-idade de bovinos da raça Guzerá.....	81
Tabela 10	Médias e desvios-padrão para as estimativas individuais dos parâmetros do modelo Gompertz ajustado a dados de peso-idade de bovinos da raça Guzerá.....	82

SUMÁRIO

1	INTRODUÇÃO.....	14
2	REFERENCIAL TEÓRICO.....	17
2.1	Curvas de crescimento.....	17
2.2	Modelos não-lineares.....	18
2.3	Modelos não-lineares para curvas de crescimento.....	19
2.3.1	Interpretação biológica dos parâmetros.....	22
2.4	Problemas de convergência em estudos de curvas de crescimento.....	24
2.5	Regressão isotônica.....	26
2.6	Transformação de dados via regressão isotônica (isotonização).....	28
3	MATERIAL E MÉTODOS.....	30
3.1	Estimação dos parâmetros dos modelos não-lineares.....	30
3.1.1	Método dos quadrados mínimos ordinários.....	30
3.1.2	Processos iterativos e método dos quadrados mínimos.....	32
3.2	Avaliadores da qualidade de ajuste.....	33
3.2.1	Quadrado médio do erro (QME).....	34
3.2.2	Coefficiente de determinação (R^2).....	34
3.2.3	Porcentagem de convergência (%C).....	35
3.3	Regressão isotônica.....	35
3.3.1	Ordenação simples.....	36
3.3.2	O maior minorante convexo.....	39
3.3.2.1	Propriedades do maior minorante convexo.....	42
3.3.3	O algoritmo PAVA.....	45
3.3.4	Redução de erros (A regressão isotônica como um mecanismo para transformação de dados).....	48
3.3.5	Isotonização de dados de curvas de crescimento.....	48
3.3.6	Implementação computacional da regressão isotônica.....	53
3.4	Estudo de simulação de dados.....	53
3.5	Aplicação em dados reais.....	57
4	RESULTADOS E DISCUSSÃO.....	59
4.1	Estudo de simulação de dados.....	59
4.2	Aplicação em dados reais.....	71
5	CONCLUSÕES.....	84
	REFERÊNCIAS	85
	ANEXOS.....	88

1 INTRODUÇÃO

O estudo do crescimento de animais é extremamente importante para a determinação de técnicas adequadas de manejo, bem como para o melhoramento genético, uma vez que possibilita a identificação de indivíduos que crescem com maior eficiência e tal identificação pode garantir, na maioria das vezes, o sucesso da atividade. A tendência geral do crescimento para a maioria das espécies de animais de interesse econômico é uma curva na forma de S estendido, denominada curva de crescimento e esta representa o desenvolvimento dos animais em todas as fases de sua vida.

Modelos de regressão não-linear têm se mostrado adequados para descrever estas curvas, pois apresentam parâmetros que podem ser interpretados biologicamente. Estes são caracterizados principalmente pelo peso na maturidade, que representa o peso na idade adulta e pela velocidade de crescimento, ou taxa de maturidade, que representa uma medida de precocidade. Diferentes modelos não-lineares são apresentados na literatura para descrever curvas de crescimento animal e cada um destes modelos apresenta suas vantagens e desvantagens sob o ponto de vista estatístico. Alguns modelos bastante utilizados são: Logístico, Gompertz, Brody, Von Bertalanffy e Richards.

Por se tratar de modelos não-lineares, geralmente os métodos de estimação caracterizam-se como processos iterativos. Dentre estes métodos, destaca-se o algoritmo de Gauss-Newton, cujo emprego é amplamente utilizado em estudos de curvas de crescimento. Um dos problemas frequentemente encontrados em tais estudos diz respeito à não convergência do algoritmo, que impossibilita a obtenção das estimativas dos parâmetros e, conseqüentemente, a identificação de indivíduos com maior eficiência de crescimento.

De acordo com McManus, Louvandini e Campos (2010), a ausência de convergência é um problema típico de estudos de curvas de crescimento, pois ao analisar curvas de peso e de altura de quatro grupos genéticos de cavalos, verificaram que o modelo Richards não convergiu para nenhuma das variáveis analisadas; que o modelo Logístico não convergiu ao ajustar curvas de peso e o de Gompertz não convergiu no ajuste de altura para a maioria dos grupos considerados.

Ainda em relação a este problema, Rodrigues et al. (2006) utilizaram técnicas de Simulação Monte Carlo e avaliaram os modelos Logístico e Logístico Generalizado em situações envolvendo perdas de observações e concluíram que a região de perda mais danosa para a convergência é ao redor do ponto de inflexão. De forma geral, vale a pena ressaltar que não apenas perdas de observações implicam problemas de convergência. Outro fator importante é a presença de oscilações na trajetória esperada da curva, caracterizadas muitas vezes por uma perda de peso abrupta dos animais, decorrente da influência de efeitos ambientais como falta de nutrientes e/ou presença de enfermidades.

Alterações na forma da curva causadas por fatores externos e caracterizadas por decréscimos do peso comprometem a convergência dos modelos de regressão não-lineares utilizados na descrição dos fenômenos.

Assim, faz-se necessário o desenvolvimento de procedimentos de estimação que contemplem tal fato e que, de alguma forma, considerem a naturalidade da resposta esperada nos experimentos. Entre os procedimentos disponíveis, destaca-se a aplicação da técnica de regressão isotônica como uma transformação apropriada para dados referentes a análises de regressões Probit e Logit.

Sucintamente, o procedimento estatístico conhecido como regressão isotônica -- cuja teoria foi sumarizada por Barlow et al. (1972) -- contempla a ordenação existente na variável resposta no processo de estimação e, para isso,

se utiliza de uma área um tanto complexa da matemática, denominada análise convexa.

Diante do exposto, este trabalho teve por objetivo propor uma metodologia de transformação de dados, via análise de regressão isotônica, para estudos de curvas de crescimento cujos dados apresentam distúrbios caracterizados por decréscimos de pesos em determinadas faixas de idades. Procurou-se avaliar a eficiência da metodologia baseada em regressão isotônica em relação ao aumento da convergência e da qualidade do ajuste do modelo, bem como propor um procedimento iterativo de isotonização com a finalidade de obter uma transformação ótima para os dados.

2 REFERENCIAL TEÓRICO

2.1 Curvas de crescimento

O crescimento de animais tem uma forte relação com a quantidade e a qualidade da carne. Assim, conhecer o processo de ganho de massa corporal do animal é de fundamental importância, pois esse conhecimento possibilita o controle da produção de carne, otimizando os diversos manejos de criação e, conseqüentemente, melhorando lucratividade do empreendimento criatório. Este ganho pode ser influenciado pela alimentação, por condições climáticas, pelo estado sanitário e pelas características genéticas associadas aos animais (GOTTSCHALL, 1999).

Uma forma prática e eficiente de se analisar o crescimento do animal é por meio do estudo de suas curvas de crescimento, as quais descrevem uma relação funcional entre peso e idade (SILVA; AQUINO; OLIVEIRA, 2001). Geralmente, tal relação é representada por modelos de regressão não-lineares, os quais, em sua maioria, apresentam os seguintes parâmetros (SILVA et al., 2004): peso assintótico superior, que representa o peso médio na maturidade, ou peso adulto; taxa de maturidade, que representa a velocidade de crescimento, de forma que quanto mais alto for o seu valor, mais precoce é o animal e vice-versa (BROWN; FITZHUGH JR.; CARTWRIGHT, 1976); ponto de inflexão, que se refere ao momento em que o animal muda de uma fase de crescimento acelerada para uma fase de crescimento inibitória; e por último a constante integração, o qual não apresenta uma interpretação biológica direta.

Segundo Fitzhugh Jr. (1976), os seguintes requisitos devem ser atendidos para que um modelo de regressão não-linear descreva adequadamente a relação peso-idade: interpretação biológica dos parâmetros, “alta qualidade” de ajuste e facilidade de convergência. De forma geral, para que um modelo não-

linear possa ser usado para descrever curvas de crescimento ele deve conter parâmetros cujas interpretações sejam relevantes sob o ponto de vista biológico, devendo apresentar ajustes com pequenos desvios e também proporcionar “altas taxas” de convergência, uma vez que modelos não-lineares requerem métodos iterativos para estimação dos parâmetros.

2.2 Modelos não-lineares

Buscar uma possível relação entre uma variável dependente e uma ou mais variáveis independentes é uma rotina comum quando se realizam análises estatísticas. Isto pode ser feito por meio de modelos de regressão, divididos em em duas classes distintas: lineares e não-lineares. Existem várias diferenças entre estas classes de modelos. Segundo Souza (1998), modelos de regressão não-linear com respostas univariadas y_t são da forma

$$y_t = f(x_t, \theta) + \varepsilon_t, \quad t = 1, \dots, n,$$

em que a função resposta $f(x, \theta)$ tem forma funcional conhecida, x_t é um vetor k dimensional formado por observações de variáveis exógenas, $\theta \in \Theta$ é um vetor de parâmetros p dimensional e ε_t é um erro experimental não observável diretamente.

Draper e Smith (1998) classificaram os modelos como:

a) modelos lineares: aqueles que são lineares em relação aos parâmetros, sendo aqueles em que a derivada da função em relação aos parâmetros não dependem dos parâmetros, $\frac{\partial}{\partial \theta_j} f(X, \theta) = g(X)$, para $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$;

em que n é o número total de observações e p é o número de parâmetros do modelo;

b) modelos linearizáveis: aqueles que podem ser transformados em lineares por meio de uma transformação. Como exemplo, cita-se o modelo $Y = a^X \cdot e$, no qual o erro é multiplicativo e aplicando-se logaritmo a igualdade, tem-se:

$$\begin{aligned}\ln Y &= \ln(a^X \cdot e) \\ \ln Y &= \ln a^X + \ln e \\ \ln Y &= X \ln a + \ln e.\end{aligned}$$

Fazendo $Z = \ln Y$, $b = \ln a$ e $e^* = \ln e$, o modelo pode ser reescrito como $Z = bX + e^*$ e assim verifica-se facilmente que este modelo é linear, pois $\frac{\partial Z}{\partial b} = X = g(X)$.

c) modelos não-lineares: aqueles que não se enquadram nos casos a) e b). Por exemplo, o modelo $Y = a^X + e$, no qual o termo de erro e é aditivo, verifica-se que não existe transformação capaz de torná-lo linear e a derivada da função em relação ao parâmetro depende do parâmetro, pois $\frac{\partial f}{\partial a} = Xa^{X-1} \ln(a) = g(X, a)$.

Outro exemplo de modelo não linear é $Y = a(1 - be^{-Kt}) \cdot e$, onde percebe-se que as derivadas da função em relação aos parâmetros continuam dependendo dos parâmetros.

De modo geral, um modelo é dito não-linear quando ele não é linear em relação aos parâmetros e nem pode ser linearizado por meio de transformações, uma vez que admite uma estrutura de erros aditiva.

2.3 Modelos não-lineares para curvas de crescimento

Segundo Richards (1959 apud FREITAS, 2005), o estudo de curvas de crescimento iniciou-se com o modelo de Von Bertalanffy para estudos metabólicos. Tal modelo é dado por:

$$y = \left\{ \frac{\eta}{k} - \left(\frac{\eta}{k} - y_0^{-m} \right) e^{-(1-m)Kt} \right\}^{\frac{1}{1-m}} \quad (1)$$

em que y é o peso do animal; η e k , as constantes de anabolismo e catabolismo, respectivamente; y_0 é o valor do peso do animal no nascimento. Ainda de acordo com Freitas (2005), o gráfico do modelo (1) converge assintoticamente para $\left(\frac{\eta}{k} \right)^{\frac{1}{1-m}} = A$, correspondendo ao peso limite do animal.

Para $m = 1$, o consumo de oxigênio é proporcional ao peso do animal e para $m = \frac{2}{3}$, o metabolismo do animal é proporcional à área da superfície corporal.

Freitas (2005) explica que o fato de o modelo acima descrito proporcionar resultados irrealistas de η e k para situações em que $m > 1$ ocasionou sua reescrita. Elevando os dois membros a $(1 - m)$, temos:

$$y^{1-m} = A^{1-m} - \beta e^{-\tilde{K}t}, \quad \text{em que } A^{1-m} = \left(\frac{\eta}{k} \right), \quad \beta = \frac{\eta}{k} - y_0^{(1-m)} = A^{1-m} - y_0^{(1-m)}$$

e $\tilde{K} = (1 - m)K$.

Mais tarde, este modelo foi reparametrizado para:

$$y^{1-m} = A^{1-m} (1 \pm b e^{-\tilde{K}t}) \quad (2)$$

em que $b = \pm \beta / A^{1-m}$, sendo o sinal (-) quando $m < 1$ e sinal (+) para $m > 1$.

Richards (1959 apud FREITAS, 2005) classificou os modelos a partir de (2) observando m da seguinte forma:

- Para $0 < m < 1$ tem-se um modelo intermediário entre Brody e Gompertz;
- Para $1 < m < 2$ o formato está situado entre Gompertz e Logístico;

- Para $m > 2$, resulta em um modelo similar ao logístico, porém assimétrico, com ponto de inflexão que pode ser alterado para um valor maior que $y = \frac{A}{2}$.

Na Tabela 1, constam algumas propriedades dos modelos Brody, Gompertz, Logístico, Richards e Von Bertalanffy.

Tabela 1 Propriedades dos modelos não-lineares (y): Brody (1), Gompertz (2), Logístico (3), Richards (4) e Von Bertalanffy (5). Taxa de crescimento instantânea ($\partial y / \partial t$), taxa de maturidade absoluta ($A^{-1}[\partial y / \partial t]$), taxa de crescimento instantânea relativa ($(\partial y / \partial t) / y$) e ponto de inflexão ((y_i, t_i))

Modelo y =	$\partial y / \partial t$	$A^{-1}(\partial y / \partial t)$	$(\partial y / \partial t) / y$	(y_i, t_i)
1. A(1 - bC₁) $\partial y / \partial A = 1 - bC_1$ $\partial y / \partial b = -AC_1$ $\partial y / \partial K = AbtC_1$	$AbKC_1$	bKC_1	$AbKC_1 / y$	não possui
2. Aexp(-bC₁) $\partial y / \partial A = \exp(-bC_1)$ $\partial y / \partial b = -yC_1$ $\partial y / \partial K = ybtC_1$	$bKyC_1$	KyC_1 / A	$bKyC_1$	$A / e; (\log b) / K$
3. A/(1 + bC₁) $\partial y / \partial A = 1 / (1 + bC_1)$ $\partial y / \partial b = -yC_1 / (1 + bC_1)$ $\partial y / \partial K = ybt / (1 + bC_1)C_1$	$ybK / (1 + bC_1)C_1$	$bKC_1 / (1 + bC_1)^2$	$bK / (1 + bC_1)C_1$	$A / 2; (\ln b) / K$
4. A/(1 - bC₁)^m $\partial y / \partial A = (1 - bC_1)^m$ $\partial y / \partial b = -ymC_1 / (1 - bC_1)$ $\partial y / \partial K = ymbtC_1 / (1 - bC_1)$ $\partial y / \partial m = y \ln(1 - bC_1)$	$ymbKC_1 / (1 - bC_1)$	$mbKC_1 / (1 - bC_1)^{m-1}$	$mbKC_1 / (1 - bC_1)$	$Am^{1/(1-m)}; \ln(b(1-m)^{-1}) / K$
5. A(1 - bC₁)³ $\partial y / \partial A = (1 - bC_1)^3$ $\partial y / \partial b = -3AC_1(1 - bC_1)^2$ $\partial y / \partial K = 3^A bt(1 - bC_1)^2 C_1$ $C_1 = \exp(-Kt)$	$3AbKC_1(1 - bC_1)^2$	$3bKC_1(1 - bC_1)^2$	$3ybKC_1(1 - bC_1)$	$8A / 27; \log(3b) / K$

Fonte: Freitas (2005)

2.3.1 Interpretação biológica dos parâmetros

Uma vantagem de utilizarem-se os modelos não-lineares para descrever curvas de crescimento é o fato de alguns parâmetros apresentarem interpretação biológica.

Com relação aos modelos apresentados na Tabela 1, o parâmetro A representa o peso assintótico ou peso máximo. Se o peso adulto do animal não é atingido, este parâmetro reflete o peso nas últimas pesagens.

A constante b não possui interpretação biológica, porém é de suma importância para modelar a curva sigmoideal. Freitas (2005) mostra como a constante b se relaciona com a forma da curva nas seguintes expressões:

- (a) $A(1 - b) \leq y \leq A$ para Brody;
- (b) $A(1 - b)^3 \leq y \leq A$ para Von Bertalanffy;
- (c) $A/(1 - b) \leq y \leq A$ para Logístico;
- (d) $Ae^{-b} \leq y \leq A$ para Gompertz;
- (e) $A(1 - b)^m \leq y \leq A$ para Richards.

Observa-se ainda que, quando $m \rightarrow 1$, como no caso do modelo Gompertz, tem-se $b = \left(\frac{y_0}{A}\right)^{1-m}$. A constante K , ou índice de maturidade, é a razão da taxa de crescimento máxima em relação ao tamanho adulto, que expressa a taxa de redução na taxa de crescimento relativa, em t^{-1} , em que valores menores indicam taxa de crescimento relativa mais rápida.

Além dos parâmetros implícitos nos modelos mencionados, isto é, A , b e K , funções destes parâmetros também podem ser de interesse zootécnico. Dentre estas, Freitas (2005) destaca a taxa de crescimento absoluta instantânea (TCAI), a taxa de crescimento instantânea relativa (TCIR), o ponto de inflexão (PI), a

taxa de maturidade absoluta (TMA) e a taxa de crescimento em um intervalo qualquer (TC). Todas as descrições a seguir foram obtidas de Freitas (2005).

A taxa de crescimento absoluta instantânea (TCAI) estima o incremento no peso para cada unidade de tempo t . Para o modelo geral (2), a TCAI é dada por:

$$\frac{\partial y}{\partial t} = \mp K A b e^{-Kt} (1-m)^{-1} (1 \pm b e^{-Kt})^{\frac{m}{1-m}} \quad (3)$$

A taxa de crescimento instantânea relativa (TCIR) no tempo t em relação ao peso do indivíduo nesse particular tempo t é obtida do modelo geral (2), de acordo com a variação de m , por:

$$y_t^{-1} \left(\frac{\partial y_t}{\partial t} \right) = \mp K b e^{-Kt} [(1-m)(1 \pm b e^{-Kt})]^{-1}, \quad (\mp \text{ para } m > 1; \quad \pm \text{ para } m < 1) \quad (4)$$

O ponto de inflexão (PI) é o ponto em que a TCAI passa de crescente para decrescente, sendo dado pela equação $\frac{y_t}{A} = m^{\frac{1}{1-m}}$. Nota-se que a constante m determina a proporção do peso final (A) no qual o PI ocorre. O modelo Brody não possui ponto de inflexão, pois este é indefinido para $m = 0$. Por isso, é comum utilizar este modelo para descrever o crescimento pós-natal ou fase de auto-inibição do crescimento.

A taxa de maturidade absoluta (TMA) é a razão entre $\frac{\partial y_t}{\partial t}$ e o peso assintótico A , ou seja, $A^{-1}(\partial y_t / \partial t)$. O gráfico $A^{-1}(\partial y_t / \partial t)$ versus y_t / A representa a taxa de troca em peso do tamanho proporcional em relação ao tamanho global, variando na escala de 0 a 1. A área delimitada pela curva é igual a $\frac{k}{2m+2}$ e possibilita aplicações importantes. Quando se ajusta uma curva a um conjunto de animais, m permanece constante e a área depende apenas da

constante K . Assim, por meio da associação de K e o peso limite do animal, é possível identificar na população os animais mais pesados e mais jovens. Da mesma forma, quando diferentes curvas são ajustadas a um conjunto de animais, pode-se estabelecer uma equivalência entre os K 's e comparar estas curvas variando-se apenas o valor de m .

A taxa de crescimento (TC) no intervalo t_i e t_j ($i < j$) para duas avaliações quaisquer i e j no animal, é dada por: $\frac{1}{t_j - t_i} \int_{t_i}^{t_j} \left(\frac{\partial y}{\partial t} \right) \partial t = \frac{y_j - y_i}{t_j - t_i}$,

enquanto a taxa de maturidade absoluta média é

$$\frac{A^{-1}}{t_j - t_i} \int_{t_i}^{t_j} \left(\frac{\partial y}{\partial t} \right) \partial t = A^{-1} \left(\frac{y_j - y_i}{t_j - t_i} \right).$$

2.4 Problemas de convergência em estudos de curvas de crescimento

Para obter as estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não linear, vários métodos iterativos são propostos na literatura. Os mais utilizados são: o método de Gauss-Newton, o método "Steepest-Descent", ou método do Gradiente e o método de Marquardt, que fazem uso das derivadas parciais da função esperança $f(x_i, \theta)$ com relação a cada parâmetro. Outro método iterativo também bastante utilizado é o de Gauss-Newton modificado ou método de linearização, o qual usa resultados de mínimos quadrados lineares em uma sucessão de passos, convertendo o problema de uma regressão não-linear para uma série de regressões lineares. Para tanto, faz-se uma expansão em série de Taylor até o 1º grau e depois minimiza a soma de quadrados residual (OGLIARI, 1998).

Uma vez que algoritmo Gauss-Newton se caracteriza como um processo iterativo, a constatação de sua convergência é de extrema importância, pois se

esta não é atingida não se tem estimativas para os parâmetros dos modelos não-lineares. Às vezes, os softwares estatísticos podem até apresentar estas estimativas na presença de não convergência, porém, nestas situações, tais estimativas são caracterizadas por valores absurdos, ou fora da realidade, visto que o algoritmo converge para o mínimo local e não global.

Em relação à importância da avaliação da convergência, Silveira (2010) relata que em estudos de curvas de crescimento individuais via ajuste de modelos não-lineares o percentual de convergência é um critério relevante, principalmente quando há interesse em utilizar as estimativas dos parâmetros em programas de melhoramento genético animal. Isto porque, se um modelo apresenta qualidade de ajuste muito alta mas percentual de convergência muito baixo, o melhorista terá um número reduzido de animais, caracterizados somente por aqueles na qual a convergência foi verificada, para serem utilizados em programas de seleção para eficiência de crescimento.

Segundo McManus, Louvandini e Campos (2010) a ausência de convergência é um problema sério em estudos de curvas de crescimento, pois ao estudarem curvas de peso e altura de quatro grupos genéticos de cavalos, os autores tiveram problemas de convergência com modelos de três (Logístico e Gompertz) e quatro (Richards) parâmetros. Em um estudo de simulação Monte Carlo e alheios a este fato, Rodrigues et al. (2006) avaliaram os modelos Logístico e Logístico Generalizado em situações envolvendo diferentes números de dados, as quais foram determinadas por perdas de observações em diferentes locais da curva e verificaram que o local de perda mais problemático para a convergência é em torno do ponto de inflexão.

A ausência de convergência não depende somente da perda de observações (MENDES et al., 2009; SILVEIRA, 2010), pois outro aspecto importante é a presença de oscilações, ou variações irrelevantes (HUSSIAN et al., 2004), na trajetória esperada da curva. Estes comportamentos inesperados

são caracterizados principalmente por decréscimos de peso que ocorrem no decorrer da curva e são devidos à influência de efeitos ambientais como falta de nutrientes, presença de enfermidades, toxidez de alimentos, falhas de manejo em geral, dentre outros.

O decréscimo não esperado do peso no decorrer do tempo tem sido observado em diferentes trabalhos envolvendo curvas de crescimento na área de Zootecnia. Dentre eles Mendes et al. (2009), trabalhando com bovinos da raça Hereford criados no sul do país; Silveira (2010), analisando dados de cordeiros cruzados Dorper x Rabo Largo criados na Bahia; e Sousa et al. (2010), com novilhos da raça Guzerá criados em Minas Gerais.

De forma geral, em todos os trabalhos citados acima, embora se esperasse um comportamento natural crescente do peso em função da idade, alterações na forma da curva causadas por fatores externos comprometeram a ocorrência deste comportamento natural e culminaram em problemas de convergência dos modelos de regressão não-lineares. Em outras áreas da Ciência, como Farmacologia, Medicina e Entomologia, a presença de comportamentos inesperados, ou não naturais, principalmente em dados de dose-resposta tem impulsionado pesquisadores a desenvolver técnicas estatísticas específicas, que levam em conta a trajetória esperada da curva no processo de estimação. Dentre estas técnicas destaca-se a regressão isotônica.

2.5 Regressão isotônica

Em várias situações práticas, devido a alguma informação “a priori”, os parâmetros de interesse apresentam algum tipo de ordenação. Assim, deve-se buscar um modelo que preserve esta característica de ordem. O método para encontrar este modelo é denominado Regressão Isotônica (BARLOW et al., 1972), sendo que o termo isotônica (ou monotônica) refere-se ao fato de a

variável resposta aumentar com o aumento da variável independente. Se a variável resposta decresce com o aumento da variável preditora, pode-se usar o termo antitônica.

Segundo Souza (2010), é sabido que alguns modelos de regressão requerem, por pressupostos teóricos, uma ordenação parcial ou total dos números correspondentes à variável dependente. Tal fato ocorre em diversas situações experimentais como em ensaios de dose-resposta, nos quais a variável independente x pode representar a dose de determinado inseticida e a probabilidade de uma resposta (por exemplo, a morte de um inseto) é y . É de se esperar que se a dose x aumenta, a proporção de insetos mortos não diminuirá. De interesse específico deste trabalho, outro exemplo são as curvas de crescimento, nas quais a variável independente x representa o tempo de vida e a variável dependente y representa o peso (ou estatura) do animal. Também se espera que o peso sempre aumente com o tempo até determinado limite biológico e com isso a variável dependente seja sempre crescente. Porém, isso pode não ocorrer na prática, visto que fatores ambientais desconhecidos podem interferir no ganho de peso, provocando, em alguns períodos, instabilidade e/ou perda de peso.

Na análise de regressão usual não se usa uma informação *a priori* sobre uma possível relação de ordem na variável resposta. Na teoria da regressão isotônica essa informação é utilizada de forma essencial. Logo, espera-se que a eficiência do ajuste seja aumentada quando se faz o uso da mesma. A regressão isotônica utiliza esta informação em um contexto bastante geral, considerando até mesmo situações em que a ordenação entre os elementos é parcial (BARLOW et al., 1972),

Tendo em vista os problemas de convergência mencionados no item 2.4, a idéia de se utilizar regressão isotônica decorre do fato de se propor um pré-ajuste dos dados, geralmente caracterizado como uma transformação da variável

dependente. Este processo geralmente é denominado de isotonização (RAMSAY, 1998), ou ainda de normalização, conforme sugestão de Hussian et al. (2004).

2.6 Transformação de dados via regressão isotônica (isotonização)

Em amplo estudo sobre a aplicação de procedimentos de transformação de dados em análises de regressão envolvendo funções monotônicas, Ramsay (1998) descreve uma transformação baseada em regressão isotônica como uma técnica viável para pré-correção das observações e aplicada a diferentes conjuntos de dados, como de crescimento e de acompanhamento de concentração de ozônio. Outra transformação sugerida pelo autor foi baseada em regressão não paramétrica spline.

Xia et al. (2000) modificaram a metodologia de Ramsay para que a mesma fosse utilizada como um método ótimo de transformação de dados em estudos de séries temporais. Os autores aplicaram a metodologia proposta no ajuste de modelos autorregressivos à famosa série de dados denominada *canadian lynx data* (número de lincos observados diariamente entre os anos de 1821 e 1943) e constataram que a transformação sugerida foi mais robusta em relação à influência de dados discrepantes (outliers).

Hussian et al. (2004) utilizaram uma transformação de dados, a qual denominaram normalização, em estudos envolvendo ajuste de modelos de regressão não-linear para descrever a concentração média de nitrogênio em função dos meses e anos observados. Para tanto, utilizaram uma generalização do algoritmo PAVA (pool-adjacent-violators algorithm) e verificaram que tal transformação melhorou o ajuste dos modelos e, principalmente, a interpretação dos parâmetros sobre o ponto de vista biológico. Segundo os autores, esta melhora se deveu ao fato de a transformação adotada ter suprimido variações irrelevantes nos dados coletados.

Várias áreas da medicina utilizam o ajuste de modelos específicos para estudar curvas de interesse. Dentre estas áreas, destaca-se a de reprodução, mais especificamente, a de análise hormonal. Na área em questão, é comum a presença de trajetórias anormais para a concentração de hormônios em função do tempo. Por exemplo, Brumback e Rice (1998) utilizaram regressão spline com coeficientes aleatórios para descrever o comportamento da progesterona durante o ciclo menstrual e a razão para se usar modelos mais complexos é a presença de vários picos decrescentes (vales) de progesterona, comportamento este que se afasta da trajetória esperada e que compromete o ajuste de modelos de regressão usuais. Neste contexto, Gunn e Dunson (2005) aplicaram a transformação baseada em regressão isotônica a dados de concentração de progesterona e conseguiram assim eliminar os distúrbios caracterizados pelos picos decrescentes presentes no conjunto original de dados e assim foi possível utilizar modelos de regressão usuais, os quais foram ajustados pela metodologia bayesiana via “Gibbs Sampler”.

Ao estudar os modelos AMMI (“*additive main effects and multiplicative interaction model*”) utilizados na análise e interpretação da interação genótipo \times ambiente proveniente de um grupo de cultivares de milho e trigo testados em vários ambientes, Araújo (2005) comenta que vários autores verificaram o problema de superestimação e subestimação de autovalores. Este autor constatou que a transformação dos dados via regressão isotônica se mostrou altamente eficiente para contornar o problema em questão.

De forma geral, as significativas melhoras da precisão no processo de estimação relatadas por Xia et al. (2000), Hussian et al. (2004) e Araújo (2005) vão ao encontro dos resultados explicitados pelo teorema apresentado por Robertson, Wright e Dykstra (1988, p. 41), o qual garante uma redução do erro no ajuste de modelos a conjuntos de dados isotonzados.

3 MATERIAL E MÉTODOS

3.1 Estimação dos parâmetros dos modelos não-lineares

Para ajustar os modelos de regressão não-linear Gompertz, Logístico e Von Bertalanffy apresentados na Tabela 1 do item 2.3.1 a dados de crescimento (simulados e reais, conforme será apresentado nos 3.4 e 3.5) utilizou-se o método dos quadrados mínimos ordinários, cujas soluções foram obtidas por meio do processo iterativo de Gauss-Newton.

3.1.1 Método dos quadrados mínimos ordinários

Considerando a função $f(\mathbf{x}, \boldsymbol{\theta})$ como uma generalização para qualquer um dos modelos (Gompertz, Logístico e Von Bertalanffy) mencionados no item anterior e adotando a notação matricial, tem-se:

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}^0) + \boldsymbol{\varepsilon},$$

em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad f(\mathbf{x}, \boldsymbol{\theta}^0) = \begin{bmatrix} f(\mathbf{x}, \theta_1^0) \\ f(\mathbf{x}, \theta_2^0) \\ \vdots \\ f(\mathbf{x}, \theta_n^0) \end{bmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \text{sendo } \boldsymbol{\theta}^0 \text{ um vetor de valores}$$

iniciais para os parâmetros a serem estimados.

A soma dos quadrados dos erros aleatórios (SQE) deverá ser minimizada por $\boldsymbol{\theta}$, portanto a função de mínimos quadrados para um modelo não-linear é dada por:

$$\text{SQE}(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2,$$

a qual pode ser representada matricialmente por:

$$\text{SQE}(\boldsymbol{\theta}) = [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]' [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]$$

Segundo Souza (1998), em modelos não-lineares não se pode fazer afirmações gerais sobre as propriedades dos estimadores de quadrados mínimos, tais como não tendenciosidade e variância mínima, exceto para grandes amostras, os chamados resultados assintóticos. Para uma melhor compreensão do processo de obtenção destes estimadores, utilizou-se a seguinte notação de diferenciação matricial:

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f_1(\boldsymbol{\theta}) \\ f_2(\boldsymbol{\theta}) \\ \vdots \\ f_n(\boldsymbol{\theta}) \end{bmatrix}, \text{ e } \mathbf{F}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix},$$

em que:

$\mathbf{f}(\boldsymbol{\theta})$ é uma função vetor coluna $n \times 1$ de um argumento p dimensional $\boldsymbol{\theta}$, e $\mathbf{F}(\boldsymbol{\theta})$ é a matriz Jacobiana de $\mathbf{f}(\boldsymbol{\theta})$. Dessa forma, o estimador de mínimos

quadrados, $\hat{\boldsymbol{\theta}}$, satisfaz a equação $\left. \frac{\partial \text{SQE}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$, a qual representa a

minimização de interesse. Sendo,

$$\frac{\partial \text{SQE}(\theta)}{\partial \theta'} = \frac{\partial}{\partial \theta'} [y - f(\theta)]' [y - f(\theta)] = -2 [y - f(\theta)]' F(\theta), \quad \text{tem-se:}$$

$$F'(\hat{\theta}) [y - f(\hat{\theta})] = 0.$$

Portanto, o sistema de equações normais (SEN) é dado por:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_p} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_p} \end{bmatrix} \cdot \begin{pmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\theta}) \\ f_2(\hat{\theta}) \\ \vdots \\ f_n(\hat{\theta}) \end{bmatrix} \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

3.1.2 Processos iterativos e método dos quadrados mínimos

Para o SEN não-linear acima não existe uma solução explícita, sendo assim a solução para o sistema deve ser obtida por meio de processos iterativos. Um dos métodos iterativos é a linearização da função não-linear, chamado Método de Gauss-Newton, que se resume ao seguinte procedimento.

Seja o modelo não-linear $y_i = f(x_i, \theta) + \varepsilon_i$ e $\hat{\theta}_0$ um valor tal que $F'(\hat{\theta}_0) [Y - f(\hat{\theta}_0)] \approx 0$. Aproximando $f(\hat{\theta})$ pelo ponto $\hat{\theta}_0$ por uma TSA (*Taylor Series Expansion*) de 1ª ordem, tem-se:

$$f(\hat{\theta}) \approx f(\hat{\theta}_0) + F(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0) \quad (5)$$

$$F'(\hat{\theta})[Y - f(\hat{\theta})] \approx 0 \quad (6)$$

Aplicando (5) em (6): $F'(\hat{\theta})[Y - f(\hat{\theta}_0) - F(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0)] \approx 0$ e

multiplicando à esquerda, ambos os lados da igualdade, por $[F'(\hat{\theta})]^{-1}$, obtém-se:

$$Y - f(\hat{\theta}_0) - F(\hat{\theta}_0)\hat{\theta} + F(\hat{\theta}_0)\hat{\theta}_0 \approx 0.$$

Logo, $F(\hat{\theta}_0)\hat{\theta} \approx F(\hat{\theta}_0)\hat{\theta}_0 + [Y - f(\hat{\theta}_0)]$. Multiplicando novamente à esquerda,

ambos os lados da igualdade, por $[F(\hat{\theta}_0)]^{-1}$, verifica-se que:

$$\hat{\theta} \approx \hat{\theta}_0 + [F(\hat{\theta}_0)]^{-1}[Y - f(\hat{\theta}_0)].$$

Fazendo $\hat{\theta} = \hat{\theta}_{k+1}$ e $\hat{\theta}_0 = \hat{\theta}_k$, tem-se para a k-ésima iteração, a expressão (7), a qual representa o processo iterativo conhecido como Gauss-Newton:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + [F(\hat{\theta}_k)]^{-1}[Y - f(\hat{\theta}_k)] \quad (7)$$

Este processo iterativo prossegue até que algum critério adotado para convergência seja atingido.

3.2 Avaliadores da qualidade de ajuste

Uma vez obtidas as estimativas dos parâmetros dos modelos de regressão não-linear por meio dos procedimentos apresentados nos itens 3.1.1 e 3.1.2, o próximo passo é a comparação dos modelos ajustados, de forma que seja possível indicar aquele que melhor representou o comportamento dos dados de crescimento analisados. Tal comparação é realizada por meio de avaliadores da qualidade de ajuste de modelos de regressão não-linear.

Segundo Silveira (2010), na literatura especializada, a maioria dos trabalhos envolvendo comparação de modelos de crescimento utiliza apenas o Quadrado Médio do Erro (QME) e o coeficiente de determinação (R^2) como avaliadores da qualidade de ajuste, e isto pode ser explicado pelo fato destes avaliadores serem apresentados diretamente nos arquivos de resultados de softwares estatísticos usuais como SAS[®] e R. Além destes dois avaliadores, especialmente em situações nas quais se tem ajustes individuais, a porcentagem de convergência (%C) configura-se como um importante avaliador, uma vez que quantifica a complexidade do modelo de regressão não-linear. A relevância da %C já foi discutida no item 2.4. Este trabalho também optou por estes avaliadores -- QME, R^2 e %C --, que são apresentados a seguir.

3.2.1 Quadrado médio do erro (QME)

Este avaliador simplesmente expressa a variância residual proveniente do ajuste do modelo considerado e, logicamente, comparando os vários modelos, quanto menor o valor do QME, mais adequado é o modelo em questão. Sua expressão é dada por (SOUZA, 1998, p. 347):

$$QME = \frac{SQE(\hat{\theta})}{n - p},$$

em que: $SQE(\hat{\theta})$ é a soma de quadrados do erro apresentada no item 3.1.1, n é o número de observações e p o número de parâmetros do modelo considerado.

3.2.2 Coeficiente de determinação (R^2)

De forma geral, o R^2 expressa a proporção da variação total observada nos dados amostrais que foi explicada pelo modelo de regressão adotado. Sua

expressão pode ser apresentada de forma simplificada por (SOUZA, 1998, p. 360):

$$R^2 = 1 - \frac{SQE(\hat{\theta})}{SQT},$$

em que: SQT é soma de quadrados total corrigida pela média, ou seja:

$$SQT = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

3.2.3 Porcentagem de convergência (%C)

É usado quando se tem ajustes individuais, ou seja, o interesse reside em ajustar uma curva para dados de cada animal. Sendo assim, é possível observar qual modelo apresenta maior facilidade de convergência, que é dado pela porcentagem de ajustes que convergiram.

3.3 Regressão isotônica

A regressão isotônica geralmente é feita em dois tipos de conjuntos de dados: os conjuntos com uma ordenação simples das observações e os conjuntos com uma quase ordenação das observações. Embora esta última seja mais geral (SOUZA, 2010), é muito mais complexa. Portanto, no presente trabalho, serão consideradas apenas as teorias de regressão isotônica fundamentadas na ordenação simples.

3.3.1 Ordenação simples

Seja $X = \{x_1 \prec x_2 \prec \dots \prec x_k\}$ um conjunto finito, onde a relação binária (\prec) entre os elementos de X é uma relação de ordem simples se as seguintes propriedades forem satisfeitas:

1. é reflexiva: $x \prec x$ para todo $x \in X$;
2. é transitiva: $x, y, z \in X, x \prec y$ e $y \prec z$ então $x \prec z$;
3. é anti-simétrica: $x, y \in X, x \prec y$ e $y \prec x$ então $x = y$; e
4. todo e qualquer elemento de X é comparável: $x, y \in X$, implica que, $x \prec y$ ou $y \prec x$.

Conjuntos simplesmente ordenados podem ser interpretados geometricamente como os elementos de uma fila.

Seja $X = \{x_1 \prec x_2 \prec \dots \prec x_k\}$ um conjunto finito com uma ordem simples. Uma função $f : X \rightarrow R$ é isotônica se $f(x_1) \leq f(x_2) \leq \dots \leq f(x_k)$, sendo que (\prec) indica a relação existente no conjunto da variável independente e (\leq) no conjunto da variável dependente.

Para o caso da ordenação simples, a regressão isotônica é definida como segue.

Definição 1. Seja g uma função em X . Uma função g^* é uma regressão isotônica de g com peso ω se e somente se g^* for uma função isotônica e minimiza

$$\sum_{x \in X} [g(x) - f(x)]^2 \omega(x) \quad (8)$$

onde f varia entre todas as funções isotônicas em X . Portanto, g^* é uma solução de mínimos quadrados restritos para a expressão (8).

Para compreender melhor a definição 1, seja $X = \{x_1, x_2, \dots, x_k\}$, onde $x_1 < x_2 < \dots < x_k$. Para $x_i \in X$, tomemos $y_j(x_i)$, com $j = 1, 2, \dots, m(x_i)$, uma amostra i.i.d. de uma distribuição com média desconhecida $\mu(x_i)$. Assumindo-se que $\mu(x)$ é uma função linear de x , uma forma de se estimar este parâmetro é através da regressão linear, que consiste em minimizar $\sum_{x \in X} \sum_{j=1}^{m(x)} [y_j(x) - f(x)]^2$, com $f(x)$ variando na classe das funções lineares f .

Seja $\bar{y}(x) = \frac{1}{m(x)} \sum_{j=1}^{m(x)} y_j(x)$, $x \in X$. Uma vez que

$$\sum_{j=1}^{m(x)} [y_j(x) - f(x)]^2 = \sum_{j=1}^{m(x)} [y_j(x) - \bar{y}(x)]^2 + m(x) [\bar{y}(x) - f(x)]^2$$

um problema equivalente é minimizar

$$\sum_{x \in X} [\bar{y}(x) - f(x)]^2 m(x) \quad (9)$$

na classe de funções lineares f de X .

Se não existe restrição para μ , a estimativa de mínimos quadrados poderia ser obtida através da minimização de (9) na classe de funções arbitrárias f de X e a solução é a média amostral \bar{y} . Agora, se se conhece ou se assume que μ é não decrescente em x ; isto é, é isotônica em relação a uma ordem simples de X , a estimativa de mínimos quadrados de μ pode ser obtida por minimização da equação (9) ponderada pelos pesos, na classe de funções não decrescentes f de X , a classe de funções isotônicas com relação à ordem simples de X . A solução é denominada *regressão isotônica simples*.

Uma vez apresentada de forma detalhada a definição 1, suponha agora que sejam tomadas amostras aleatórias de p populações normais com médias

$\mu(x_1), \mu(x_2), \dots, \mu(x_p)$ e variância comum σ^2 . Para a i -ésima amostra, seja n_i o tamanho amostral e sejam $Y_{i1}, Y_{i2}, \dots, Y_{in}$ as observações, de forma que $\bar{Y}_i = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i}$ é o estimador de máxima verossimilhança (EMV), sem restrição para o vetor de médias populacionais $\mu = (\mu(x_1), \mu(x_2), \dots, \mu(x_p))$. Suponha também que as médias populacionais são não-decrescentes, isto é, $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Pode-se desejar satisfazer esta restrição e ainda estimar estas médias, mas devido à variabilidade amostral, a amostra das médias pode não estar ordenada de forma crescente.

Tomando-se o negativo do logaritmo da função de verossimilhança, é possível obter as estimativas que satisfazem à restrição de ordem. Dessa forma,

maximizar a função de verossimilhança $\prod_{i=1}^p \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_{ij}-\mu_i)^2}{2\sigma^2}}$ é o mesmo

que minimizar o termo

$$\sum_{i=1}^p \sum_{j=1}^{n_i} [Y_{ij} - \mu(x_i)]^2 \quad (10)$$

sujeito a $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Somando e subtraindo \bar{Y}_i em (10), tem-se

$\sum_{i=1}^p \sum_{j=1}^{n_i} [(Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \mu(x_i))]^2$, que equivale a:

$$\sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + 2 \sum_{i=1}^p \sum_{j=1}^{n_i} [(Y_{ij} - \bar{Y}_i)(\bar{Y}_i - \mu(x_i))] + \sum_{i=1}^p (\bar{Y}_i - \mu(x_i))^2 n_i$$

Observa-se que a primeira parcela desta soma não envolve $\mu(x_i)$, ainda que a segunda parcela seja nula. Assim, o estimador de máxima-verossimilhança

(EMV) restrito minimiza $\sum_{i=1}^p (\bar{Y}_i - \mu(x_i))^2 n_i$, sujeito à condição $\mu(x_1) \leq \mu(x_2) \leq \dots \leq \mu(x_p)$. Então, em um conjunto finito $X = \{x_1, x_2, \dots, x_p\}$, com $x_1 \prec x_2 \prec \dots \prec x_p$, onde g e ω são definidas em X por $g(x_i) = Y_i$ e $\omega(x_i) = n_i$, fica claro que o EMV restrito de $\mu = (\mu(x_1), \mu(x_2), \dots, \mu(x_p))$ é dado pela regressão isotônica μ^* de \bar{Y} com peso ω .

Faz-se necessário apresentar uma abordagem geométrica da teoria da regressão isotônica, úteis e essenciais à obtenção de algoritmo para seu cálculo.

3.3.2 O maior minorante convexo

Para obter uma interpretação gráfica da regressão isotônica e considerando a ordenação simples $x_1 \prec x_2 \prec \dots \prec x_k$, deve-se plotar os pontos

$P_j = (W_j, G_j)$, $j = 0, 1, 2, \dots, k$, em que $G_j = \sum_{i=1}^j g(x_i) \omega(x_i)$ é a soma

acumulada da função g ponderada pelos pesos ω e $W_j = \sum_{i=1}^j \omega(x_i)$ é a soma acumulada dos pesos.

O gráfico dos pontos $P_j = (W_j, G_j)$, $j = 0, 1, 2, \dots, k$, ($P_0 = (0, 0)$) constitui o diagrama de somas acumuladas (CSD) da função dada g com pesos ω . O declive do segmento que une os pontos P_{j-1} e P_j é justamente $g(x_j)$, $j = 1, 2, \dots, k$.

O maior minorante convexo (GCM) é formado por segmentos de reta, ligando o ponto P_{j-1}^* ao ponto P_j de modo que todos os pontos do CSD fiquem

acima ou no próprio GCM. Graficamente, o GCM é o maior diagrama convexo, porém menor do que o CSD.

O valor da regressão isotônica g^* em um ponto x_j é a inclinação do GCM no ponto P_j^* com abscissa $W_j = \sum_{i=1}^j \omega(x_i)$, onde P_j^* é um ponto do GCM ou é a projeção do ponto P_j sobre o GCM.

Para um maior entendimento dos conceitos CSD e GCM apresentados, considere o conjunto ilustrativo de dados conforme apresentado por Barlow et al. (1972):

j	$\omega(x_j)$	W_j	$g(x_j)$	G_j	G_j^*	g_j^*
1	1	1	-2	-2	-2	-2
2	2	3	5/2	3	-8/5	1/5
3	3	6	-4/3	-1	-1	1/5
4	2	8	1	1	1	1

em que: $W_j = \sum_{i=1}^j \omega(x_i)$, $G_j = \sum_{i=1}^j g(x_i)\omega(x_i)$, $G_j^* = \sum_{i=1}^j g^*(x_i)\omega(x_i)$ e $j = 1, 2, 3, 4$.

Sabe-se que a inclinação em P_j de CSD é dada por $\frac{G_j - G_{j-1}}{W_j - W_{j-1}} = g(x_j)$

e que

a inclinação em P_j^* de GCM é $\frac{G_j^* - G_{j-1}^*}{W_j - W_{j-1}} = g^*(x_j)$.

Como a regressão isotônica é a inclinação do ponto no GCM, podemos calcular:

$$g^*(x_1) = \frac{-2-0}{1-0} = -2.$$

Os pontos P_2 e P_3 pertencem à mesma reta. Logo, seus coeficientes angulares são iguais: $g^*(x_2) = g^*(x_3) = \frac{-1-(-2)}{6-1} = \frac{1}{5}$ e

$$g^*(x_4) = \frac{1-(-1)}{8-6} = 1.$$

Na Figura 1, são apresentados os gráficos CSD e GCM para a situação ilustrada.

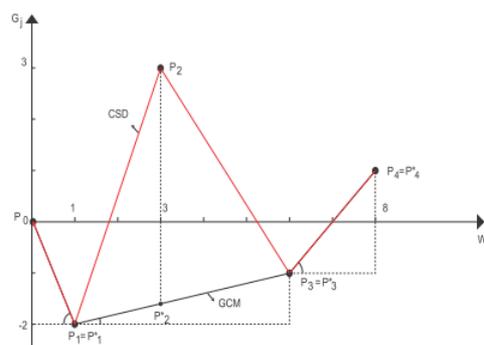


Figura 1 Gráficos CSD e GCM para o conjunto de dados ilustrativos apresentados por Barlow et al. (1972)

A partir do diagrama apresentado na Figura 1 é possível construir o gráfico da regressão isotônica, que é apresentado na Figura 2.

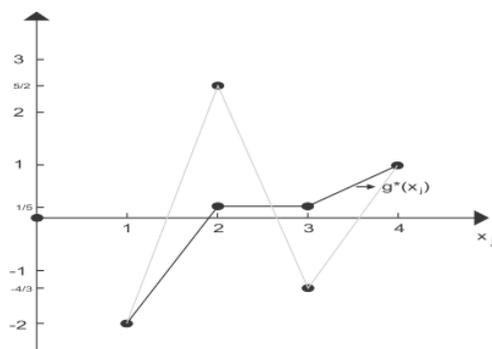


Figura 2 Gráfico da Regressão Isotônica para o conjunto de dados ilustrativos apresentados por Barlow et al. (1972)

3.3.2.1 Propriedades do maior minorante convexo

Será útil expressar analiticamente algumas propriedades do maior minorante convexo.

(i) Os pontos P_k do GCM e CSD coincidem, isto é, $G_k^* = G_k$;

(ii) Se para algum índice i do GCM o ponto P_{i-1}^* estiver estritamente abaixo do CSD em relação a P_{i-1} , então o ponto P_{i-1}^* admite a mesma inclinação no GCM em relação ao ponto à sua esquerda e à sua direita. Isto é: $G_{i-1}^* < G_{i-1} \Rightarrow g_i^* - g_{i-1}^* = 0$, para $i = 1, 2, \dots, k$.

(iii) Se $P_r = P_r^*$, $P_s = P_s^*$, se $P_r P_s = P_r^* P_s^*$, o segmento de reta unindo os pontos P_r e P_s , constituindo um lado do GCM e $P_s P_t = P_s^* P_t^*$ o lado adjacente à direita e se $r < j < s$, então, algebricamente, a inclinação de $P_j P_s$ é menor do que a de $P_j^* P_s^*$. Se $s < j < t$, então a inclinação de $P_s P_j$ é maior do

que a de $P_s^* P_j^*$. Isto é, se $g^*(x)$ tem um valor constante a para $x_r < x \leq x_s$ e o valor constante $b > a$ para $x_s < x \leq x_r$, então

$$\frac{\sum_{i=j+1}^s g(x_i) \varpi(x_i)}{\sum_{i=j+1}^s \varpi(x_i)} = \frac{G(x_s) - G(x_j)}{W(x_s) - W(x_j)} \leq a \quad \text{para } r \leq j < s, \text{ e}$$

$$\frac{\sum_{i=s+1}^j g(x_i) \varpi(x_i)}{\sum_{i=s+1}^j \varpi(x_i)} = \frac{G(x_j) - G(x_s)}{W(x_j) - W(x_s)} \geq b \quad \text{para } s < j \leq s.$$

Em particular, $g(x_s) \leq g^*(x_s) = a < b = g^*(x_{s+1}) < g(x_{s+1})$.

A Figura 3 ilustra a interpretação geométrica desta última propriedade (iii).

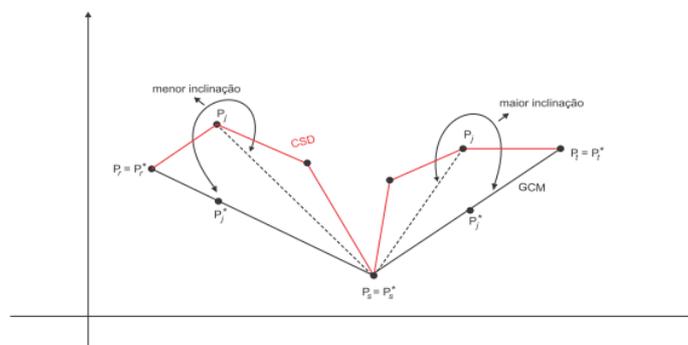


Figura 3 Diagrama de somas acumuladas e seus maiores minorantes convexas

De forma geral, o diagrama mostrado na Figura 3 apresenta grande utilidade para o desenvolvimento do Teorema 1, que se configura como um dos mais importantes da teoria de regressão isotônica.

Teorema 1. Se X é um conjunto com uma ordenação simples, a inclinação $g^*(x_i)$ de GCM fornece a regressão isotônica de g . De fato, se f é isotônica em X , então

$$\sum_x [g(x) - f(x)]^2 \omega(x) \geq \sum_x [g(x) - g^*(x)]^2 \omega(x) + \sum_x [g^*(x) - f(x)]^2 \omega(x) \quad (11)$$

Assim, o teorema em questão preconiza que a regressão isotônica é única.

Prova Teorema 1: Desenvolvendo os dois lados da desigualdade (11), tem-se:

$$\begin{aligned} \sum_x [g(x)]^2 \omega(x) - 2 \sum_x [g(x)f(x)]\omega(x) + \sum_x [f(x)]^2 \omega(x) &\geq \sum_x [g(x)]^2 \omega(x) - \\ &- 2 \sum_x [g(x)g^*(x)]\omega(x) + \sum_x [g^*(x)]^2 \omega(x) + \\ &+ \sum_x [g^*(x)]^2 - 2 \sum_x [g^*(x)f(x)]\omega(x) + \sum_x [f(x)]^2 \omega(x), \end{aligned}$$

subtraindo agora o segundo termo da inequação (11) nos dois lados da desigualdade, a expressão resultante é dada por:

$$\begin{aligned} -2 \sum_x [g(x)f(x)]\omega(x) + 2 \sum_x [g(x)g^*(x)]\omega(x) - 2 \sum_x [g^*(x)]^2 \omega(x) + 2 \sum_x [g^*(x)f(x)]\omega(x) &\geq 0 \\ \therefore 2 \sum_x [g(x) - g^*(x)][g^*(x) - f(x)]\omega(x) &\geq 0. \end{aligned}$$

Uma vez que na expressão anterior o 2 é uma constante positiva, a inequação acima pode ainda ser escrita como:

$$\sum_x [g(x) - g^*(x)][g^*(x) - f(x)]\omega(x) \geq 0, \quad (12)$$

para toda função isotônica f em X . Assim, a soma parcial da expressão (12) produz

$$\sum_{i=1}^k [g(x_i) - g^*(x_i)][g^*(x_i) - f(x_i)]\omega(x) = \sum_{i=1}^k \{[f(x_i) - f(x_{i-1})] -$$

$$[g^*(x_i) - g^*(x_{i-1})] \times [G_{i-1} - G_{i-1}^*] \omega(x_i) + [g^*(x_k) - f(x_k)] [G_k - G_k^*],$$

em que: $x_0 = f(x_0) = g^*(x_0) = G_0 = G_0^* = 0$.

Da propriedade (i) do GCM, temos que o último termo da direita é zero e de (ii) temos que $[g^*(x_i) - g^*(x_{i-1})] [G_{i-1} - G_{i-1}^*] = 0$, para $i = 1, 2, \dots, k$. Uma vez que $G_{i-1}^* \leq G_{i-1}$ e $f(x_{i-1}) \leq f(x_i)$, a desigualdade (12) prevalece. Para verificarmos a unicidade da regressão isotônica, suponhamos que g_1 é uma outra regressão isotônica em X tal que $\sum_x [g(x) - g_1(x)]^2 \omega(x) \leq \sum_x [g(x) - f(x)]^2 \omega(x)$ para cada função isotônica f em X . Portanto, g_1 e g^* são ambas isotônicas. Logo,

$$\sum_x [g(x) - g_1(x)]^2 \omega(x) = \sum_x [g(x) - g^*(x)]^2 \omega(x).$$

Mas, de (11), desde que g_1 seja uma isotônica, segue-se que:

$$\sum_x [g(x) - g_1(x)]^2 \omega(x) \geq \sum_x [g(x) - g^*(x)] \omega(x) + \sum_x [g^*(x) - g_1(x)] \omega(x),$$

fato este que implica em: $\sum_x [g^*(x) - g_1(x)] \omega(x) = 0$. Dessa forma,

$g^*(x) = g_1(x)$, para $x \in X$.

Um algoritmo eficiente para o cálculo da regressão isotônica, obtido a partir das propriedades geométricas do GCM é o seguinte.

3.3.3 O algoritmo PAVA

Um algoritmo amplamente utilizado para calcular a regressão isotônica para uma ordem simples é o PAVA (*pool-adjacent-violators algorithm*). Robertson, Wrigth e Dykstra (1988 apud ARAÚJO, 2005), Ayer et al. (1955)

foram os primeiros a apresentar o algoritmo em questão e a descreverem o processo de isotonização que se segue.

O algoritmo inicia-se com a função $g(x)$ e, se $g(x)$ é isotônica, então $g(x) = g^*(x)$. Caso contrário deve existir algum índice i tal que $g(x_{i-1}) > g(x_i)$, no qual $g(x_{i-1})$ é denominado violador. Estes dois valores são substituídos pelas suas médias ponderadas, as quais são denotadas por $Av(i-1, i)$, isto é:

$$Av(i-1, i) = \frac{g(x_{i-1})\omega(x_{i-1}) + g(x_i)\omega(x_i)}{\omega(x_{i-1}) + \omega(x_i)}. \quad (13)$$

Após o cálculo desta média ponderada, os pesos $\omega(x_{i-1})$ e $\omega(x_i)$ serão substituídos pela soma $\omega(x_{i-1}) + \omega(x_i)$ e assim subsequentemente. Se esse for um novo conjunto, com $p-1$ valores isotônicos, então $g^*(x_{i-1}) = g^*(x_i) = (Av(i-1, i))$ e $g^*(x_j) = g(x_j)$, com $j = 1, 2, \dots, k-1$. Caso contrário, esse processo é repetido até que se obtenha um conjunto com valores isotônicos, sendo que o valor de $g^*(x_i)$ é a média ponderada do bloco em que o violador está contido.

Para ilustrar a aplicação do algoritmo PAVA, Bhattacharyya e Klotz (1966 apud BARLOW et al., 1972) usaram dados anuais referentes ao número de dias a partir de uma data fixa para o Lago Mendota (Madison-WI, EUA) começar a congelar. Os dados foram coletados a partir de 23 de novembro de 1854 e esta coleta foi realizada durante 111 anos. O estudo teve como finalidade analisar a hipótese de tendência de aquecimento e, para tanto, os autores adotaram um modelo simples com y_i observações (ou seja, dados observados a cada ano) independentes de uma distribuição normal com médias desconhecidas μ_i , $i = 1, 2, \dots, 111$ e variância comum σ^2 . Desta feita, para construir o teste da

razão de verossimilhança com o intuito de testar nenhuma tendência de aquecimento (H_0) contra a presença da mesma (H_1) adotou-se:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_{111}$$

$$H_1 : \mu_1 \leq \mu_2 \leq \dots \leq \mu_{111}.$$

Tendo em vista a necessidade de se obter as estimativas de máxima verossimilhança para as médias sob H_0 e H_1 , o estimador de máxima verossimilhança foi definido pela regressão isotônica μ^* de $\hat{\mu}$. Uma vez aplicado o algoritmo PAVA para os dez primeiros anos, tem-se os seguintes resultados:

x	1	2	3	4	5	6	7	8	9	10
Pesos	1	1	1	1	1	1	1	1	1	1
$g(x)$	25	13	2	15	14	21	9	33	25	15
$g^*(x)$	13,3		14,5			15,0		24,3		

em que x é o ano e $g(x)$ é o número de dias que o lago levou para congelar.

Observando os valores de $g(x)$, temos que $g(1) > g(2)$ logo detecta-se o primeiro violador, dessa forma a média ponderada destes dois valores é:

$$Av(1,2) = \frac{g(1)\omega(1) + g(2)\omega(2)}{\omega(1) + \omega(2)} = \frac{25 \times 1 + 13 \times 1}{1 + 1} = 19,$$

Portanto, este bloco é substituído pelo valor médio 19. Continuando a análise, detecta-se outro violador, pois $19 > 2$ e repete-se o processo novamente:

$$\frac{19 \times 2 + 2 \times 1}{2 + 1} = \frac{40}{3} = 13,3.$$

De acordo com o exemplo ilustrativo apresentado, nota-se que o objetivo do algoritmo PAVA é detectar violadores e substituir os blocos por suas médias ponderadas, até que se obtenha um conjunto de valores isotônicos.

Assim, tem-se que $g^*(1) = g^*(2) = g^*(3) = 13,3$; $g^*(4) = g^*(5) = 14,5$; $g^*(6) = g^*(7) = 15$ e $g^*(8) = g^*(9) = g^*(10) = 24,3$.

3.3.4 Redução de erros (A regressão isotônica como um mecanismo para transformação de dados)

Robertson, Wright e Dykstra (1988) mostram que a isotonização dos dados pode reduzir os erros associados a processos de estimação em vários aspectos, fornecendo evidências da melhora da qualidade de ajuste. Tal afirmação é justificada pelo seguinte teorema apresentado pelos autores.

Teorema 2. Seja $X = \{x_1, \dots, x_n\}$ e $\theta: X \rightarrow R$ uma função isotônica. Seja ainda $\hat{\theta}$ um estimador qualquer desta função isotônica θ . Se $\hat{\theta}^*$ é a regressão isotônica de $\hat{\theta}$, então $\sum_{x \in X} (\theta(x) - \hat{\theta}^*(x))^2 \omega(x) \leq \sum_{x \in X} (\theta(x) - \hat{\theta}(x))^2 \omega(x)$.

A prova deste teorema encontra-se em Robertson, Wright e Dykstra (1988, p. 41). Baseado neste resultado pode-se propor a isotonização como forma de melhorar a qualidade do ajuste de modelos aos dados de crescimento.

3.3.5 Isotonização de dados de curvas de crescimento

Visando à definição de isotonização como uma transformação de dados baseada em regressão isotônica no item 2.6 e uma vez realçada sua eficiência para tratar dados com distúrbios (comportamentos não naturais das variáveis dependentes em relação ao esperado), esta pesquisa propõe a aplicação de tal técnica na solução de problemas de convergência em estudos de curvas de crescimento animal via modelos de regressão não linear. Assim, pretende-se

corrigir os dados originais de pesos tomados em uma determinada idade, menores que aqueles tomados em idades precedentes, ou seja, pretende-se torná-los não decrescentes.

Para o desenvolvimento da proposta relatada, considere a observação y_i referente ao peso de um animal tomado em um tempo x_i , sendo $i=1,2,\dots,n$. Considere também uma função $f(\mathbf{x}, \boldsymbol{\theta})$ que generaliza um modelo de crescimento para todas as expressões dos modelos Gompertz, Logístico e Von Bertalanffy apresentados na Tabela 1 do item 2.3.1. Assim, o ajuste destes modelos ao conjunto de dados originais de peso-idade (\mathbf{y}), sendo que $\mathbf{y} = [y_1, y_2, \dots, y_n]$, por meio do método dos Quadrados Mínimos Ordinários consiste em assumir $\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, em que $\mathbf{x} = [x_1, x_2, \dots, x_n]$ e $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]$, $\varepsilon_i \sim N(0, \sigma^2)$.

A primeira proposta de isotonização diz respeito à aplicação da regressão isotônica com pesos iguais. Esta primeira proposta transforma os valores originais de pesos y_i em valores y_i^* , sendo estes últimos crescentes, considerando-se, subsequentemente, cada data de avaliação. Dessa forma, os modelos de crescimento foram ajustados ao conjunto de dados $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_n^*]$. Portanto, assumiu-se que $\mathbf{y}^* = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$.

A segunda proposta de isotonização considera a aplicação da regressão isotônica com pesos diferentes, isto é, cujos pesos $\omega(y_i)$ são diferentes. Neste contexto, os pesos adotados no presente trabalho foram dados por $\omega(y_i) = 1/|y_i - y_i^*|$, se $y_i \neq y_i^*$. Esta definição garante que os valores transformados via regressão isotônica, agora denotados por y_i^{**} , sejam não decrescentes, pois quanto menor o valor $\omega(y_i)$, mais intensa é a correção

exercida. Assim, os modelos de crescimento foram ajustados ao conjunto de dados $\mathbf{y}^{**} = [y_1^{**}, y_2^{**}, \dots, y_n^{**}]$ e, dessa forma, assumiu-se que $\mathbf{y}^{**} = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$.

As três situações mencionadas como os dados originais com distúrbios, transformados via regressão isotônica com pesos iguais e via regressão isotônica com pesos diferentes podem ser facilmente visualizadas na Figura 4.

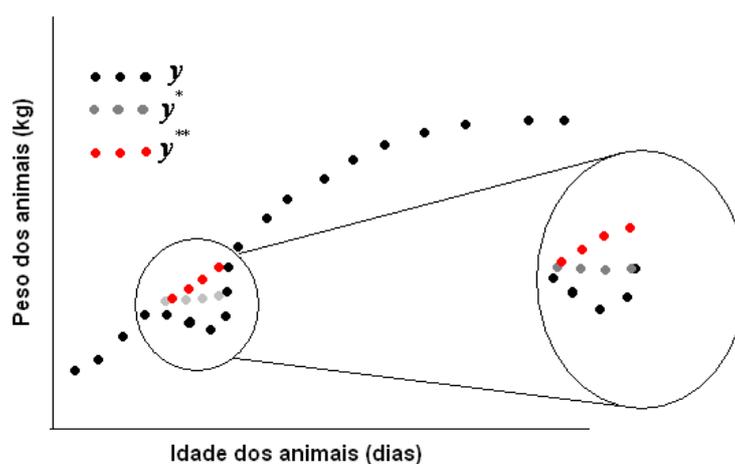


Figura 4 Representação gráfica dos processos de isotonização efetuados

Percebe-se na figura 4 que os dados originais apresentaram um distúrbio caracterizado por uma oscilação na trajetória. Percebe-se ainda que por meio da regressão isotônica com pesos iguais houve uma correção nesta trajetória e utilizando-se uma regressão isotônica com pesos diferentes, houve uma correção maior como uma tentativa de recuperar a trajetória esperada da curva.

Para ilustrar melhor o efeito de se variar os pesos, considere os dados da tabela abaixo, que contém dados fictícios, sobre os quais foram aplicadas técnicas de regressão isotônica com pesos iguais e com pesos diferentes.

Dados originais	3	3	3	2	1	2
Reg. Isot. (pesos iguais)	2,33	2,33	2,33	2,33	2,33	2,33
Pesos	1,5	1,5	1,5	3	0,75	3
Reg. Isot. (pesos diferentes)	2,33	2,33	2,33	2,375	2,5833	2,7916

Analisando esta tabela, é fácil perceber que ao dado de menor magnitude (1) foi associado o menor peso (0,75), o que acarretou uma maior correção.

Além das propostas mencionadas anteriormente, o presente trabalho também teve como objetivo propor um procedimento iterativo baseado em regressão isotônica cuja meta foi a obtenção de uma transformação ótima para os dados. Tal procedimento leva em consideração sucessivas aplicações da regressão isotônica com pesos diferentes, cujos pesos em cada iteração (repetição) k são definidos por $\omega(y_i) = 1/|y_i - y_i^{**(k)}|$, para $y_i \neq y_i^{**(k)}$. Vale ressaltar que k representa as iterações computadas a partir da primeira isotonização ponderada (y^{**}), sendo $k=1,2,\dots,m$. Ao adotar o peso em questão, preconiza-se o fato de que a cada iteração um peso diferente é considerado e este se altera até que diferenças entre o peso computado em uma iteração k e uma iteração $k+1$ não se verifiquem. A Figura 5 tem por objetivo elucidar este processo iterativo proposto.

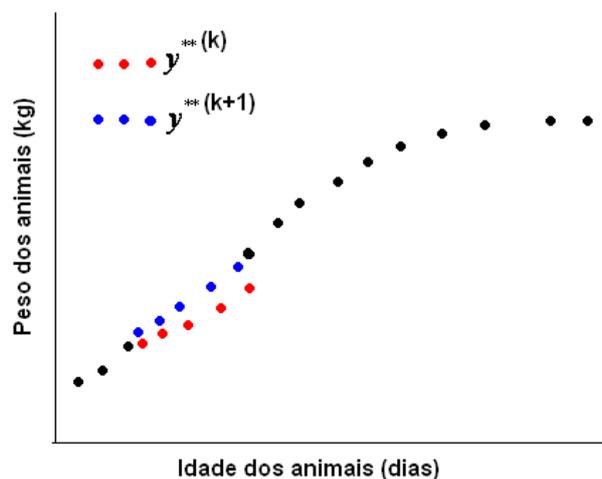


Figura 5 Representação gráfica do processo iterativo baseado em regressão isotônica com pesos diferentes

Da mesma forma que apresentado anteriormente para a primeira e segunda proposta de isotonização, assume-se que a cada iteração k , tem-se um novo conjunto de dados transformados $\mathbf{y}^{**}(k) = [y_1^{**}(k), y_2^{**}(k), \dots, y_n^{**}(k)]$ e, portanto, os modelos de crescimento considerados foram ajustados a cada um destes conjuntos por meio da suposição $\mathbf{y}^{**}(k) = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$. Dessa maneira, em se tratando de estudos de simulação de dados, é possível avaliar a eficiência deste procedimento iterativo por meio do cálculo do Erro Quadrático Médio (EQM) para cada parâmetro em cada uma das iterações, sendo a referida eficiência verificada pela redução do EQM ao longo das iterações. Por outro lado, em se tratando de dados reais, tal eficiência pode ser analisada por meio da redução do QME e do R^2 com o decorrer do aumento das iterações.

É importante comentar que a elaboração desta última proposta está fundamentada no Teorema apresentado por Robertson, Wright e Dykstra (1988,

p. 41), que garante uma redução do erro de ajuste de modelos a dados tratados, ou transformados, mediante regressão isotônica. Com base nesta idéia, o objetivo da proposta em questão é exercer subsequentes aplicações da regressão isotônica com pesos diferentes a um conjunto de dados até que a melhora causada por tais transformações não seja mais significativa, culminando em uma transformação otimizada e, conseqüentemente, proporcionando a maior redução do erro no ajuste do modelo de interesse aos dados sob tal transformação.

3.3.6 Implementação computacional da regressão isotônica

A regressão isotônica pode ser implementada em diferentes softwares estatísticos, entre os quais o software livre R (R DEVELOPMENT CORE TEAM, 2009) que apresenta tal implementação de forma mais simples e compacta, sendo, por este motivo, utilizado no presente trabalho (anexo A). Neste software, o pacote requerido para a consideração de pesos iguais é o *isotone* e a função usada foi a *isoreg*. Já para a consideração de pesos diferentes, o pacote requerido foi o *cir* e a função usada foi a *cir.pava*.

No software SAS[®], a regressão isotônica pode ser implementada via PROC NLP. Porém, esta implementação não se apresenta de forma simples e prática como aquela mostrada pelo software R.

3.4 Estudo de simulação de dados

Visando avaliar a eficiência da regressão isotônica para corrigir distúrbios em curvas de crescimento animal, foi realizado um estudo de simulação de dados via técnica Monte Carlo. O estudo de simulação em questão consistiu em gerar valores de y_i por meio da adoção de valores fixos para os parâmetros dos modelos (Gompertz, Logístico e Von Bertalanffy). Foram geradas vinte e uma observações longitudinais, sendo $x_i = 0, 30, 60, \dots, 600$,

pois, na prática, em experimentos envolvendo curvas de crescimento animal geralmente as pesagens são obtidas mensalmente. Para cada observação y_i foi gerado um valor aleatório relacionado ao termo de erro, o qual foi especificado como $e_i \sim N(0,3)$. Ao todo, foram simuladas 1000 curvas de crescimento diferentes, sendo estas diferenças garantidas pelo número aleatório gerado para o erro. Em termos práticos, estas 1000 curvas podem corresponder a curvas de 1000 animais diferentes pertencentes a um mesmo rebanho.

Para inserir distúrbios na curva de crescimento, a fim de que a mesma não apresentasse a trajetória crescente sigmóide esperada, primeiro calculou-se a derivada discreta para cada ponto de cada curva simulada, ou seja,

$$y'_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}.$$

Tomando-se o valor máximo da derivada discreta de cada curva, o valor de x_i correspondente a este ponto de máximo representa o tempo no qual ocorre o ponto de inflexão da curva.

Tal procedimento foi adotado visto que distúrbios de crescimento geralmente ocorrem em torno do ponto de inflexão, pois de acordo com Gottschall (1999) é onde o animal cresce com maior velocidade e, conseqüentemente, demanda maior qualidade de manejo (geralmente maior demanda nutricional). Assim, quando o manejo nesta fase de crescimento não é adequado, o animal pode reduzir seu peso abruptamente, pois suas necessidades diárias de alimentos não são atingidas justamente quando ele mais precisa. De forma geral, a redução de peso que caracteriza os distúrbios de crescimento já mencionados se deve a este descontrole nutricional.

Além do aspecto prático apresentado acima, o ponto de inflexão também está relacionado aos aspectos estatísticos da análise de curva de crescimento, uma vez que perdas de observações ou presença de observações discrepantes em torno deste ponto prejudicam drasticamente o ajuste de curvas de crescimento (RODRIGUES et al., 2006).

No estudo de simulação e de acordo com a já relatada importância do ponto de inflexão, uma função de distúrbio foi inserida, tendo como referência este ponto. Para tanto, foi utilizada a seguinte função paramétrica:

$$f(\mu, \alpha, \beta, x) = -\beta \cdot \exp\left\{\frac{-(x - \mu)^2}{\alpha}\right\},$$

em que: μ é o parâmetro que controla a posição, coincidente com o valor do tempo no qual ocorre o ponto de inflexão, sendo α o parâmetro que controla a extensão e β o parâmetro que controla a intensidade do distúrbio.

Dessa forma, a função de distúrbio foi simplesmente somada à curva de crescimento original, fato que culminou com o surgimento de um distúrbio, conforme apresentado na Figura 6.

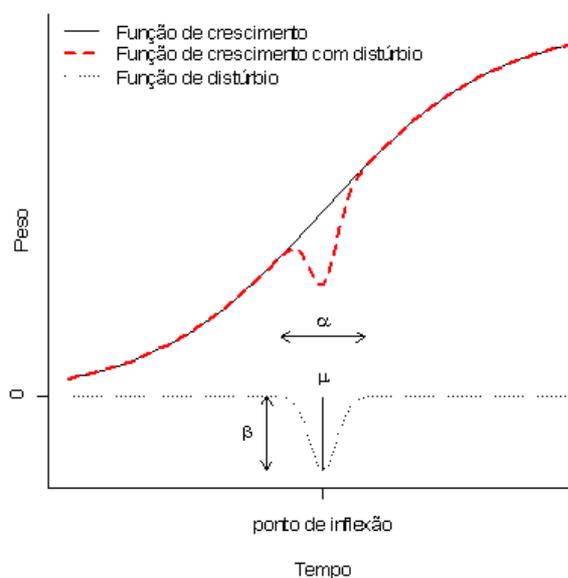


Figura 6 Esquema ilustrativo da inserção da função de distúrbio na curva de crescimento original

Na sequência, buscou-se corrigir este distúrbio através do uso da regressão isotônica com pesos iguais por meio da função *isoreg* do pacote *isotone* do software R. Após esta primeira isotonização, definiram-se os pesos da regressão isotônica com pesos diferentes como sendo o inverso do módulo da diferença entre os dados originais e os dados corrigidos pela regressão isotônica com pesos iguais. Uma vez obtidos tais pesos, realizou-se uma correção dos dados originais via regressão isotônica com pesos diferentes e por meio da função *cir.pava* do pacote *cir* do software R. Vale ressaltar que os procedimentos de isotonização foram efetuados separadamente para cada uma das 1000 curvas de crescimento geradas.

Após a especificação das três situações a serem consideradas (dados originais com distúrbio, dados corrigidos pela regressão isotônica com pesos iguais e dados corrigidos pela regressão isotônica com pesos diferentes), realizaram-se ajustes de cada um dos três modelos adotados (Gompertz, Logístico e Von Bertalanffy) para cada uma das três situações citadas a fim de determinar a eficiência dos procedimentos de isotonização em relação à porcentagem de convergência. Estes ajustes foram realizados por meio da função *nls* do R, que estimou os parâmetros mediante método dos mínimos quadrados ordinários via algoritmo de Gauss-Newton. Nos ajustes em questão, os valores iniciais requeridos para se iniciar o processo iterativo de estimação foram os mesmos valores dos parâmetros usados para simular os dados de peso-idade.

Além da porcentagem de convergência, foi ainda calculado o Erro Quadrático Médio (EQM) para cada um dos parâmetros dos três modelos em todas as três situações acima mencionadas. Considerando $j=1,2,\dots,1000$ repetições da simulação, as expressões do EQM para os parâmetros A, b e K foram dadas respectivamente por:

$$\text{EQM}(A) = \frac{\sum_{j=1}^{1000} (A - \hat{A}_j)^2}{1000}, \quad \text{EQM}(b) = \frac{\sum_{j=1}^{1000} (b - \hat{b}_j)^2}{1000} \quad \text{e}$$

$$\text{EQM}(K) = \frac{\sum_{j=1}^{1000} (K - \hat{K}_j)^2}{1000}, \text{ em que } A, b \text{ e } K \text{ foram os valores paramétricos}$$

considerados na simulação das curvas.

Tendo em vista a abordagem iterativa apresentada no item 3.3.4, que consistiu em aplicar subsequentemente várias regressões isotônicas ponderadas, é importante comentar que a eficiência da metodologia proposta também foi verificada pelo cálculo do EQM para cada parâmetro. Porém, neste contexto, os EQM's foram calculados em cada uma das iterações consideradas. No presente trabalho, em razão do custo computacional envolvido, assumiu-se o número máximo de apenas 10 iterações. Conforme já comentado no item citado, a eficiência deste método pode ser facilmente verificada por um decréscimo do EQM em cada iteração.

3.5 Aplicação em dados reais

Os dados utilizados se referem às observações de peso-idade de animais da raça Guzerá, cedidos pela Associação Mineira de Criadores de Zebu (AMCZ) com sede em Curvelo – MG. Tais dados foram obtidos em uma prova de ganho de peso (PGP), realizada na fazenda Meleiro, situada no município de Curvelo - MG. As avaliações dos animais foram realizadas obedecendo à metodologia das PGP definidas pela Associação Brasileira dos Criadores de Zebu (ABCZ).

Foram avaliados 45 machos da raça Guzerá recém-desmamados, com peso médio inicial de 219,9 kg e desvios padrão de 38,05 kg e idade média de 325,8 dias com desvio padrão de 28,0 dias. A PGP foi basicamente desenvolvida

com os animais consumindo pastagens de *Brachiaria brizantha* e suplementação múltipla, além de receberem o tratamento sanitário adequado, onde os animais recém desmamados são mantidos em condições uniformes de manejo e alimentação por um período de 294 dias. Durante as estações da seca e das águas, os animais receberam sal proteinado e vermifugação habitual. Os animais foram pesados na entrada da PGP no dia 27 de junho de 2009 e, após um período de adaptação de 70 dias, foram então pesados novamente dando início ao teste de performance propriamente dito, sendo que as avaliações ocorreram a cada 56 dias durante os 224 dias de prova. Assim, foram consideradas seis pesagens por animal: uma inicial e outras cinco decorrentes da prova de ganho peso.

Os modelos Logístico, Von Bertalanffy e Gompertz (Tabela 1) foram ajustados aos dados de peso-idade por meio da função *nls* do R, mediante método dos mínimos quadrados ordinários e via algoritmo iterativo de Gauss-Newton. Nestes ajustes, os valores iniciais requeridos para dar início ao processo iterativo de estimação foram obtidos de Sousa et al. (2010), quem também analisaram tais dados sob o enfoque de curvas de crescimento via ajuste de modelos de regressão não linear.

4 RESULTADOS E DISCUSSÃO

4.1 Estudo de simulação de dados

A proposta de isotonização de dados referentes a estudos envolvendo curvas de crescimento animal -- cujo objetivo principal é melhorar a porcentagem de convergência (%C) e a qualidade dos ajustes em situações que consideram curvas atípicas relacionadas à perda de peso dos animais -- foi avaliada por meio da simulação Monte Carlo. As três situações consideradas (dados originais com distúrbio, dados corrigidos via regressão isotônica com pesos iguais e via regressão isotônica com pesos diferentes) nesta simulação podem ser visualizadas nas Figuras 7, 8 e 9, as quais retratam, respectivamente, as curvas de crescimento geradas por meio dos modelos Logístico, Von Bertalanffy e Gompertz para seis indivíduos aleatoriamente amostrados da população base de 1000 indivíduos.

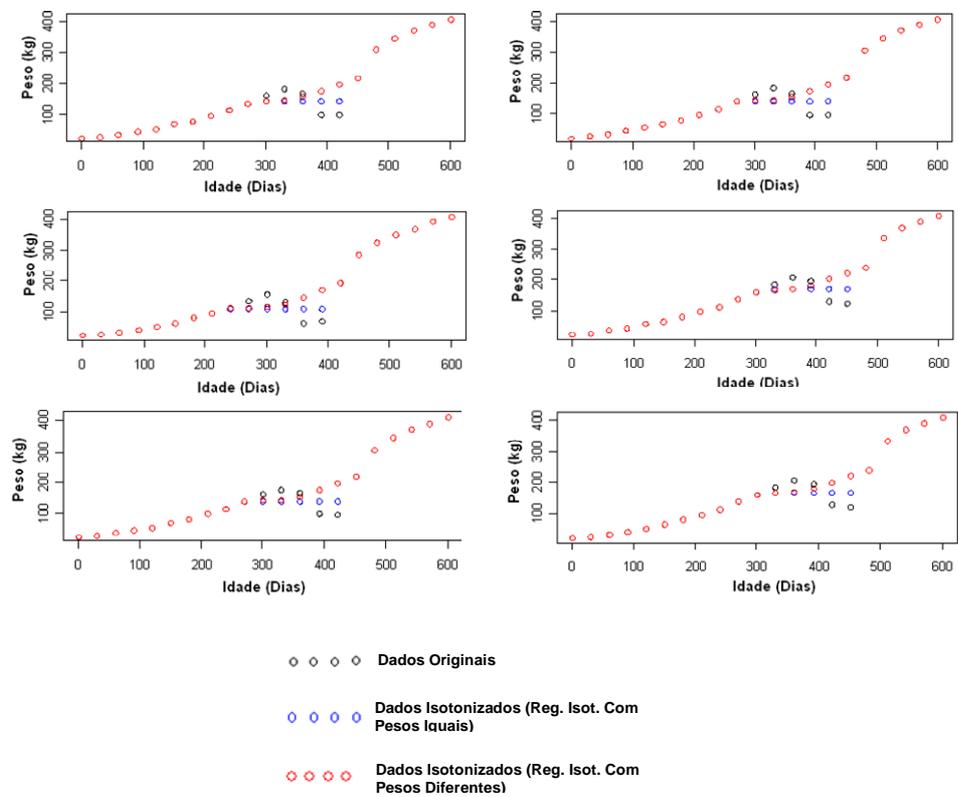


Figura 7 Curvas de crescimento de seis animais geradas via modelo Logístico

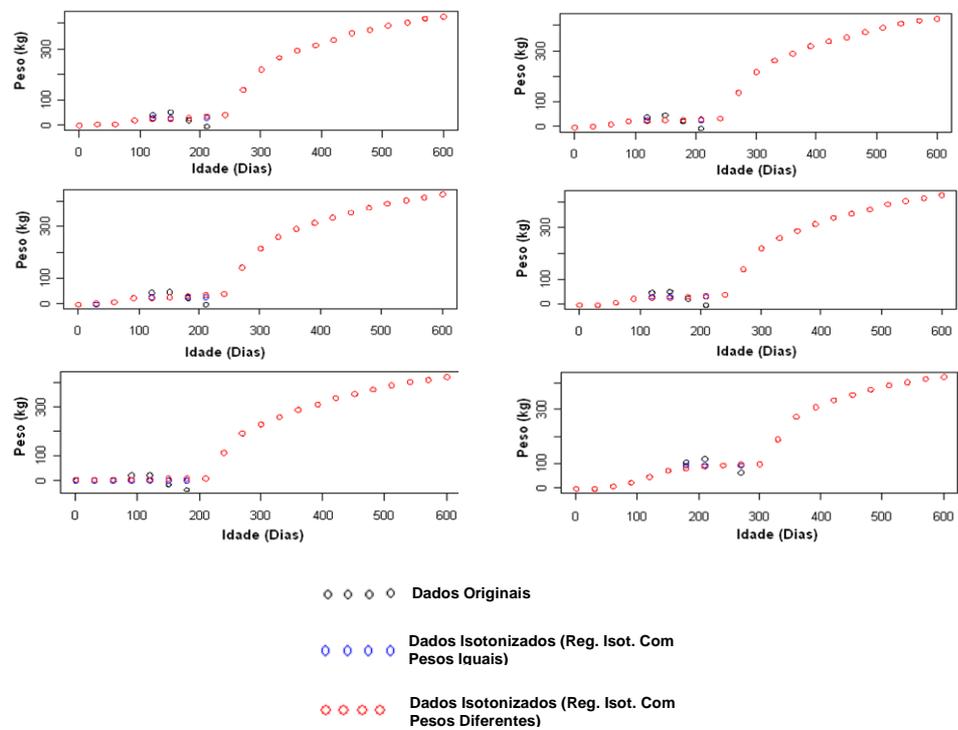


Figura 8 Curvas de crescimento de seis animais geradas via modelo Von Bertalanffy

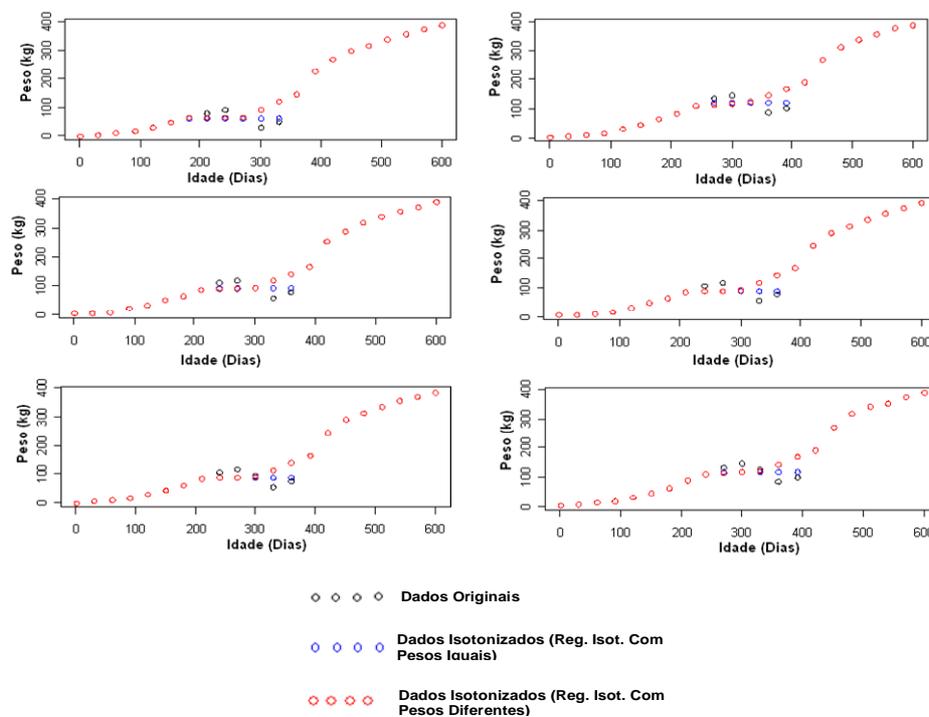


Figura 9 Curvas de crescimento de seis animais geradas via modelo Gompertz

Os gráficos mostrados nas Figuras 7, 8 e 9 explicitam a idéia de transformação de dados via regressão isotônica, permitindo, de certa forma, recuperar a trajetória natural da curva representativa do fenômeno em estudo que, neste caso, trata-se do crescimento de animais de interesse Zootécnico. Propostas como a de Hussian et al. (2004) – que envolvem o ajuste de modelos de regressão não-linear para descrever a concentração média de nitrogênio em função do tempo, ou como a de Gunn e Dunson (2005), referente ao ajuste de modelos de regressão para descrever a concentração de progesterona no decorrer

de cada ciclo menstrual – já foram aplicadas com sucesso na literatura estatística.

O estudo de simulação de dados providenciou importantes resultados em relação à proposta de se utilizar a isotonização de dados em estudos de curvas de crescimento animal. As Tabelas 2, 3 e 4 mostram tais resultados para os modelos Logístico, Bertalanffy e Gompertz, respectivamente. Nestas tabelas, as notações referentes aos conjuntos de dados para os quais os modelos foram ajustados respeitam as definições apresentadas no item 3.3.4.

Tabela 2 Resultado do processo de simulação para o modelo Logístico

Situação (transformação dos dados)	% C	EQM (A)	EQM (b)	EQM (K)
Originais (y)	50	529786605,6	426037,2492	8,170361e-06
Reg. Isot. Pesos Iguais (y^*)	50	2192547,1	1959,7055	5,725936e-06
Reg. Isot. Pesos Diferentes (y^{**})	100	539594,4	265,4507	5,994183e-06

Tabela 3 Resultado do processo de simulação para o modelo Von Bertalanffy

Situação (transformação dos dados)	% C	EQM (A)	EQM (b)	EQM (K)
Originais (y)	80	12552,609	0,1001136	7,243446e-07
Reg. Isot. Pesos Iguais (y^*)	100	7568,218	0,1116972	4,646744e-07
Reg. Isot. Pesos Diferentes (y^{**})	100	5849,840	0,1117990	3,311908e-07

Tabela 4 Resultado do processo de simulação para o modelo Gompertz

Situação (transformação dos dados)	% C	EQM (A)	EQM (b)	EQM (K)
Originais (y)	50	1311510,1	1,5368543	5,110558e-06
Reg. Isot. Pesos Iguais (y [*])	80	2625149,8	0,0943173	7,153625e-06
Reg. Isot. Pesos Diferentes (y ^{**})	100	406188,3	0,3113374	4,321749e-06

No âmbito geral, os resultados mostrados nas Tabelas 2, 3 e 4 sugerem que o procedimento de isotonização dos dados originais de crescimento na presença de alterações não naturais -- aqui definidas como distúrbios -- mostrou-se eficiente em relação ao aumento da %C para todos os modelos estudados. Tal eficiência foi evidente para a transformação baseada na regressão isotônica com pesos diferentes, a qual propiciou uma %C de 100% para todos os três modelos adotados.

Nota-se que a maior eficácia da isotonização foi constatada quando os modelos Logístico (Tabela 2) e Gompertz (Tabela 4) foram utilizados, pois a %C foi apenas de 50% para estes dois modelos, considerando-se os dados originais. Do ponto de vista prático, aplicando-se estes modelos na descrição de curvas de crescimento com distúrbios referentes a animais de um determinado rebanho, obter-se-iam informações de eficiência de crescimento por meio da interpretação biológica dos parâmetros apenas para a metade dos animais.

Tendo em vista os resultados da regressão isotônica com pesos iguais, nota-se que esta foi eficiente quando se considerou os modelos Von Bertalanffy (Tabela 3) e Gompertz (Tabela 4), uma vez que as %C foram de 100 e 80%,

respectivamente. Porém, esta forma simplificada de isotonização não conduziu a bons resultados quando o modelo Logístico foi usado, observando-se apenas 50% de %C, a mesma porcentagem apresentada pelo ajuste, considerando os dados originais.

Em resumo, os resultados das Tabelas 2, 3 e 4 permitem inferir que a hipótese levantada sobre a melhora da %C via isotonização dos dados realmente foi observada, pois o processo de suavização proveniente da aplicação de regressão isotônica, simples e ponderada possibilitou ajustar os três modelos considerados a um maior número de animais. Este fato pode melhorar consideravelmente a qualidade de pesquisas na área de Zootecnia relacionadas à análise de curvas de crescimento.

De acordo com Silveira (2010), na área de melhoramento genético animal, este aumento na qualidade das pesquisas diz respeito à utilização das estimativas dos parâmetros A (peso adulto) e K (taxa de maturidade) como critério de seleção de indivíduos com maior eficiência de crescimento. Sendo assim, quanto maior o número de animais com suas estimativas dos parâmetros A e K, maior a probabilidade de serem encontrados indivíduos geneticamente superiores.

Nas Tabelas 2, 3 e 4, é possível observar uma melhora da precisão na estimação dos parâmetros dos modelos de crescimento ao se usar a isotonização, observação já verificada e relatada por outros autores (XIA et al., 2000; HUSSIAN et al., 2004, ARAÚJO, 2005).

Para todos os modelos adotados, é possível notar que o EQM para o parâmetro A (peso adulto) sofreu uma redução drástica em função da ordem da aplicação dos três conjuntos diferentes de dados, que são: dados originais, dados transformados por regressão isotônica com pesos iguais e dados transformados por regressão isotônica com pesos diferentes, evidenciando a melhoria do ajuste dos modelos nos dados isotonizados.

Quanto ao EQM do parâmetro b (parâmetro de integração), resultados semelhantes àqueles obtidos para o parâmetro A foram observados quando o modelo Logístico foi considerado (Tabela 2). Porém, tal afirmação não pode ser reafirmada para os demais modelos devido à constatação de diferenças em níveis decimais.

Também foi possível verificar uma expressiva melhora na precisão da estimação do parâmetro K com o uso da regressão isotônica com pesos diferentes em relação ao uso dos dados originais e àqueles provenientes da aplicação da regressão isotônica com pesos iguais. A utilização dos dados corrigidos pela regressão isotônica com pesos iguais foi mais eficiente que a utilização dos dados originais, considerados os modelos Logístico e Von Bertalanffy. No entanto, a mesma observação não pôde ser verificada em relação ao modelo Gompertz.

Na avaliação de um procedimento iterativo baseado em regressão isotônica cuja meta foi a obtenção de uma transformação ótima para os dados, conforme Robertson, Wright e Dykstra (1988), as Figuras 10, 11 e 12 mostram a eficiência do procedimento proposto mediante observação do EQM para os parâmetros dos modelos Logístico, Von Bertalanffy e Gompertz.

De forma geral, em todas as figuras apresentadas (10, 11 e 12), observa-se um comportamento decrescente do EQM para todos os parâmetros em todos os modelos adotados em função do aumento do número de iterações. É importante realçar que a estabilização do EQM é clara e permite inferir que o total de 10 iterações realmente foi suficiente para que se obtivesse uma transformação ótima dos dados, baseada em sucessivas aplicações da regressão isotônica com pesos diferentes para os modelos Logístico (Figura 10) e Von Bertalanffy (Figura 11). Porém, para o modelo Gompertz, esta estabilização não foi facilmente observada, exigindo, conseqüentemente, um maior número de iterações ao se trabalhar com este modelo.

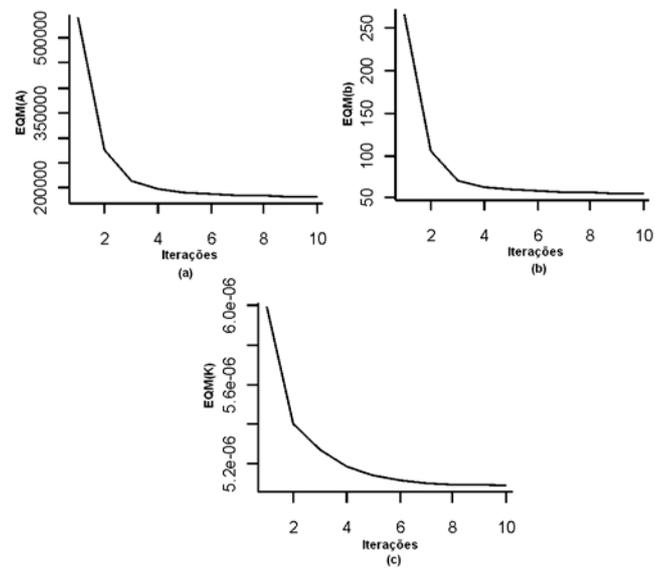


Figura 10 Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Logístico ao longo de 10 iterações consideradas pela metodologia proposta

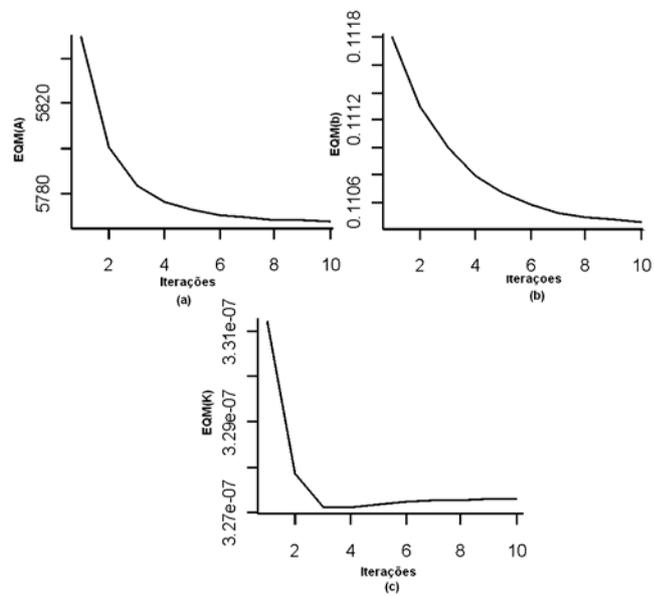


Figura 11 Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Von Bertalanffy ao longo de 10 iterações consideradas pela metodologia proposta

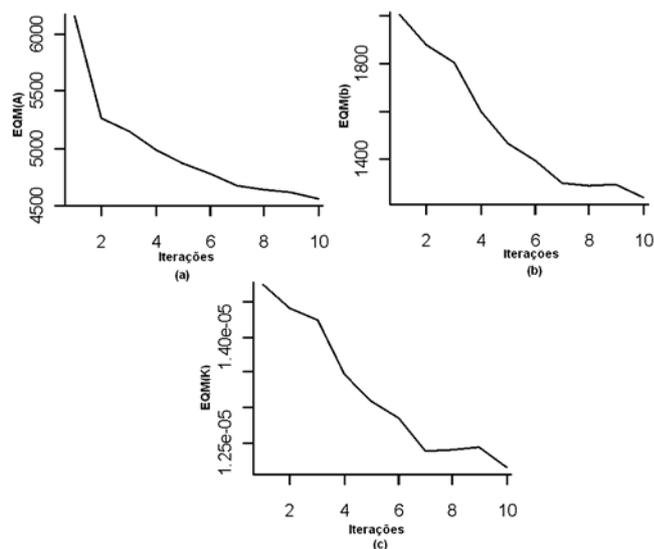


Figura 12 Comportamento do EQM para os parâmetros A (a), b (b) e K (c) do modelo Gompertz ao longo de 10 iterações consideradas pela metodologia proposta

A discussão acima traz à tona o Teorema apresentado por Robertson, Wright e Dykstra (1988, p. 41), o qual preconiza um decréscimo para o EQM ao se aplicar a regressão isotônica em relação à sua não aplicação. Tal comportamento pode ser evidenciado no processo iterativo proposto no presente trabalho que possibilitou constatar e avaliar este decréscimo em sucessivas isotonizações.

Além disso, o estudo de simulação empregado foi útil para verificar que a redução do EQM -- e, conseqüentemente, o aumento da qualidade da estimação -- não é constante para diferentes modelos e nem para diferentes parâmetros destes modelos. Assim, as diferenças observadas nos gráficos das Figuras 10, 11 e 12 podem ser melhor exploradas e discutidas como se segue.

Observa-se que para os parâmetros A, Figuras 10 (a), 11 (a) e 12 (a) e b; Figuras 10 (b), 11 (b) e 12 (b), o EQM se estabiliza primeiro para o modelo Logístico, depois para o modelo Von Bertalanffy e, por último, para o modelo Gompertz, cuja estabilização não é evidente. Para o parâmetro K, o comportamento do EQM é um pouco diferente, uma vez que a estabilização é primeiramente verificada para o modelo Von Bertalanffy, seguido pelo modelo Logístico e pelo modelo Gompertz, cuja estabilização do EQM também não é evidenciada.

Em termos da avaliação da magnitude do EQM, menores valores foram sempre observados ao se considerar o ajuste do modelo Von Bertalanffy e para o modelo Logístico, independentemente do parâmetro avaliado.

Caso seja de interesse o estabelecimento de uma ordem referente à plausibilidade dos modelos em estudo, o modelo Von Bertalanffy seria o mais plausível, conforme discutido no parágrafo anterior, seguido pelo Logístico, que apresenta menores valores de EQM para os parâmetros b e K em relação ao modelo Gompertz. Este último apenas apresenta menor valor de EQM para o parâmetro A, em relação ao modelo Logístico e, além disso, ao considerar o número máximo de 10 iterações, a estabilização do EQM não se verifica para tal modelo.

Quanto à indicação do modelo Von Bertalanffy como o mais plausível, tendo em vista os resultados do estudo de simulação, é importante ressaltar que outros estudos de curvas de crescimento também elegeram este modelo e exaltaram suas qualidades. Silveira (2010) relatou que o modelo Von Bertalanffy foi o mais indicado para descrever curvas de crescimento de três grupos genéticos de cordeiros de corte. Guedes et al. (2004) relataram ser este o modelo que melhor descreveu as curvas de crescimento de cordeiros das raças Santa Inês e Bergamácia e Freitas (2005) reporta que o modelo Von Bertalanffy

e o modelo Logístico são os mais versáteis dentre os modelos de três parâmetros para descrever curvas de crescimento de espécies de interesse econômico.

4.2 Aplicação em dados reais

A aplicação da proposta de isotonização de dados em estudos de curvas de crescimento envolvendo dados reais teve por objetivo melhorar a porcentagem de convergência (%C) e a qualidade de ajuste de modelos de regressão não-linear em uma situação que considera curvas atípicas. Esta situação é caracterizada pela perda de peso em determinadas idades quando bovinos da raça Guzerá são avaliados em uma prova de ganho de peso a pasto. Tendo em vista os dados em questão, as três situações consideradas (dados originais com distúrbio, dados corrigidos via regressão isotônica com pesos iguais e via regressão isotônica com pesos diferentes) podem ser visualizadas na Figura 13 para seis animais aleatoriamente amostrados do rebanho que consistia de 45 animais.

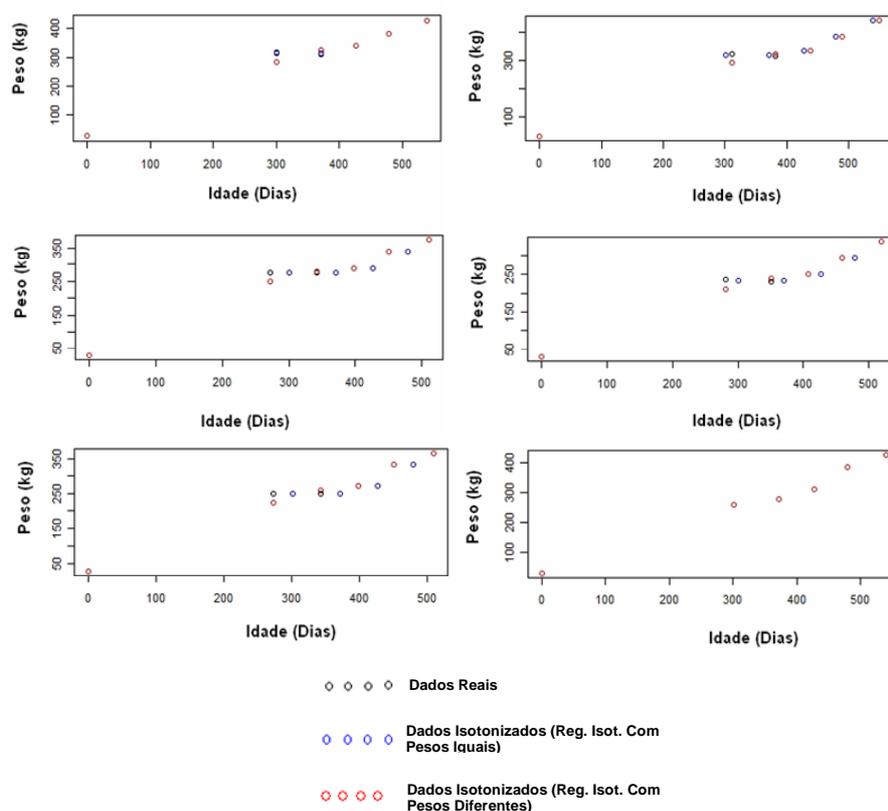


Figura 13 Curvas de crescimento de seis animais da raça Guzerá sob prova de ganho de peso a pasto

O gráfico apresentado na Figura 13 expõe a idéia de transformação de dados via regressão isotônica, o que possibilita, de forma geral, uma recomposição da trajetória natural da curva de crescimento dos animais. Nesta figura, é possível notar que houve perda de peso para alguns animais realmente, uma vez que são explícitas as diferenças nos pesos observados ao se considerar as três situações distintas já mencionadas.

A aplicação das metodologias propostas aos dados reais de crescimento de bovinos da raça Guzerá forneceu resultados relevantes em relação à melhora da porcentagem de convergência (%C) e da qualidade de ajuste dos modelos adotados (QME e R^2). As Tabelas 5, 6 e 7 mostram estes resultados para os modelos Logístico, Bertalanffy e Gompertz, respectivamente. Nas tabelas em questão, as notações estão de acordo com as definições apresentadas no item 3.3.4 e os valores de QME e R^2 são representados pelas médias dos valores obtidos dos ajustes individuais dos 45 animais.

Tabela 5 Resultado da aplicação das metodologias propostas, considerando o ajuste do modelo Logístico a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	% C	QME	R^2
Originais (y)	100	924,7541	0,9597324
Reg. Isot. Simples (y^*)	100	864,0190	0,9622322
Reg. Isot. Ponderada (y^{**})	100	630,1000	0,9718794

Tabela 6 Resultado da aplicação das metodologias propostas, considerando o ajuste do modelo Von Bertalanffy a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	% C	QME	R²
Originais (y)	80	125021,6	0,9650946
Reg. Isot. Simples (y [*])	82	121601,8	0,9696846
Reg. Isot. Ponderada (y ^{**})	84	132956,8	0,9794715

Tabela 7 Resultado da aplicação das metodologias propostas, considerando o ajuste do modelo Gompertz a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	% C	QME	R²
Originais (y)	100	6204,444	0,9707605
Reg. Isot. Simples (y [*])	100	6169,423	0,9730842
Reg. Isot. Ponderada (y ^{**})	100	6482,893	0,9816000

De acordo com os resultados apresentados nas Tabelas 5, 6 e 7 e ainda que o conjunto de dados utilizados apresentasse distúrbios na forma da curva de crescimento (Figura 13), não foi possível perceber melhoras na %C decorrente da aplicação dos métodos de isotonização propostos. Isto porque pôde-se observar 100% de convergência para todas as três situações empregadas (dados originais, corrigidos pela regressão isotônica com pesos iguais e corrigidos pela regressão isotônica com pesos diferentes) nos modelos Logístico e Gompertz.

Apenas uma melhora discreta na %C em relação à aplicação da isotonização foi observada para o modelo Von Bertalanffy e uma vez que as %C observadas para tal modelo foram todas menores que as apresentadas pelos demais, verifica-se que o modelo Von Bertalanffy apresenta dificuldades de convergência para o conjunto de dados em questão, dificuldades estas que não foram constatadas para os modelos Logístico e Gompertz.

Quanto ao QME, os menores valores foram observados para os ajustes dos modelos Von Bertalanffy e Gompertz, considerando os dados corrigidos pela regressão isotônica com pesos iguais. Porém, ao se utilizar o modelo Logístico, observou-se uma redução no QME em relação à seguinte ordem de situações consideradas: dados originais, dados transformados via regressão isotônica com pesos iguais e dados transformados via regressão isotônica com pesos diferentes. Dessa forma, estes resultados referentes ao ajuste do modelo Logístico realmente expressam o comportamento esperado do EQM em relação às metodologias propostas no presente trabalho.

Quanto ao R^2 , nota-se que os valores deste avaliador aumentaram em função da aplicação das técnicas de isotonização para todos os modelos considerados, embora este aumento tenha sido de baixa magnitude. Assim, tendo em vista este avaliador, é possível inferir que as metodologias propostas neste trabalho contribuíram para a melhoria da qualidade de ajuste dos modelos de regressão não-linear adotados.

De forma geral, consideradas as médias de QME e R^2 provenientes dos valores observados para as três situações em cada modelo, observa-se que a média de QME proporcionada pelo modelo Logístico (806,29) realmente foi muito inferior àquelas apresentadas pelos modelos Von Bertalanffy (126.526,70) e Gompertz (6.285,58). Portanto, em relação ao critério QME, o modelo Logístico é o mais indicado para descrever os dados de crescimento considerados. Em relação ao R^2 , pequenas diferenças foram observadas, sendo a

maior média apresentada pelo modelo Gompertz (0,9751), seguida pelos modelos Von Bertalanffy (0,9714) e Gompertz (0,9646).

Na ponderação dos resultados médios de QME e R^2 , pode-se sugerir o modelo Logístico como o mais adequado para descrever o peso dos animais Guzerá em função da idade, ressalvadas as condições experimentais relatadas neste trabalho. Este resultado concorda em parte com o relato de Freitas (2005), quem sugeriu os modelos Von Bertalanffy e Logístico como os mais adequados para estudos de curvas de crescimento na área de Zootecnia. Porém, o referido resultado não concorda com os obtidos por Sousa et al. (2010), que ao trabalhar com o mesmo conjunto de dados indicou o modelo Von Bertalanffy como o mais adequado. Algumas justificativas para esta não concordância podem ser dadas pelo fato de Sousa et al. (2010) ter usado outro software, no caso o SAS[®] via procedimento Model e também por não terem utilizado as transformações nos dados via regressão isotônica.

O método iterativo de isotonização proposto com o intuito de obter uma transformação ótima para dados de peso-idade com distúrbios (item 3.3.4) também foi aplicado aos dados reais. Entretanto, diferentemente da aplicação aos dados simulados, quando o comportamento do EQM (Figuras 10, 11 e 12) foi analisado, nesta aplicação foram analisados os comportamentos do QME conforme figura 14 e do R^2 , Figuras 15.

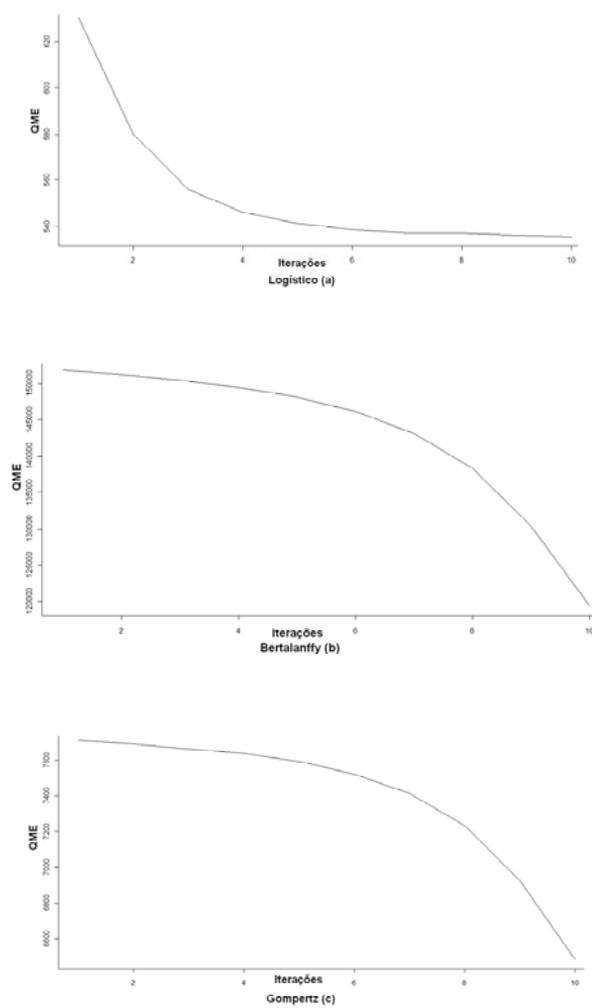


Figura 14 Comportamento do QME para os modelos Logístico, Von Bertalanffy e Gompertz ao longo de 10 iterações consideradas pela metodologia proposta

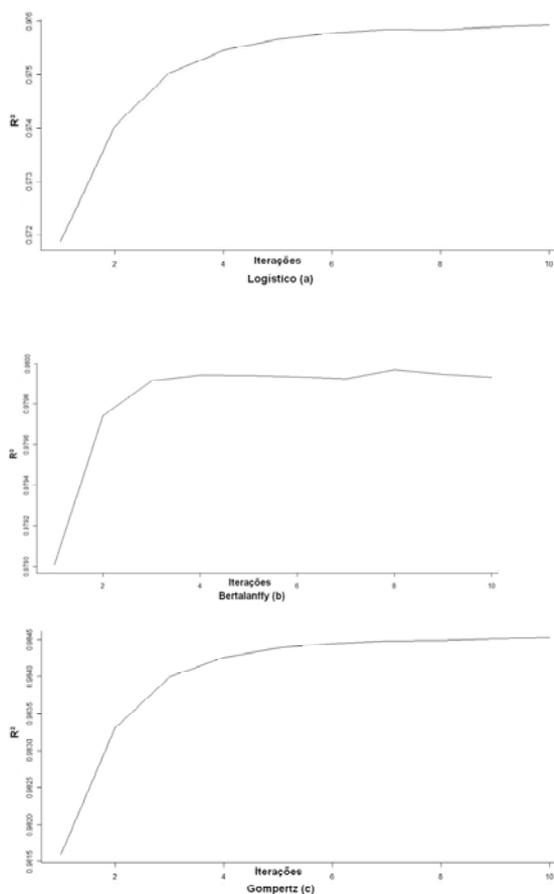


Figura 15 Comportamento do R^2 para os modelos Logístico, Von Bertalanffy e Gompertz ao longo de 10 iterações consideradas pela metodologia proposta

De forma geral, observa-se na Figura 14 um comportamento decrescente do QME para todos os modelos adotados em função do aumento do número de iterações, indicando, assim, a eficiência do método iterativo proposto. É importante realçar que, para o modelo Logístico, (a) a estabilização do QME é evidente e permite inferir que o total de 10 iterações foi suficiente para a

obtenção de uma transformação eficiente dos dados reais de crescimento de bovinos Guzerá, baseada em sucessivas aplicações da regressão isotônica com pesos diferentes. Porém, para os modelos Von Bertalanffy (b) e Gompertz (c), esta estabilização não é observada, sugerindo que um número maior de iterações é necessário para que se obtenha uma transformação ótima dos dados considerados.

Em relação ao R^2 , nota-se na Figura 15 que os valores deste avaliador da qualidade de ajuste tenderam a se estabilizar em torno de 5 iterações para todos os modelos adotados. Este comportamento também elucida a eficiência da metodologia de isotonização iterativa proposta neste trabalho, embora os aumentos constatados nos valores de R^2 em cada iteração sejam de pequena magnitude. Pode-se ainda comentar que esta pequena magnitude está relacionada ao pequeno grau de distúrbio dos dados reais de crescimento, ou seja, pela baixa perda de peso dos animais. Acredita-se que quanto maior a intensidade dos distúrbios, mais visível seja o aumento do R^2 ao longo das iterações executadas.

As Tabelas 8, 9 e 10 mostram, respectivamente, as médias, seus desvios-padrão e coeficientes de variação das estimativas dos parâmetros dos modelos de crescimento Logístico, Von Bertalanffy e Gompertz ajustados aos dados individuais de peso-idade de bovinos Guzerá sob prova de ganho de peso a pasto. As notações das situações contempladas nas tabelas em questão seguem as definições do item 3.3.4. Além disso, é importante lembrar que ao considerar a regressão isotônica com pesos diferentes e iterativa ($\mathbf{y}^{**(10)}$), o número máximo de 10 iterações não garantiu a obtenção de uma transformação ótima quando se ajustam os modelos Von Bertalanffy e Gompertz em função da não estabilização do QME observada na Figura 14.

Tabela 8 Médias, desvios-padrão e coeficiente de variação para as estimativas individuais dos parâmetros do modelo Logístico ajustado a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	Parâmetros		
	A	b	K
Originais (y)	384,7524	7,4107	0,0087
	(59,4221)	(1,1151)	(0,0014)
	15,4%	15,0%	16,1%
Reg. Isot. Pesos Iguais (\hat{y}^*)	385,3174	7,4631	0,0087
	(59,3915)	(1,1079)	(0,0014)
	15,4%	14,8%	16,1%
Reg. Isot. Pesos Diferentes (\hat{y}^{**})	393,8557	7,9059	0,0084
	(59,8319)	(1,2186)	(0,0012)
	15,2%	15,4%	14,3%
Reg. Isot. Iterativa ($\hat{y}^{**(10)}$)	404,9524	8,3876	0,0082
	(61,9374)	(1,4701)	(0,0012)
	15,3%	17,5%	14,6%

Tabela 9 Médias, desvios-padrão e coeficiente de variação para as estimativas individuais dos parâmetros do modelo Von Bertalanffy ajustado a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	Parâmetros		
	A	b	K
Originais (y)	541,8035	0,5858	0,0035
	(253,3900)	(0,0356)	(0,0010)
	46,7%	6,1%	28,6%
Reg. Isot. Pesos Iguais (y^*)	540,5017	0,5864	0,0034
	(249,9100)	(0,0347)	(0,0010)
	46,2%	5,9%	29,4%
Reg. Isot. Pesos Diferentes (y^{**})	570,1498	0,5985	0,0032
	(239,0700)	(0,0317)	(0,0008)
	41,9%	5,3%	25,0%
Reg. Isot. Iterativa ($y^{**(10)}$)	603,9307	0,6067	0,0030
	(232,4100)	(0,0388)	(0,0009)
	38,5%	6,4%	30,0%

Tabela 10 Médias, desvios-padrão e coeficiente de variação para as estimativas individuais dos parâmetros do modelo Gompertz ajustado a dados de peso-idade de bovinos da raça Guzerá

Situação (transformação dos dados)	Parâmetros		
	A	b	K
Originais (y)	450,3442	2,4424	0,0049
	(114,2862)	(0,1898)	(0,0011)
	25,4%	7,8%	22,4%
Reg. Isot. Pesos Iguais (y [*])	451,4651	2,4496	0,0049
	(113,9200)	(0,1872)	(0,0011)
	25,2%	7,6%	22,4%
Reg. Isot. Pesos Diferentes (y ^{**})	469,6984	2,5173	0,0046
	(111,4700)	(0,1875)	(0,0009)
	23,7%	7,4%	19,6%
Reg. Isot. Iterativa (y ^{**⁽¹⁰⁾})	469,6984	2,5173	0,0046
	(111,4702)	(0,1871)	(0,0009)
	23,7%	7,4%	19,6%

De acordo com os resultados apresentados nas Tabelas 8, 9 e 10 e assumindo como referência o modelo Logístico, apontado pelos resultados das Tabelas 5, 6 e 7 como o mais indicado para descrever os dados em questão, nota-se que as estimativas obtidas para o parâmetro A (peso adulto) tendem a assumir valores mais condizentes com a realidade à medida que se utiliza os métodos de isotonização propostos. Quanto ao termo condizente com a realidade, é importante comentar que este foi definido de acordo com as estimativas obtidas em outros trabalhos envolvendo curvas de crescimento de machos da raça Guzerá, trabalhos estes que são escassos na literatura devido a

dificuldade de se coletar os dados de crescimento. Dentre estes, Tedeschi et al. (2000) observaram o valor de 457 kg e Sousa et al. (2010) o valor de 466 kg. No presente trabalho, tendo em vista que o modelo Logístico ajustado aos dados corrigidos via isotonização iterativa proporcionou o valor 404,9524, pode-se inferir que tal valor esteja condizente com a realidade da raça Guzerá.

De forma geral, verifica-se ainda que os valores dos coeficientes de variação (CV) tenderam a ser mais homogêneos entre os parâmetros A, b e K para o modelo Logístico (Tabela 8), No entanto, os valores dos CV's foram maiores para os parâmetros A e K em relação ao parâmetro b para os demais modelos (Tabelas 9 e 10). O fato de as estimativas geradas pelo modelo Logístico serem mais homogêneas que aquelas geradas pelos modelos Von Bertalanffy e Gompertz contribui de alguma forma para a suposição de que realmente tal modelo é o mais indicado para descrever os dados de crescimento observados.

Finalizando a discussão das Tabelas 8, 9 e 10, vale ressaltar que, exceto para o parâmetro b do modelo Von Bertalanffy, para todos os outros parâmetros e todos os outros modelos, os valores de CV tenderam a decrescer à medida que técnicas de isotonização foram utilizadas. Esta descrição reforça ainda mais a hipótese de que a regressão isotônica realmente pode aumentar a qualidade do ajuste de modelos de regressão não-linear a dados de curvas de crescimento com distúrbios provenientes da perda de peso em determinados períodos.

5 CONCLUSÕES

A transformação de dados via análise de regressão isotônica com pesos iguais e pesos diferentes -- possibilita o aumento da porcentagem de convergência e da qualidade dos ajustes dos modelos de regressão não-linear Logístico, Von Bertalanffy e Gompertz a dados de crescimento simulados que apresentam distúrbios caracterizados por decréscimos de pesos em determinadas faixas de idades.

O método de isotonização com base em processo iterativo possibilitou a obtenção de uma transformação eficiente para dados de crescimento com distúrbios.

A realização do estudo de simulação Monte Carlo possibilitou observar a eficiência das metodologias uma vez que todos os processos de isotonização adotados resultaram em maior porcentagem de convergência e menor erro quadrático médio (EQM) para os parâmetros dos modelos avaliados. Este estudo também possibilitou detectar diferenças na intensidade da eficiência de tais metodologias ao considerar diferentes modelos.

A aplicação das metodologias de isotonização em dados reais de peso-idade de bovinos da raça Guzerá também propiciou resultados bons e mais evidentes, considerando-se o ajuste do modelo Logístico. O modelo em questão foi o mais adequado para descrever o crescimento dos animais, uma vez que apresentou alta porcentagem de convergência (100%) e boa qualidade de ajuste, medida por meio do quadrado médio do erro (QME) e do coeficiente de determinação (R^2).

REFERÊNCIAS

- ARAÚJO, L. B. de. **Métodos de correção de autovalores e regressão isotônica nos modelos AMMI**. 2005. 75 f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2005.
- BARLOW, R. E. et al. **Statistical inference under order restrictions: the theory and application of isotonic regression**. London: J. Wiley, 1972. 388 p.
- BROWN, J. E.; FITZHUGH JR., H. A.; CARTWRIGHT, T. C. A comparison of nonlinear models for describing weight-age relationships in cattle. **Journal of Animal Science**, Champaign, v.42, n. 4, p. 810-818, Apr. 1976.
- BRUMBACK, B. A.; RICE, J. A. Smoothing spline models for the analysis of nested and crossed samples of curves. **Journal of the American Statistical Association**, Alexandria, v. 93, n. 443, p. 961-976, Sept. 1998.
- DRAPER, N. R.; SMITH, H. A. **Introdução aos modelos de regressão linear e não-linear**. 3. ed. New York: J. Wiley, 1998. 736 p.
- FITZHUGH JR., H. A. Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, Champaign, v. 42, n. 4, p.1036-1051, Apr.1976.
- FREITAS, A. R. de. Curvas de crescimento na produção animal. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 34, n. 3, p. 786-795, maio/jun. 2005.
- GOTTSCHALL, C. S. Impacto nutricional na produção de carne-curva de crescimento. In: LOBATO, J. F. P.; BARCELLOS, J. O. J.; KESSLER, A. M. **Produção de bovinos de corte**. Porto Alegre: EDIPUCRS, 1999. p. 169-192.
- GUEDES, M. H. P. et al. Estudo das curvas de crescimento de cordeiros das raças Santa Inês e Bergamácia considerando heterogeneidade de variâncias. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 2, p. 381-388, mar./abr. 2004.
- GUNN, L. H.; DUNSON, D. B. A transformation approach for incorporating monotone or unimodal constraints. **Biostatistics**, New York, v. 3, n. 6, p. 434-449, July 2005.

HUSSIAN, M. et al. Monotonic regression of assessment of trends in environmental quality data. In: EUROPEAN CONGRESS ON COMPUTATIONAL METHODS IN APPLIED SCIENCES AND ENGINEERING, 1., 2004, Jyväskylä. **Proceedings**...Jyväskylä: ECCOMAS, 2004. p. 24-28.

MCMANUS, C.; LOUVANDINI, H.; CAMPOS, V. Curvas de crescimento não-lineares para peso e altura em quatro grupos genéticos de cavalos. **Ciência Animal Brasileira**, Goiânia, v. 11, n. 1, p. 80-89, abr. 2010.

MENDES, P. N. et al. Análise da curva de crescimento difásica de fêmeas hereford por meio da função não linear de Gompertz. **Ciência Animal Brasileira**, Goiânia, v. 10, n. 2, p. 454-461, jun. 2009.

OGLIARI, P. J. **Modelos não-lineares para dados longitudinais provenientes de experimentos em blocos casualizados**. 1998. 156 f. Tese (Doutorado em Estatística e Experimentação) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 1998.

RAMSAY, J. O. Estimating smooth monotone functions. **Royal Statistical Society**, New York, v. 2, n. 60, p. 365-375, Jan. 1998.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Viena: R Foundation for Statistical Computing, 2009. Disponível em: <<http://www.R-project.org>>. Acesso em: 25 jun. 2010.

ROBERTSON, T.; WRIGHT, F. T.; DYKSTRA, R. L. **Order restricted statistical inference**. New York: J. Wiley, 1988.

RODRIGUES, A. et al. Avaliação de modelos não-lineares de crescimento na presença de dados incompletos via simulação Monte Carlo. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 51., 2006, Botucatu. **Programas e Resumos**. Botucatu: Universidade Estadual Paulista, Instituto de Biociências, 2006. 1 CD-ROM.

SAS INSTITUTE. **Statistical analysis system user's guide**: version 9. Cary: 2003.

SILVA, F. F.; AQUINO, L. H.; OLIVEIRA, A. I. G. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado Nelore. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 5, p.1195-1205, set./out. 2001.

SILVA, N. A. M. da et al. Curvas de crescimento e influência de fatores não-genéticos sobre as taxas de crescimento de bovinos da raça Nelore. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 3, p. 647-654, maio/jun. 2004.

SILVEIRA, F. G. da. **Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte**. 2010. 59 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2010.

SOUSA, R. C. et al. Curva de crescimento de bovinos da raça Guzerá sob prova de ganho de peso à pasto. In: SIMPÓSIO BRASILEIRO DE MELHORAMENTO ANIMAL, 8., 2010, Maringá. **Anais...** Maringá: Sbma, 2010. p.1-3.

SOUZA, E. M. de. **Teoria da regressão isotônica e algumas relações com a Teoria do Probit, Logit, Delineamento Up and Down e Delineamento da Moeda Viciada**. 2010. 122 f. Exame de Qualificação (Doutorado em Estatística) - Universidade Federal de Lavras, Lavras, 2010.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: Embrapa, 1998. 489 p.

TEDESCHI, L. O. et al. 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**, Viçosa, MG, v. 5, n. 29, p.1578-1587, set./out. 2000.

XIA, Y. et al. On the estimation of an instantaneous transformation for time series. **Royal Statistical Society**, New York, v. 2, n. 62, p.383-397, Jan. 2000.

ANEXOS

Anexo A. Rotina no software R para simulação e análise de dados de crescimento por meio do modelo Logístico, considerando transformações de dados obtidas por regressão isotônica.

```

library(isotone)
library(graphics)
library(cir)

#parâmetros do modelo de crescimento
a=500; b=20; k=0.0075; # valores paramétricos
tf=600 # tempo final
int=30 # intervalo entre as pesagens
var_erro=3 # variância do erro
nexp=1000 # Número de experimentos

n= length(seq(0,tf,int)) # Número de observações
y=matrix(0,nexp,n)
x=matrix(sort((rep((seq(0,tf,int)),nexp))),nexp,n)
e=matrix(0,nexp,n)

#curva original (sem inclusão da função de distúrbio)
#set.seed(23456) #fixando semente aleatória

for(j in 1:nexp)
{
  for(i in 1:n)
  {
    e[j,i]= rnorm(1,0,sqrt(var_erro))
    y[j,i]=(a/(1+b*exp(-k*x[j,i]))) + e[j,i]
  }
}

#curva da derivada discreta para obtenção do ponto de inflexão

y1=matrix(0,nexp,n)
for(j in 1:nexp)
{
  for(i in 2:n)
  {
    y1[j,i]=(y[j,i]-y[j,(i-1)])/(x[j,i]-x[j,(i-1)])
  }
}
aux=cbind(rep(seq(0,tf,int),nexp),matrix(t(y1),n*nexp,1),sort(rep(seq(,nexp),n)))
aux1=aggregate(x = aux[,1], by = list(aux[,2], aux[,3]), FUN = "max")

pf=matrix(0,nexp,1) # armazenar os pontos de inflexão
for(j in 1:nexp)
{
  pf[j]=aux1$x[j*n]
}

par(mfrow=c(3,3))
plot(x[1,],y1[1,])
plot(x[2,],y1[2,])
plot(x[3,],y1[3,])
plot(x[4,],y1[4,])
plot(x[5,],y1[5,])

```

```

plot(x[6,],y1[6,])
plot(x[7,],y1[7,])
plot(x[8,],y1[8,])
plot(x[9,],y1[9,])

# inserção da função de distúrbio

ra=180 # parâmetro beta: controla a intensidade do distúrbio
sc=1700 # parâmetro alfa: controla a extensão do distúrbio

y2=matrix(0,nexp,n)
y3=matrix(0,nexp,n)
y4=matrix(0,nexp,n)
y5=matrix(0,nexp,n)
y6=matrix(0,nexp,n)
y7=matrix(0,nexp,n)
y8=matrix(0,nexp,n)
y9=matrix(0,nexp,n)
y10=matrix(0,nexp,n)
y11=matrix(0,nexp,n)
y12=matrix(0,nexp,n)
y13=matrix(0,nexp,n)
y14=matrix(0,nexp,n)
y15=matrix(0,nexp,n)
y16=matrix(0,nexp,n)
y17=matrix(0,nexp,n)
y18=matrix(0,nexp,n)
y19=matrix(0,nexp,n)
y20=matrix(0,nexp,n)
y21=matrix(0,nexp,n)
y22=matrix(0,nexp,n)
y23=matrix(0,nexp,n)

for(j in 1:nexp)
{
  for(i in 1:n)
  {
    y2[j,i]=((1+b*exp(-k*x[j,i]))) + e[j,i]) + (-ra*exp(-(0.95*x[j,i]-
pf[j,])^2/(sc)))
  }
}
# método iterativo de isotonização

y3[j,]=isoreg(x[j,],y2[j,])$yf
y4[j,]=1/(abs((y3[j,]-y2[j,])))
y5[j,]=cir.pava(y2[j,],x[j,],y4[j,])
y6[j,]=1/(abs((y5[j,]-y2[j,])))
y7[j,]=cir.pava(y2[j,],x[j,],y6[j,])
y8[j,]=1/(abs((y7[j,]-y2[j,])))
y9[j,]=cir.pava(y2[j,],x[j,],y8[j,])
y10[j,]=1/(abs((y9[j,]-y2[j,])))
y11[j,]=cir.pava(y2[j,],x[j,],y10[j,])
y12[j,]=1/(abs((y11[j,]-y2[j,])))
y13[j,]=cir.pava(y2[j,],x[j,],y12[j,])
y14[j,]=1/(abs((y13[j,]-y2[j,])))
y15[j,]=cir.pava(y2[j,],x[j,],y14[j,])
y16[j,]=1/(abs((y15[j,]-y2[j,])))
y17[j,]=cir.pava(y2[j,],x[j,],y16[j,])
y18[j,]=1/(abs((y17[j,]-y2[j,])))
y19[j,]=cir.pava(y2[j,],x[j,],y18[j,])
y20[j,]=1/(abs((y19[j,]-y2[j,])))
y21[j,]=cir.pava(y2[j,],x[j,],y20[j,])
y22[j,]=1/(abs((y21[j,]-y2[j,])))
y23[j,]=cir.pava(y2[j,],x[j,],y22[j,])
}
# gráficos das curvas originais com distúrbio, das curvas corrigidas via regressões
# isotônicas simples e ponderada para os 6 primeiros indivíduos

```

```

par(mfrow=c(3,3))
plot(x[1,],y2[1,])
lines(x[1,],y3[1,],col="blue",type="p")
lines(x[1,],y5[1,],col="red",type="p")
.
.
plot(x[6,],y2[6,])
lines(x[6,],y3[6,],col="blue",type="p")
lines(x[6,],y5[6,],col="red",type="p")

#primeiro resultado: comparação reg original x reg isot comum x reg isot pond

aux2=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y2),n*nexp,1),sort(rep(seq(,n
exp),n))))
aux3=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y3),n*nexp,1),sort(rep(seq(,n
exp),n))))
aux4=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y5),n*nexp,1),sort(rep(seq(,n
exp),n))))

nconver0 =
  sapply(1:nexp,
    function(qual, silent=FALSE){
      da = subset(aux2, X3==qual)
      n0 = try(nls(X2-a/(1+b*exp(-k*X1)), data=da,
        start=list(a=500, b=20, k=0.0075),
      trace=FALSE),silent=silent)
      # se n0 for não nulo porque houve convergência
      if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
    }
  )
pconv_y2=100*sum(!is.na(nconver0[1,]))/nexp

eq_y2 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2, function(x){(c(x)-c(a, b,
k))^2})
eqm_y2 = apply(eq_y2, 1, mean)

nconver0 =
  sapply(1:nexp,
    function(qual, silent=FALSE){
      da = subset(aux3, X3==qual)
      n0 = try(nls(X2-a/(1+b*exp(-k*X1)), data=da,
        start=list(a=500, b=20, k=0.0075),
      trace=FALSE),silent=silent)
      # se n0 for não nulo porque houve convergência
      if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
    }
  )
pconv_y3=100*sum(!is.na(nconver0[1,]))/nexp
eq_y3 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2, function(x){(c(x)-c(a, b,
k))^2})
eqm_y3 = apply(eq_y3, 1, mean)

nconver0 =
  sapply(1:nexp,
    function(qual, silent=FALSE){
      da = subset(aux4, X3==qual)
      n0 = try(nls(X2-a/(1+b*exp(-k*X1)), data=da,
        start=list(a=500, b=20, k=0.0075),
      trace=FALSE),silent=silent)
      # se n0 for não nulo porque houve convergência
      if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
    }
  )
pconv_y4=100*sum(!is.na(nconver0[1,]))/nexp

```

```

eq_y4 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2, function(x){(c(x)-c(a, b,
k))^2})
eqm_y4 = apply(eq_y4, 1, mean)

result1=cbind(rbind(pconv_y2,pconv_y3,pconv_y4),rbind(eqm_y2,eqm_y3,eqm_y4))
result1

#segundo resultado: verificação do EQM para processo iterativo

aux5=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y5),n*nexp,1),sort(rep(seq(,n
exp),n))))
aux6=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y7),n*nexp,1),sort(rep(seq(,n
exp),n))))
aux7=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y9),n*nexp,1),sort(rep(seq(,n
exp),n))))
aux8=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y11),n*nexp,1),sort(rep(seq(,
nexp),n))))
aux9=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y13),n*nexp,1),sort(rep(seq(,
nexp),n))))
aux10=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y15),n*nexp,1),sort(rep(seq(
,nexp),n))))
aux11=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y17),n*nexp,1),sort(rep(seq(
,nexp),n))))
aux12=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y19),n*nexp,1),sort(rep(seq(
,nexp),n))))
aux13=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y21),n*nexp,1),sort(rep(seq(
,nexp),n))))
aux14=data.frame(cbind(rep(seq(0,tf,int),nexp),matrix(t(y23),n*nexp,1),sort(rep(seq(
,nexp),n))))

nconver0 =
  sapply(1:nexp,
    function(qual, silent=FALSE){
      da = subset(aux5, X3==qual)
      n0 = try(nls(X2~a/(1+b*exp(-k*X1)), data=da,
        start=list(a=500, b=20, k=0.0075),
trace=FALSE),silent=silent)
      # se n0 for nÃo nulo Ã porque houve convergÃncia
      if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
    }
  )
eq_y5 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2, function(x){(c(x)-c(a, b,
k))^2})
eqm_y5 = apply(eq_y5, 1, mean)

.
.
.

nconver0 =
  sapply(1:nexp,
    function(qual, silent=FALSE){
      da = subset(aux14, X3==qual)
      n0 = try(nls(X2~a/(1+b*exp(-k*X1)), data=da,
        start=list(a=500, b=20, k=0.0075),
trace=FALSE),silent=silent)
      # se n0 for nÃo nulo Ã porque houve convergÃncia
      if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
    }
  )
eq_y14 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2, function(x){(c(x)-c(a, b,
k))^2})
eqm_y14 = apply(eq_y14, 1, mean)

```

```
result2=rbind(eqm_y5, eqm_y6, eqm_y7, eqm_y8, eqm_y9, eqm_y10, eqm_y11, eqm_y12,  
eqm_y13 , eqm_y14)  
  
par(mfrow=c(2,2))  
plot(seq(1,10),result2[,1],type="l")  
plot(seq(1,10),result2[,2],type="l")  
plot(seq(1,10),result2[,3],type="l")
```