



UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Rayane Santos Leite

**Uma introdução a regressão isotônica em  
modelos não lineares na dinâmica do potássio  
em solos não saturados**

Campina Grande - PB

Dezembro de 2015

Rayane Santos Leite

**Uma introdução a regressão isotônica em modelos não lineares na dinâmica do potássio em solos não saturados**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Ana Patricia Bastos Peixoto

Campina Grande - PB

Dezembro de 2015

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

L533i Leite, Rayane Santos.

Uma introdução a regressão isotônica em modelos não lineares na dinâmica do potássio em solos não saturados [manuscrito] / Rayane Santos Leite. - 2015.  
37 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2015.

"Orientação: Profa. Dra. Ana Patricia Bastos Peixoto, Departamento de Estatística".

1. Modelos não lineares. 2. Regressão isotônica. 3. Solo. 4. Dinâmica do solo. I. Título.

21. ed. CDD 519.536

Rayane Santos Leite

## Uma introdução a regressão isotônica em modelos não lineares na dinâmica do potássio em solos não saturados

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 11 de Dezembro de 2015.

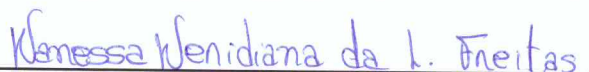
### BANCA EXAMINADORA



Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Patricia Bastos Peixoto  
Universidade Estadual da Paraíba



Prof. Dr. Gustavo Henrique Esteves  
Universidade Estadual da Paraíba



Prof<sup>a</sup>. Msc. Wanessa Weridiana da Luz  
Freitas  
Universidade Estadual da Paraíba

*A Jesus de Nazaré,  
Aos meus avós Margarida (in memoriam) e Pedro,  
Aos meus pais Redi e Fátima,  
Ao meu irmão Lucas.  
Com amor, DEDICO.*

# Agradecimentos

A Deus, agradeço pelo dom da vida, pelo seu infinito amor para comigo e por todas as bênçãos ofertadas, pois sem ele nada sou.

Agradeço aos meus pais Redi Ferreira Leite e Maria de Fátima Santos Leite, por serem meus maiores exemplos, pelas orações em meu favor, incentivos incondicionais, orientações e preocupações para que eu pudesse sempre percorrer o caminho certo. E o amor a mim dedicado foi imprescindível neste percurso de formação. As palavras não expressam a minha gratidão por terem feito do meu sonho o de vocês.

Agradeço a minha avó materna Margarida (“in memoriam”) por todo legado que deixou por ser um exemplo de inspiração pra minha vida, pelos valorosos ensinamentos de grande importância na minha educação.

Ao meu avô materno Pedro pelo amor e incentivo que me foi essencial e que sempre me deu todo apoio financeiro que estava ao seu alcance.

Aos avós paternos Luis e Antônia por todo o carinho para comigo.

Ao meu irmão Lucas Santos Leite por todo seu amor e carinho que me dedica e por sempre se preocupar e acreditar nos meus sonhos.

A minha tia e madrinha Sira sua contribuição foi muito valiosa neste percurso, obrigada por sempre ser presente em minha vida.

A minha família sou grata.

Ao meu namorado, Vicente, por todo amor, carinho, paciência e compreensão, expressos em palavras de apoio que me foram muito importantes.

A professora Dr<sup>a</sup>. Ana Patricia Bastos Peixoto, que dedicou seu valiosíssimo tempo em me orientar, pelos incentivos nas mais diversas circunstâncias, apoios e confiança em mim. Agradeço porque esta confiança se fez presente em muitas horas que eu mesma não conseguia acreditar ser possível a realização dos meus objetivos. Você é mais do que uma professora ou uma orientadora, se tornou uma amiga, deixo aqui minha admiração pela profissional extremamente qualificada que és e pela forma humana que conduziu minha orientação. Foi ótimo aprender com você um pouco de estatística e outros valores que não estão contidos nos livros.

Aos professores Ana Patricia, Tiago, Gustavo, João Gil, Ricardo, Nyedja, Vitória, Nathielly, Giselly, Érica e Silvio, pela contribuição na minha vida acadêmica e por tamanha influência na minha futura vida profissional.

Aos meus amigos de curso Aline e Leomir, a quem aprendi a amar e construir laços,

obrigada por todos os momentos vivenciados, pelas horas de estudo, dedicação, emoções, brincadeiras, e por toda a cumplicidade. Em vocês encontrei verdadeiros irmãos. Obrigada pela paciência, pelo abraço, pelo sorriso, e pela mão sempre disposta a me estender, sem vocês esta caminhada jamais seria a mesma.

A minha grande amiga de infância e irmã Ana Stefella que mesmo na distância sempre se fazia presente em minha vida, colaborando com palavras de carinho.

A minha primeira amiga em Campina Grande Jordanya e que espero levar pra vida toda, pelos conselhos, disponibilidades, que sempre me deu um ombro amigo quando eu precisei nas mais adversas situações ocasionadas pela vida, por toda a colaboração para que eu concluísse esta caminhada tão árdua a qual ela sempre me deu muita força.

Aos amigos, Bruno, Jakciane e Juliana que sempre me animou, dando-me palavras de esteio.

A Universidade Estadual da Paraíba pela oportunidade de realizar este curso, e ao Departamento de Estatística.

Aos meus amigos de classe Mario e Manoel pelas as horas de dificuldades e alegrias, enfim todos os momentos vivenciados juntos.

A Sônia, Arnete e Pablo que sempre se demonstraram prestativos para comigo.

A todos os alunos que cursei disciplinas com eles, e aos que não cursei, mas tive o prazer de conhece-los.

A todos os nominados, e não nominados que de alguma forma contribuíram para realização deste trabalho, meu muito obrigado.

*Mesmo que as montanhas oscilassem e as colinas se abalassem,  
jamais meu amor te abandonará e jamais meu pacto de paz vacilará,  
diz o Senhor que se compadeceu de ti.  
(Isaías 54:10)*



# Resumo

Vários modelos computacionais têm sido difundidos nos últimos tempos, visando descrever o movimento de solutos por meio do perfil do solo, porém é possível observar que existe grande dificuldade em envolver todos os parâmetros pertinentes ao meio ambiente, para que o modelo possa realmente prever o processo de deslocamento e retenção dos solutos na natureza. A regressão isotônica é uma técnica de transformação de dados que possibilita a incorporação de um conhecimento prévio, em que a variável dependente é não decrescente em relação à variável independente. Verificou-se no presente caso a importância do ajuste do modelo logístico com três parâmetros para representar a dinâmica do solo, e o teor de potássio presente no perfil, servindo para inferir sobre os aspectos que constituem a dinâmica, como o transporte dos solutos. O uso da metodologia de transformação de dados por meio da análise de regressão isotônica, possibilitou um aumento da porcentagem de convergência e a qualidade dos ajustes do modelo de regressão não linear logístico a dados de crescimento que apresentam distúrbios caracterizados por decréscimos dos teores do soluto ao longo do tempo. A visualização gráfica explicita a ideia de transformação de dados via regressão isotônica, a qual permitiu de certa forma, recuperar a trajetória natural da curva representativa do fenômeno em estudo, entretanto, não houve mudanças expressivas nos valores encontrados.

**Palavras-chaves:** Dinâmica do solo. Modelo não linear. Regressão isotônica.

# Abstract

Several computational models have been widespread in recent years, aiming to describe the movement of solutes through the listing soil, but it can see that there is great difficulty in involving all parameters relevant to the environment, so that the model can actually predict process shift and retention of solutes in nature. The isotonic regression is a data processing technique that enables incorporassão prior knowledge where the dependent variable is nondecreasing with respect to an independent variable. It was found in this case the importance of the logistic model fit with three parameters to represent the dynamics of soil, and the present potassium content in the profile, serving to infer on the aspects that make up the dynamics, such as the transport of solutes. Given the above, this study aimed to use the isotonic regression in the context of non-linear models in order to increase the quality of model adjustment. The use of the data processing method by isotonic regression analysis allowed an increase in the percentage of convergence and the quality of settings of the non-linear regression logistic model the growth data show disorders characterized by reduction of the solute contents the over time. The graphical visualization explains the data processing idea via isotonic regression, which allowed a certain way, recover the natural trajectory of the representative of the phenomenon under study curve, however, there were no significant changes in the values found.

**Key-words:** Dynamic soil. Nonlinear model. Isotonic regression.

# Lista de ilustrações

Figura 1 – Ajuste do teor de potássio em relação a profundidade nos solos LVA (a) e NV (b) para os logístico com 3 parâmetros. . . . .	26
Figura 2 – Gráfico dos resíduos projetados para o teor de potássio nos solos LVA e NV ao longo da profundidade. . . . .	26
Figura 3 – Gráfico das distâncias de Cook e elementos da diagonal da matriz de projeção <i>versus</i> a ordem das observações para do teor de potássio nos solos LVA e NV ao longo da profundidade. . . . .	27
Figura 4 – Ajuste da regressão isotônica do teor de potássio nos solos LVA e NV ao longo da profundidade. . . . .	27
Figura 5 – Ajuste da regressão antitônica e diagrama da soma cumulativa com menor majorant côncava do teor de potássio no solo LVA ao longo da profundidade. . . . .	28
Figura 6 – Ajuste da regressão antitônica e diagrama da soma cumulativa com menor majorant côncava do teor de potássio no solo NV ao longo da profundidade. . . . .	28

# Lista de tabelas

Tabela 1 – Estimativas dos parâmetros do modelo logístico com três parâmetros, erro padrão da estimativa (E.P.E.), valores $p$ para o teste $t$ e intervalos de confiança (IC) de 95%, para o teor de potássio ao longo da profundidade nos solos LVA e NV. . . . .	25
---	----

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>12</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>13</b>
2.1	Marco Histórico . . . . .	13
2.2	Dinâmica do soluto no solo . . . . .	14
2.3	Modelos de regressão não lineares . . . . .	15
2.3.1	Procedimentos de otimização . . . . .	16
2.3.2	Método de Gauss Newton . . . . .	16
2.3.3	Análise de diagnóstico . . . . .	19
2.4	Regressão isotônica . . . . .	19
<b>3</b>	<b>MATERIAL E MÉTODOS</b> . . . . .	<b>23</b>
3.1	Isotonização de dados de curva de crescimento . . . . .	24
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>25</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>30</b>
<b>6</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> . . . . .	<b>31</b>
	<b>APÊNDICES</b>	<b>33</b>
	<b>APÊNDICE A – ROTINA R UTILIZADA PARA OBTENÇÃO DO AJUSTE DO MODELO LOGÍSTICO</b> . . . . .	<b>34</b>

# 1 Introdução

O transporte dos solutos é alvo de investigação de um grande número de modelos de simulação do movimento dos solutos no perfil do solo. Os modelos estatísticos surgem como ferramentas úteis, possíveis de serem aplicados na definição de quando e de como se deve proceder diante de situações que necessitam de determinado manejo, racionalizando-se a operação e tornando-a mais eficiente em seus vários aspectos. Para tanto, existem modelos analíticos e numéricos desenvolvidos com a finalidade de prever os processos de transferência da água e de solutos entre a superfície do solo e o lençol freático.

Em análise de regressão, algumas vezes as características intrínsecas do problema ou outras considerações indicam que o modelo apropriado para a situação é um modelo não linear. Um modelo de regressão é dito não linear, se ele é não linear como função dos seus parâmetros. Neste estudo, busca-se o ajuste de modelos que não são possíveis de serem linearizados por meio do uso de transformações de variáveis. Assim, de acordo com Draper e Smith (1998), esses modelos são denominados de “intrinsecamente não lineares” e quando informações teóricas com relação ao modelo levam a um modelo intrinsecamente não linear, deve-se preferir ajustar tal modelo sempre que possível, do que ajustar um modelo linear, com menos precisão.

Em algumas situações práticas, devido a alguma informação *a priori*, as observações e método de análise desses dados de interesse apresentam algum tipo de ordenação. Assim, deve-se buscar um modelo que preserve esta característica de ordem. Uma alternativa muito interessante é a abordagem denominada regressão isotônica, sendo que o termo isotônico (ou monotônica) refere-se ao fato de que um aumento na variável independente implique em um aumento na variável resposta. Se a variável resposta diminui com o aumento da variável preditora, utiliza-se o termo antitônica. Nesse aspecto, uma alternativa que pode propiciar melhoria na qualidade de ajuste é a utilização da regressão antitônica (BRITO et al., 2012).

Diante do exposto, a proposta deste trabalho é ajustar um modelo estatístico que represente o transporte dos solutos ao longo do perfil do solo, cujos parâmetros possam ter uma interpretação biológica e utilizar a regressão isotônica, visando à definição de isotonização como uma transformação de dados, e aplicação de tal técnica na solução de problemas de convergência em estudos de curvas de crescimento via modelos de regressão não linear.

## 2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos da dinâmica do soluto no solo, bem como uma revisão sobre os modelos de regressão não lineares e regressão isotônica, utilizando-se de artigos com enfoque teóricos e práticos relacionados ao objetivo da pesquisa.

### 2.1 Marco Histórico

Em acordo com Eeden (1996), a teoria da regressão isotônica foi desenvolvida nos anos 50 enfatizando o problema de que duas probabilidades  $\theta_1$  e  $\theta_2$ , com  $\theta_1 \leq \theta_2$  precisavam ser estimadas, utilizando-se a estimação por máxima verossimilhança. Os estimadores de máxima verossimilhança irrestrito mostram-se ser inadmissíveis pelo critério de erro quadrático. Isto conduziu esta estimação em espaços paramétricos restritos. Neste contexto, surgiram diversos trabalhos na mesma década, entre estes, destaca-se o trabalho de dissertação elaborado por Constance van Eeden. O referido trabalho foi defendido na Universidade de Amsterdam e tem como título: “Testing and Estimating Ordered Parameters of Probability Distributions”, e posteriormente em quatro artigos publicados entre os anos 1956-1957 em *Indagationes Mathematicae* (LEEUEW; HORNIK; MAIR, 2009).

Os fundamentos da teoria supracitada foram expostos no livro de Barlow et al. (1972). Houve também outra contribuição apresentada no livro de Robertson, Wright e Dykstra (1988), que apresentou os resultados fornecidos em Barlow et al (1972) acrescidos de algumas contribuições posteriores. O problema da regressão quadrática é também um problema de programação quadrática, motivando o surgimento de diversos trabalhos nas áreas ligadas à informática, desde a apresentação de algoritmos mais eficientes até a apresentação de programas em softwares estatísticos convencionais. Tais como o sas, spss, excel, stata, minitab e epi info. Ressalta-se que por volta de 2002, esta teoria passa a ter também conexão com experimentos de dose-resposta que se utilizam de delineamentos sequenciais, com trabalhos voltados para a área de ensaios clínicos (SOUZA, 2011).

É válido salientar que existe uma escassez de trabalhos nesta perspectiva. Na maioria, são trabalhos de aplicação, utilizando-se da teoria apenas como ferramenta, sem uma exposição em aspectos teóricos, esta teoria foi utilizada em estudos de curvas de crescimento (RODRIGUES et al. 2010).

O termo “isotônica” refere-se ao fato de a variável resposta aumentar com o aumento da variável independente. A técnica da regressão isotônica consiste em encontrar uma

função não decrescente que minimize a soma de quadrados dos erros (BARLOW et al., 1972). A próxima grande contribuição foi o livro de Robertson et al. (1988), que apresentou os resultados fornecidos em Barlow et al. (1972) acrescidos de algumas contribuições posteriores.

Corroborando com esta discussão Souza, Chaves e Muniz (2012), apontam que a teoria da regressão isotônica é uma ferramenta estatística natural para estudos de modelos de regressão em que os parâmetros devem ter uma restrição de ordem. As aplicações desta teoria podem ser relativamente amplas e seu estudo do ponto de vista teórico necessita de conceitos matemáticos relativamente sofisticados.

A parte de regressão isotônica deste trabalho foi estudada com base no livro clássico *Statistical Inference under Order Restrictions: The Theory Application of Isotonic Regression*, de Barlow et al. (1972). Então, as demonstrações na parte de isotônica que serão expressas posteriormente, podem ser compreendidas como uma contribuição didática à teoria.

## 2.2 Dinâmica do soluto no solo

O termo solo refere-se à camada externa e agricultável da superfície da terra, cuja origem é determinada pela ação de processos físicos, químicos e biológicos na rocha, dando origem a um material natural, sólido e poroso que abriga em seus poros quantidades variadas de soluções. As diferentes ações desses processos resultam em tipos distintos de solos, os quais apresentam uma disposição diferente com relação a esse espaço poroso e que por sua vez formam canais naturais, de onde se originam as interações entre a água e as partículas sólidas do solo, resultando-se em forças de atração e fixação da água no solo, diminuindo sua energia em relação à água livre (MIRANDA et al., 2005).

O entendimento da qualidade do solo é fundamental tendo em vista a necessidade de adoção de estratégias para um manejo sustentável dos diversos sistemas de produção (TÓTOLA; CHAER, 2002), um objetivo muito almejado nos dias atuais. A qualidade do solo refere-se às condições ótimas para que o solo se apresente adequadamente. O funcionamento do solo depende da interação de processos químicos, físicos e biológicos, que mantêm um fluxo constante e uma natureza heterogênea (TÓTOLA; CHAER, 2002). Desta forma, inferir sobre qualidade do solo de forma numérica torna-se uma tarefa difícil, entretanto, estimativas sobre a qualidade do solo podem ser feitas dentro de um marco referencial (REICHERT; REINERT; BRAIDA, 2003).

A qualidade física do solo descreve como o solo permite a infiltração, retenção e disponibilidade de água às plantas, córregos e subsuperfície, responde ao manejo e resiste à degradação, permite as trocas de calor e de gases com a atmosfera e raízes das plantas, e permite o crescimento das raízes (REICHERT; REINERT; BRAIDA, 2003). As



chuvas que atingem a superfície do solo sofrem a ação de processos físicos tais como, a infiltração da água no solo, a retenção na superfície e o escoamento superficial. No caso da redistribuição da água no solo, parâmetros físicos, tais como a condutividade hidráulica, são extremamente importantes na medida em que auxiliam no entendimento da dinâmica da água, tornando-a fundamental para os estudos da drenagem agrícola.

Dentre os macronutrientes, depois do nitrogênio (N), o potássio (K) é o elemento requerido em maiores quantidades pelas espécies de plantas cultivadas (NACHTIGALL; VAN RAIJ, 2005). Na planta, o potássio está envolvido no metabolismo do nitrogênio, transporte de açúcar, extração, transporte e uso da água e resistência ao estresse (KRAUSS, 2005). O potássio tem funções reguladoras muito importantes, sendo necessário para ativar pelo menos sessenta enzimas. Está ligado também ao processo fotossintético em vários níveis, participa da síntese do ATP e afeta a taxa de assimilação do dióxido de carbono.

## 2.3 Modelos de regressão não lineares

Existem vários problemas em estatística e um bem comum é o estudo da relação entre duas variáveis  $X$  e  $Y$ , isso significa que procura-se uma função de  $X$  que explique  $Y$ . Há casos em que alguns momentos o interesse não se resume em conhecer se existe ligação entre duas variáveis quantitativas  $X$  e  $Y$ , mas também em saber uma provável afinidade de causa e efeito entre as variáveis. Anseia-se conhecer se  $Y$  depende de  $X$ , se isto acontece  $Y$  é chamada de variável dependente e  $X$  é chamada de variável independente.

Um modelo de regressão pode ser chamado de linear, quando supõe-se que a relação da resposta às variáveis é uma função linear de alguns parâmetros. Existem ainda, os modelos de regressão que não são funções lineares dos parâmetros chamadas de modelos de regressão não lineares.

Um modelo de regressão é considerado não linear, se o mesmo for não linear com função em relação aos seus parâmetros. Os modelos que não são viáveis a serem linearizados por meio da utilização de transformações de variáveis.

Gallant (1987) observou que a ocorrência mais natural em uma análise estatística são experimentos que guiam as variáveis respostas univariadas, indicadas por  $Y$ . Além disso, a variável resposta pode estar ligada a alguma variável explicativa, indicada por  $X$ . Então, para uma amostra aleatória de tamanho  $n$ , o modelo de regressão não linear é escrito,

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

em que  $y_i$  é a observação da variável dependente;  $f(x_i, \boldsymbol{\theta})$  é a função resposta não linear em  $\boldsymbol{\theta}$ ,  $x_i$  representa a variável independente;  $\boldsymbol{\theta}$  é um vetor de parâmetros e  $\varepsilon_i$  é o erro

experimental não observável diretamente, suposto independente e identicamente distribuído (i.i.d) de uma distribuição normal com média 0 e variância  $\sigma^2$  desconhecida.

O ajuste dos parâmetros de modelos não lineares pode ser obtido pela minimização da soma dos quadrados dos resíduos, conhecida como método dos mínimos quadrados, da por,

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2. \quad (2.2)$$

É natural que após a relação de um dado experimento  $(x_i, y_i)$ , serem observações fixas e conhecidas, então  $S(\boldsymbol{\theta})$  depende exclusivamente de  $\boldsymbol{\theta}$ . Denota-se por  $\hat{\boldsymbol{\theta}}$  os estimadores de mínimos quadrados de  $\boldsymbol{\theta}$ , ou seja, os valores de  $\hat{\boldsymbol{\theta}}$  que minimizam  $S(\boldsymbol{\theta})$ . Para determinação dos estimadores de mínimos quadrados  $\hat{\boldsymbol{\theta}}$  deve-se derivar com relação a cada  $\theta_j$  ( $j = 1, \dots, p$ ). Fazendo isto, determina-se  $p$  equações, chamadas de equações normais na forma

$$\sum_{i=1}^n \{y_i - f(x_i, \boldsymbol{\theta})\} \left[ \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0,$$

e quando  $\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j}$  não depende de  $\boldsymbol{\theta}$ , ou tem-se as equações normais de um modelo de regressão linear. Para funções esperança não lineares,  $\boldsymbol{\theta}$  estará presente em pelo menos uma das derivadas parciais de  $f(x_i, \boldsymbol{\theta})$ .

### 2.3.1 Procedimentos de otimização

Em termos práticos, o processo de estimação de um parâmetro em um modelo não linear é obtido de forma semelhante aos estimadores dos modelos lineares, por meio da minimização dos erros quadrados. Contudo, muitas vezes, a obtenção das estimativas dos parâmetros no modelo não linear pode ser problemática (SOUZA, 1998; RITZ, C.; STREIBIG, 2008), pois como em modelos multiparamétricos, as soluções das equações normais podem ser extremamente difíceis, não apresentando solução explícita, sendo necessário o uso de algum método iterativo de resolução para equações não lineares (BATES; WATTS, 1988; RATKOWSKY, 1983). Dentre os métodos iterativos propostos, os mais utilizados são o método de Gauss-Newton ou método da linearização, o método *Steepest-Descent* ou método do Gradiente e o método de Marquardt (BATES; WATTS, 1988).

### 2.3.2 Método de Gauss Newton

O método de Gauss-Newton é também conhecido como método da linearização, fazendo uso de uma aproximação em série de Taylor para aproximar o modelo de regressão não linear com termos lineares. É um dos mais populares para solucionar problemas que envolvem mínimos quadrados em modelos não lineares. Composto de iterações que são

passos que em geral guiam uma solução para o problema de regressão não linear, com a finalidade base de aceitar aproximações lineares para os  $\varepsilon_i(\theta)$ 's em cada iteração  $f(x_i, \theta)$ .

Assim, o problema de minimizar a função das somas de quadrados dos resíduos,

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

passa a ser o de diminuir na  $(k+1)$ -ésima iteração, a função  $S^{(k)}(\theta)$ , em que

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

e  $f^{(k)}(x_i, \theta)$  é a aproximação de série de Taylor de 1ª ordem de  $f(x_i, \theta)$  em torno do ponto  $\theta^{(k)}$ , ou seja, em torno de valores principiantes do parâmetro. Assim a solução geral para o  $k$ -ésimo caso é dada por

$$f^{(k)}(x_i, \theta) \cong f(x_i, \theta^{(k)}) + \sum_{j=1}^n \frac{\partial f(x_i, \theta)}{\partial \theta_j} (\theta_j - \theta_j^{(k)}).$$

Em forma matricial, dispõe-se de

$$S^{(k)}(\theta) = \|Y - f(X, \theta^{(k)}) - F(\theta^{(k)}) \cdot (\theta - \theta_j^{(k)})\|_2^2,$$

em que,  $F(\theta^{(k)})$  é a matriz jacobiana de  $f(\mathbf{X}, \theta)$  no ponto  $\theta^{(k)}$  e  $\theta - \theta^{(k)}$ .

Assim, considerando-se que  $\mathbf{Y} - f(\mathbf{X}, \theta^{(k)}) = \mathbf{E}^{(k)}$  tem-se

$$S^{(k)}(\theta) = \|\mathbf{E}^{(k)} - F(\theta^{(k)}) \cdot (\theta - \theta_j^{(k)})\|_2^2,$$

em que,  $\mathbf{E}^{(k)}$  “faz o papel” da variável dependente  $Y$  e  $F(\theta^{(k)})$  a matriz  $X$  dos modelos lineares.

Entretanto, se  $F(\theta^{(k)})$  for uma matriz posto completo, o “valor”  $\theta - \theta_j^{(k)}$  que minimiza  $S^{(k)}(\theta)$  é definida por

$$\theta - \theta_j^{(k)} = (F^{(k)T} F^{(k)})^{-1} F^{(k)T} \mathbf{E}^{(k)},$$

em que,  $F^{(k)} = F(\theta^{(k)})$  é o valor de  $\theta$  obtido na  $(K+1)$ -ésima iteração deverá ser

$$\theta^{(k+1)} = \theta^{(k)} + (\theta - \theta_j^{(k)}),$$

portanto, o algoritmo de Gauss Newton deverá ser mais eficiente quanto melhor for a aproximação

$$H(\theta) \cong 2F(\theta)^T F(\theta).$$

Está hipótese é sensata quando os resíduos  $\varepsilon_i(\theta)$  forem suficientemente pequenos, ou ainda quando o grau de linearidade do modelo é pequeno, que é, quando o modelo é próximo do linear, em que nestes casos  $B(\theta) = 0$ .

Quando o modelo é linear, o algoritmo de Gauss Newton converge para o ponto de mínimo  $\theta^*$  numa única iteração partindo-se de qualquer ponto inicial  $\theta^{(0)}$ . Entretanto, a medida que o grau de não linearidade de um modelo aumenta, o algoritmo Gauss Newton pode não convergir, e conseqüentemente, a função objeto (soma de quadrados dos resíduos) pode ter mais de um ponto de mínimo e a probabilidade de convergência para o mínimo global vai diminuindo.

Por outro lado, a taxa de convergência do algoritmo de Gauss Newton (sem nenhuma modificação) pode ser avaliada por meio da seguinte desigualdade, apresentada por Mc Keown (1980)

$$\frac{\|\theta^{(k-1)} - \theta^{(*)}\|}{\|\theta^{(k)} - \theta^{(*)}\|} < \left\| \left( F^{*T} F^{(*)} \right)^{-1} B^* \right\| < \frac{(\lambda) \max(B^*)}{\lambda \min(F^{*T} F^*)} \cong \rho,$$

em que  $F^* = F(\theta^*)$ ,  $B^* = (\theta^*)$ ,  $|\lambda| \max(B^*)$  é o maior autovalor de  $B^*$  e  $\lambda_{\min}(F^{*T} F^*)$  é o menor autovalor de  $F^{*T} F^*$  e  $\rho$  é um valor pré-fixado.

Assim, qualquer procedimento iterativo requer para sua execução, a especificação de um valor inicial. No contexto de otimização, é sempre importante tentar obter valores iniciais adequados para o problema. No caso de existirem pontos críticos, tais como mínimos locais ou pontos de celas, um valor inicial não apropriado pode levar o processo a convergir para um desses pontos e não para a solução global de interesse. Em algumas situações, utiliza-se experiências anteriores ou análises semelhantes que podem ser usadas como o valor inicial. Em outros casos, precisa-se utilizar algumas técnicas para auxiliar na obtenção desses valores (BATES; WATTS, 1988).

Dessa forma, alguns procedimentos utilizados nas aplicações de modelos não lineares para resolver o problema de encontrar valores iniciais são:

- i) O método de Hartley e Booker (1965), que consiste em dividir o conjunto de  $N$  observações em  $r$  grupos e resolver o sistema não linear obtido nessa subdivisão.
- ii) Considerar uma rede de pontos no espaço paramétrico, tomando-se para valor inicial, por exemplo, o ponto da rede cujo valor da função objeto é mínimo.
- iii) Uma outra possibilidade é o procedimento desenvolvido por Ratkowsky (1983). Esse procedimento é recomendado quando a suposição de erro multiplicativo for razoável para o modelo e se baseia no critério de mínimos quadrados ponderados.

No entanto, um procedimento mais direto de se encontrar um valor inicial, para com isso obter as estimativas dos parâmetros do modelo, considerando modelo o logístico, é esboçar  $y_n$  versus  $x_n$  e tomar o valor da assíntota desse gráfico, já que a medida quando  $x_n \rightarrow \infty$ , tem-se  $y_n \rightarrow \theta_1$ , assim, um bom valor inicial para o parâmetro  $\theta$  é o  $y_{\max}$ , que possivelmente deve estar próximo da assíntota.

### 2.3.3 Análise de diagnóstico

A metodologia de diagnóstico iniciou-se com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta. A análise de resíduos num modelo estatístico pode ser baseada nos resíduos ordinários, ou em versões padronizadas, ou em resíduos construídos a partir dos componentes da função desvio (MCCULLAGH; NELDER, 1989), ou em resíduos generalizados (COX; SNELL, 1968). As análises podem incluir gráficos informais para exibir características gerais dos resíduos, bem como testes formais detectados por meio de hipóteses. Tais procedimentos formais ou informais são complementares, e ambos têm um lugar na análise residual (COOK; WEISBERG, 1982).

Por meio dos resíduos, também pode-se encontrar possíveis pontos influentes e/ou *outliers*. Dessa forma, uma ou mais observações são ditas discrepantes (*outliers*) se seus resíduos são muito grandes em relação aos demais (DRAPER; SMITH, 1998). Já os pontos influentes são observações que, embora não apresentem resíduos grandes, podem alterar significativamente as estimativas dos parâmetros do modelo escolhido. Para detectar a presença de pontos influentes ou de *outliers*, técnicas gráficas, como diagramas de dispersão e gráficos da distância de Cook podem ser usadas.

Segundo Cordeiro e Paula (1989), embora as técnicas de diagnóstico da regressão não linear sejam simples extensões das técnicas da regressão linear, as interpretações não são diretamente aplicáveis, particularmente, em virtude de os resíduos ordinários não terem mais uma distribuição aproximadamente normal.

## 2.4 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 da 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.

Rodrigues et al. (2010) propôs uma metodologia de transformação de dados, por intermédio da análise de regressão isotônica, para estudos de curvas de crescimento cujos dados expõem 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, o estudo também propôs um procedimento iterativo de isotonização, cujo intuito foi obter uma transformação ótima para os dados. Os autores concluíram que as metodologias de

transformação de dados por meio da análise de regressão isotônica, com pesos iguais e pesos diferentes, possibilitaram aumentar a porcentagem de convergência e a qualidade dos ajustes dos modelos de regressão não linear Logístico, Von Bertalanffy e Gompertz a dados de crescimento que apresentam distúrbios caracterizados por decréscimos de pesos em determinadas faixas de idades.

Na análise de regressão usual, não se usa uma informação *a priori* sobre uma relação de ordem na variável resposta. Na teoria da regressão isotônica essa informação é essencial, portanto é de se esperar que a eficiência da análise seja aumentada quando faz-se uso da mesma.

Existe um problema comum de regressão e está relacionado a um conjunto de números pré-fixados  $\mathbf{X} = \{x_1, x_2, \dots, x_k\}$  onde  $x_1 < x_2 < \dots < x_k$ . Para  $x_i \in X$ ,  $i = 1, 2, \dots, k$  seja  $y_j(x_i)$ ,  $j = 1, 2, \dots, m(x_i)$ , uma amostra i.i.d de uma distribuição com média desconhecida  $\mu(x_i)$  (no caso de experimentos do tipo dose resposta, os valores de  $X$  seriam as doses, os valores de  $y_j(x_i)$  a resposta para a dose  $x_i$  e  $m(x_i)$  o número de repetições para cada dose). Se é assumido que  $\mu(x)$  é uma função linear de  $x$ , uma maneira de se estimar  $\mu(x)$  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 de funções lineares  $f$ . Se

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

uma vez que

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

um problema equivalente é minimizar

$$\sum_{x \in X} [\bar{y}(x) - f(x)]^2 m(x), \quad (2.3)$$

na classe de funções lineares  $f$  de  $X$  (BARLOW et al., 1972).

Se não existe restrição para  $\mu$ , a estimativa de mínimos quadrados seria obtida por meio da minimização anterior na classe de funções arbitrárias  $f$  de  $X$  e a solução seria a média amostral  $\bar{y}$ . Agora, se é conhecido ou assumido que  $\mu$  é não decrescente em  $X$ , a estimativa de mínimos quadrados de  $\mu$  poderia ser obtida por minimização da equação anterior, na classe de funções não decrescentes  $f : X \rightarrow R$ . Barlow et al. (1972) usam as seguintes definições para *função isotônica e regressão isotônica*:

**Definição 1.** *Seja  $X$  um conjunto finito de números  $\{x_1, \dots, x_k\}$ . Uma função de valor real  $f$  sobre  $X$  é isotônica se para  $x_i, x_j \in X, x_i < x_j$  implica em  $f(x_i) \leq f(x_j)$ .*

**Definição 2.** *Seja  $g$  uma função em  $X$  e  $\omega : X \rightarrow R$ , uma função positiva que será denominada peso. Uma função  $g^*$  é uma regressão isotônica de  $g$  com peso  $\omega$ , se, e somente se,  $g^*$  é uma função isotônica e minimiza*

$$\sum_{x \in X} [g(x) - f(x)]^2 \omega(x), \quad (2.4)$$

em que,  $f$  varia entre todas as funções isotônicas em  $X$ . Portanto,  $g^*$  é uma solução de mínimos quadrados restritos para a expressão (2.4).

Seja  $\mathbf{X} = \{x_1, x_2, \dots, x_k\}$  um conjunto finito. Uma relação binária  $\prec$  em  $X$  é dita uma ordem simples se são satisfeitas as propriedades abaixo:

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

**Definição 3.** *Uma relação binária  $\prec$  em  $X$  estabelece uma ordem parcial em  $X$  se ela satisfaz as condições 1, 2 e 3 acima.*

**Definição 4.** *Uma relação binária  $\prec$  em  $X$  estabelece uma quase ordem em  $X$  se ela satisfaz as condições 1 e 2 acima.*

Como pode ser observado, toda ordem simples é uma ordem parcial e toda ordem parcial é uma quase ordem. Portanto, se a teoria for realizada sobre uma quase ordem ela será mais geral do que se fosse realizada sobre uma ordem simples.

Os teoremas 1 e 2 estão descritos em Barlow et al. (1972).

**Teorema 1.** *A regressão isotônica  $g^*$  de  $g$  com peso  $\omega$  existe e é única. A regressão*

isotônica é a resolução de um problema de otimização quadrático, uma vez que minimiza a soma de quadrados. Tal fato, por si só, não seria suficiente para ser utilizado em problemas estatísticos, uma justificativa para a sua utilização é obtida através do resultado:

**Teorema 2.** *Seja uma família de distribuições indexadas em um grupo quase ordenado,  $\mathbf{X} = \{x_1, \dots, x_k\}$  e se  $\mu(x)$ , a média dessas distribuições, é suposta isotônica para o caso em que as distribuições são: binomiais, geométricas, poisson, normal ou gamma. O estimador de máxima verossimilhança restrito às funções isotônicas é obtido pela regressão isotônica com pesos que dependem da variância, da média amostral  $\bar{y}(x)$ . Um resultado, descrito em Robertson et al. (1988), que ocorre para o caso normal é:*

Se  $\bar{Y}(x_i)$  é a média de uma amostra de tamanho  $m(x_i)$  de uma população normal com média desconhecida  $\mu(x_i)$  e variâncias conhecidas  $\sigma^2(x_i)$  para  $i = 1, \dots, k$ , e assumindo que as amostras das várias populações são independentes e que é conhecido que  $\mu$  é isotônica com respeito a ordem simples,  $x_1, x_2, \dots, x_k$  em  $X$ . Seja  $\bar{Y}^*$  com pesos  $\omega(x) = \frac{n(x)}{\sigma^2(x)}$  então:

**Teorema 3.** Nas condições acima, para cada  $x \in X$ ,

$$E \left[ \left( \bar{Y}(x) - \mu(x) \right)^2 \right] \geq E \left[ \left( \bar{Y}^*(x) - \mu(x) \right)^2 \right],$$

portanto, além de  $\bar{Y}^*$  ser a função isotônica é a mais próxima possível no sentido dos mínimos quadrados da média amostral  $\bar{Y}(x)$ , que é uma propriedade global envolvendo todas as populações normais indexadas pelo conjunto  $X$ , também temos essa propriedade que, para cada população a esperança do erro quadrático de  $\bar{Y}^*$  como estimador de  $\mu(x)$  é menor ou igual à variância do estimador  $\bar{Y}$ .

Acrescentando o fato que a função  $\bar{Y}^*(.)$  é o estimador de máxima verossimilhança da função  $\mu(.)$  temos uma forte justificativa de se usar a isotonização dos dados como um estimador natural para problemas onde é conhecido que a média é uma função crescente em relação a uma quase ordem de  $X$ .



### 3 Material e métodos

Os dados que foram utilizados são referentes ao trabalho desenvolvido no Laboratório de Física do Solo, do Departamento de Engenharia de Biosistemas, da Escola Superior de Agricultura “Luiz de Queiroz”, Esalq/USP, cujo objetivo foi montar um ensaio experimental em laboratório a fim de representar o comportamento do transporte dos teores de potássio, nos solos Latossolo Vermelho Amarelo (LVA) e Nitossolo Vermelho (NV), em que a vazão de entrada e a concentração de aplicação de solutos fossem conhecidas e a partir daí, analisar cada camada de solo separadamente em instantes pré-determinados. Os solos foram coletados na camada de 0 a 0,70 m de profundidade, no município de Piracicaba-SP. Este trabalho dá continuidade as análises propostas em (PEIXOTO, 2013).

O modelo utilizado para descrever o comportamento dos solutos ao longo do perfil do solo, e passaram por isotonização foi logístico com três parâmetros. Esse modelo tem se destacado em algumas aplicações e segundo Pinheiro e Bates (2000), assume a seguinte estrutura não linear

$$f(x_n; \boldsymbol{\theta}) = \frac{\theta_1}{1 + \exp[(\theta_2 - x_n)/\theta_3]}, \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_3), \quad (3.1)$$

em que  $x_n$  é a profundidade (m) associada ao  $n$ -ésimo teor de soluto, denotado por  $y_n$ , para  $n = 1, \dots, N$ , e  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_3)^\top$  é o vetor de parâmetros desconhecidos. Se  $\theta_3 > 0$ , então  $\theta_1$  é a assíntota horizontal quando  $x \rightarrow \infty$  e 0 é o assíntota horizontal quando  $x \rightarrow -\infty$ . Se  $\theta_3 < 0$ , esses papéis são invertidos. O parâmetro  $\theta_2$  é o valor de  $x$  para o qual a resposta é  $\theta_1/2$ . Este é o ponto de inflexão da curva. O parâmetro de escala  $\theta_3$  representa a distância no eixo  $x$  entre o ponto de inflexão e o ponto em que a resposta é  $\theta_1/(1 + e^{-1}) \approx 0,73\theta_1$ .

Para a estimação dos parâmetros do modelo foi utilizada a técnica dos mínimos quadrados ordinários, com o uso do método iterativo de Gauss-Newton (BATES; WATTS, 1988). Os valores iniciais utilizados para obtenção das estimativas dos parâmetros, em cada um dos conjuntos de dados, referentes aos dois tipos de solos, foram encontradas por uma função no *software* R versão 3.2.2 (R DEVELOPMENT CORE TEAM, 2012) que gera os valores iniciais, encontrada na *self-starting nonlinear models* desenvolvido por Pinheiro e Bates (2000).

Para verificar a adequação do ajuste dos modelos aos dados foram utilizadas as técnicas de diagnóstico por meio do resíduo projetado proposto por Cook e Tsai (1985), que seguindo os autores apesar de algebricamente mais complexo, tem propriedades mais próximas das propriedades correspondentes do resíduo ordinário da regressão linear e a distância de cook.

### 3.1 Isotonização de dados de curva de crescimento

Considere as observações  $y_i$  referente aos teores de potássio tomado nas profundidades  $x_i$ ,  $i = 1, 2, \dots, n$  e o modelo de crescimento logístico (3.1). O ajuste deste modelo ao conjunto de dados originais, por meio do método dos quadrados mínimos ordinários consiste em assumir  $y = f(x, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ , em que  $\mathbf{y} = [y_1, y_2, \dots, y_n]$ ,  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  e  $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]$ ,  $\varepsilon_i \sim N(0, \sigma^2)$  e  $\boldsymbol{\theta}$  é o vetor dos parâmetros do modelo.

Os dados  $y_i$ ,  $i = 1, 2, \dots, n$  não estão necessariamente em ordem crescente em razão de variações aleatórias. Uma transformação dos dados para se obter valores crescentes é feita via regressão isotônica, denominada de isotonização de dados. Com este procedimento os valores originais  $\mathbf{y} = [y_1, y_2, \dots, y_n]$  são transformados em valores  $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_n^*]$ , crescentes em relação as profundidades das avaliações. Dessa forma, os modelos de crescimento serão ajustados ao conjunto de dados  $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_n^*]$ .

Rodrigues et al. (2010) afirma que como todo processo de regressão, a curva final é muito influenciada pela ocorrência de *outliers*. A teoria da regressão isotônica é suficientemente flexível no sentido que pode se diminuir o efeito de *outliers* através do uso da regressão isotônica com pesos diferentes. Para determinar pesos adequados foi utilizado o inverso da diferença entre o valor do dado inicial  $y_i$  e o valor correspondente após a isotonização  $y_i^*$ ,  $\omega(x_i) = 1/|y_i - y_i^*|$ , se  $y_i \neq y_i^*$ . Com estes pesos os dados originais são novamente isotonizados obtendo-se novos valores  $\mathbf{y}^{**} = [y_1^{**}, y_2^{**}, \dots, y_n^{**}]$ .

A ideia para o uso de tais pesos é que se  $y_i^*$  está longe de  $y_i$ , isto é  $y_i$  é um valor discrepante em relação ao conjunto dos dados, o peso  $\omega(x_i)$  será relativamente pequeno o que diminuirá sua influência na obtenção dos dados isotonizados, impedindo assim que estes *outliers* afetem por demais o ajuste dos dados.

A regressão isotônica foi implementada pelo *software* R versão 3.2.2 (R DEVELOPMENT CORE TEAM, 2012) que apresenta uma implementação de forma mais simples e compacta. Neste *software*, o pacote requerido para a consideração de pesos iguais é o *isotone* e a função usada foi a *isoreg* e a função *cir.pava* do pacote *cir* do *software* R para a regressão isotônica com pesos diferentes. Neste trabalho realizou-se o ajuste da regressão antitônica, cujo pacote requerido foi o *fdrtool* e a função utilizada foi a *monoreg*, considerando pesos iguais.

## 4 Resultados e discussão

As estimativas encontradas para os parâmetros do modelo obtidas por meio de aproximação assintótica, bem como os valores do erro padrão das estimativas, o valor  $p$  e os intervalos de confiança a 95%, para o teor de potássio encontrados ao longo do perfil do solo são apresentados na Tabela 1. Observa-se que todos os parâmetros foram significativos, pois se encontram dentro de intervalo de confiança assintótico de 95% e estes por sua vez, não incluem a constante zero, verificando-se que o modelo ajustado é adequado.

Tabela 1 – Estimativas dos parâmetros do modelo logístico com três parâmetros, erro padrão da estimativa (E.P.E.), valores  $p$  para o teste  $t$  e intervalos de confiança (IC) de 95%, para o teor de potássio ao longo da profundidade nos solos LVA e NV.

Solo	Parâmetros	Estimativas	E.P.E.	Valor $p$	IC (95%)
LVA	$\theta_1$	1621,0000	32,2300	<,0001	(1554,6328; 1686,9109)
	$\theta_2$	0,3289	0,0046	<,0001	(0,3193; 0,3384)
	$\theta_3$	-0,0470	0,0039	<,0001	(-0,0550; -0,0389)
NV	$\theta_1$	1926,0000	179,7000	<,0001	(1556,8321; 2294,1916)
	$\theta_2$	0,1443	0,0134	<,0001	( 0,1167; 0,1718)
	$\theta_3$	-0,0552	0,0065	<,0001	(-0,0686; -0,0418)

Na Figura 1 são apresentados os ajustes dois tipos de solos. Dos 2000 mg  $L^{-1}$  aplicados de potássio, os valores máximos encontrados para os dois solos foram semelhantes e se mantiveram próximos a 1600 mg  $L^{-1}$ . A maior diferença entre as curvas do LVA e do NV é notada por meio dos valores de concentração relativos aos cinco primeiros anéis, demonstrando que a textura dos solos diferenciou os perfis de potássio principalmente quanto à sua profundidade de alcance na coluna.

Após o ajuste dos modelos aos dados, procedeu-se algumas técnicas de diagnósticos para a validação do modelo. Primeiramente realizou-se o teste de Shapiro-Wilk para os resíduos do modelo ajustado. Deste modo a partir dos resultados do teste de Shapiro-Wilk apresentados a seguir, conclui-se que não há evidências para dizer que a distribuição dos resíduos não seja normal, considerando o nível de significância de 5%, pois o  $v p = 0,6252 > 0,05$  para o solo LVA e  $v p = 0,0762 > 0,05$  para os solo NV, não rejeitando a hipótese de normalidade dos resíduos para o modelos ajustados.

Suspeita-se da presença de valores discrepantes nos dados observados, conforme o gráfico de resíduos, convém ressaltar, no entanto, que o número de observações é muito pequeno e assim, mesmo supondo que a distribuição dos erros seja normal, a probabilidade de aparecerem valores discrepantes é grande. Como pode-se observar, existe para o solo LVA um ponto que merece atenção (observação 15), e no solo NV existe uma aparente

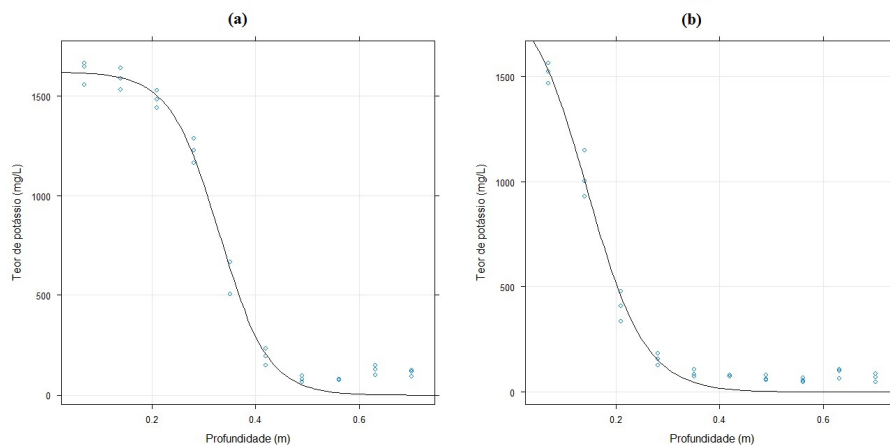


Figura 1 – Ajuste do teor de potássio em relação a profundidade nos solos LVA (a) e NV (b) para os logístico com 3 parâmetros.

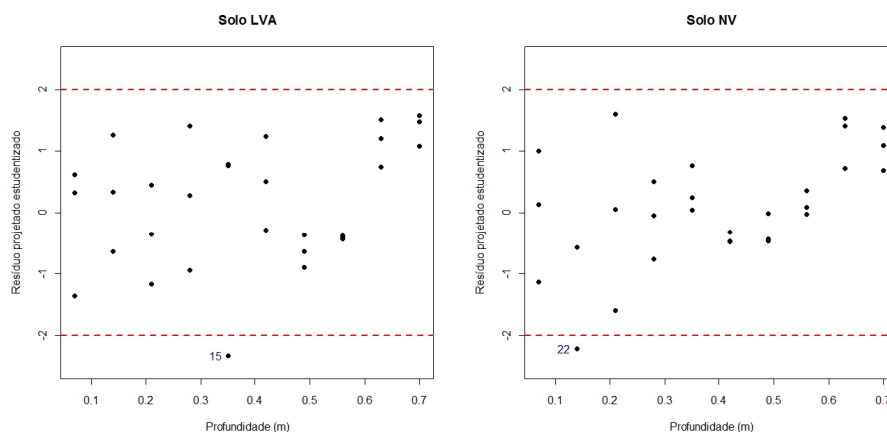


Figura 2 – Gráfico dos resíduos projetados para o teor de potássio nos solos LVA e NV ao longo da profundidade.

relação entre as observações e a profundidade, com uma observação que pode ser um possível *outlier* (observação 22). Porém, ao se fazer o ajuste sem estas observações, não houve alteração no ajuste do modelo e as diferenças entre os  $\theta$ 's, não foi maior do que 1 (em módulo), fato que não justifica a remoção destas observações (Figura 2).

Na Figura 3, as observações (4, 15) no solo LVA e (2, 11, 22 e 23) no solo NV, apesar de estarem em destaque em relação massa de dados, não são observações influêntes sobre todos os  $n$  valores ajustados, de modo que  $D_i$  foram menores que a unidade absoluta ( $< 1$  em módulo).

A proposta de isotonização de dados referentes a estudos envolvendo curvas de crescimento, cujo objetivo principal foi melhorar a porcentagem de convergência e qualidade dos ajustes em situações que consideram curvas atípicas. O procedimento de correção dos dados originais via regressão antitônica, suavizou os pontos em torno da curva (Figura 4), porém não apresentou mudanças expressivas nos valores encontrados. Rodrigues et al.

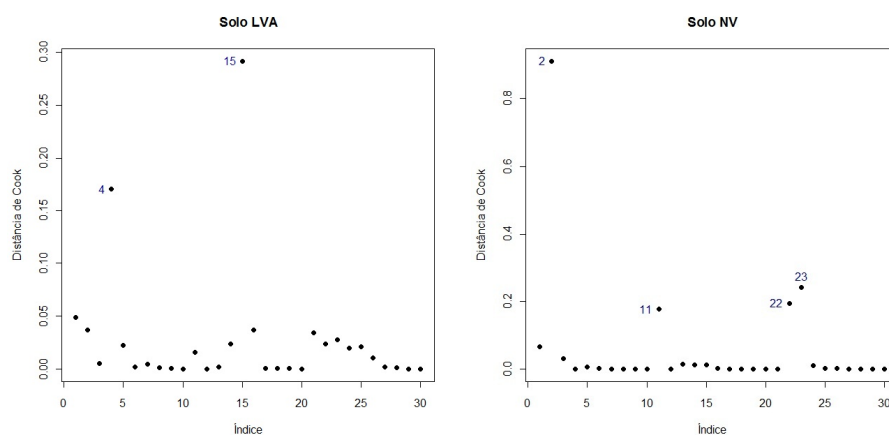


Figura 3 – Gráfico das distâncias de Cook e elementos da diagonal da matriz de projeção *versus* a ordem das observações para do teor de potássio nos solos LVA e NV ao longo da profundidade.

(2010) utilizaram uma transformação de dados via análise de regressão isotônica, para estudos de curvas de crescimento e verificaram que o uso desta metodologia possibilitou aumentar a porcentagem de convergência. Propostas semelhantes já foram aplicadas com sucesso na literatura estatística, como a de Hussian et al. (2004), que envolveu ajuste de modelos de regressão não linear para descrever a concentração média de nitrogênio em função do tempo.

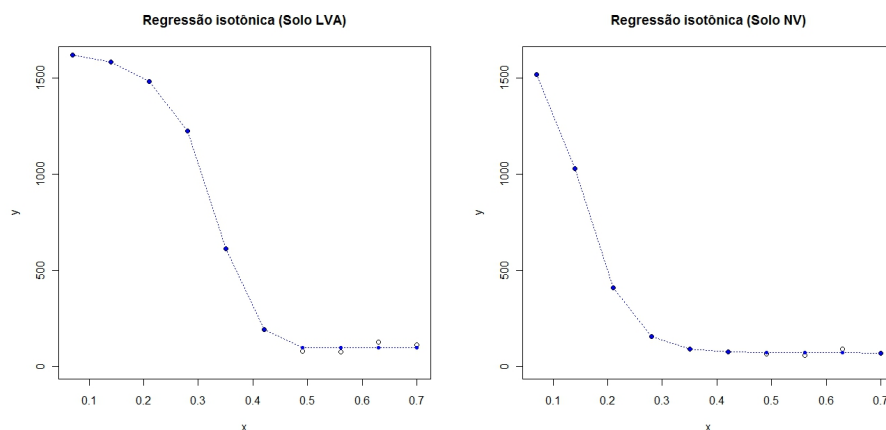


Figura 4 – Ajuste da regressão isotônica do teor de potássio nos solos LVA e NV ao longo da profundidade.

Nas Figuras 5 e 6 é possível verificar o gráfico dos pontos médios dos dados no ajuste dos pontos de dados originais em função dos pesos iguais. Vários valores  $x$  idênticos são dados como entrada, os valores  $y$  correspondentes e os pesos  $w$  são automaticamente incorporadas ao modelo. Também é possível visualizar o diagrama da soma cumulativa com a minorante mesmo côncava (regressão antitônica). Brito et al. (2012), utilizaram a transformação de dados via análise de regressão antitônica, possibilitando aumento da

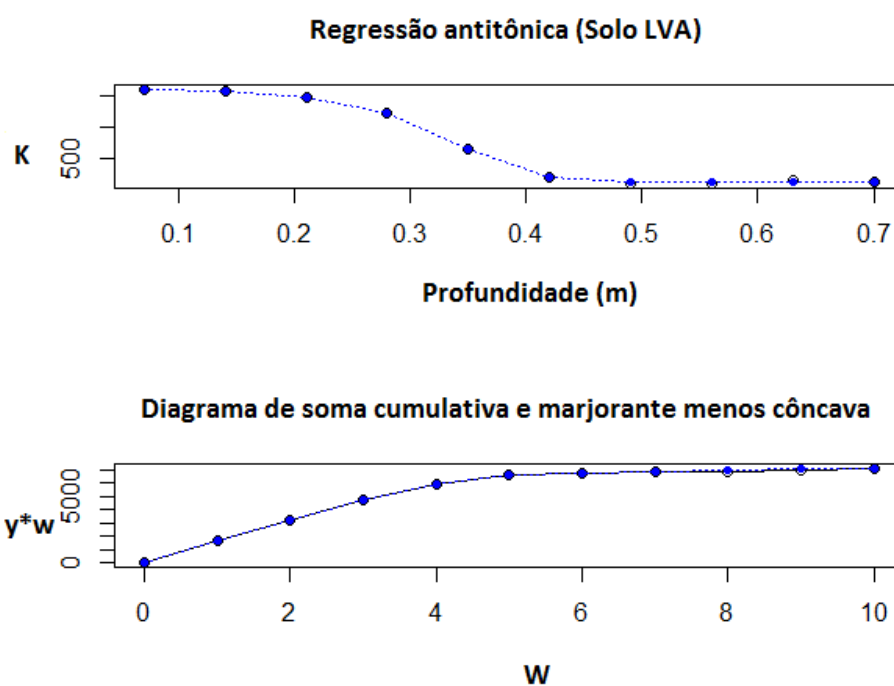


Figura 5 – Ajuste da regressão antitônica e diagrama da soma cumulativa com menor majorant côncava do teor de potássio no solo LVA ao longo da profundidade.

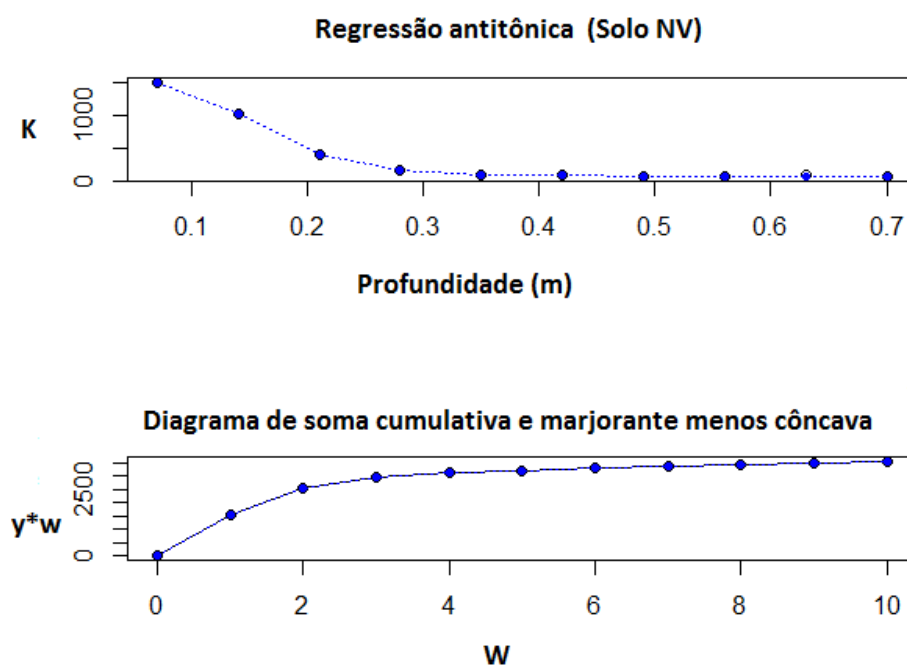


Figura 6 – Ajuste da regressão antitônica e diagrama da soma cumulativa com menor majorant côncava do teor de potássio no solo NV ao longo da profundidade.

qualidade no ajuste dos métodos da máxima curvatura modificado e do modelo linear com resposta platô.

## 5 Considerações finais

É possível destacar a importância do ajuste do modelo logístico para representar a dinâmica do solo. O referido modelo mostrou-se apropriado no ajuste do teor de potássio presente no perfil nos dois tipos de solos analisados, possibilitando inferir sobre os aspectos que constituem a dinâmica do solo, como o transporte dos solutos.

A metodologia de transformação de dados via análise de regressão isotônica com pesos iguais, possibilitou aumentar a porcentagem de convergência e a qualidade dos ajustes dos modelos de regressão não linear Logístico a dados de crescimento que apresentam distúrbios caracterizados por decréscimos dos teores do soluto ao longo do tempo.



## 6 Referências Bibliográficas

- BATES, D.M.; WATTS, D.G. Relative curvature measures of nonlinearity (with discussion). **Journal of the Royal Statistical Society**, Serie B, Methodological, London, v.42, n.1, p.1-25, 1980.
- BATES, D.M.; WATTS, D.G. **Nonlinear regression analysis and its applications**. New York: John Wiley & Sons. 1988. 365p.
- BARLOW, R. E. **Statistical inference under order restrictions: the theory and application of isotonic regression**. London: J. Wiley, 1972. 388 p.
- BRITO, M. C. M.; FARIA, G. A.; MORAIS, A. R. SOUZA, E. M.; DANTAS, J. L. L. Estimation of optimum plot size using antitonic regression. **Rev. Bras. Biom.**, São Paulo, v.30, n.3, p. 353-366, 2012.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3.ed. New York: John Willey & Sons, 1998. 706p.
- COOK, R.D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman & Hall, 1982. 280p.
- COOK, R.D.; TSAI, C.L. Residuals in Nonlinear Regression. **Biometrika**. London, v. 72, p. 23-29, 1985.
- CORDEIRO, G.M.; PAULA, G.A. Modelos de regressão para análise de dados univariados. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 17., 1989, Rio de Janeiro. **Resumos ...** Rio de Janeiro: IMPA, 1989.
- COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society**, B, Oxford, v.30, p.248-275, 1968.
- GALLANT, A.R. **Nonlinear Statistical Models**. New York: John. Wiley, 1987. 624p. (Wiley series in probability e mathematical statistics).
- 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.
- KRAUSS, A. Potassium effects on yield quality. In: YAMADA, T.; ROBERTS, T.L. (Eds.). **Potássio na agricultura brasileira**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 2005. cap.11, p.281-299.
- MCCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd ed. London: Chapman and Hall, 1989. 511p.
- MIRANDA, J.H.; DUARTE, S.N.; LIBARDI, P.L.; FOLEGATTI, M.V. Simulação do deslocamento de potássio em colunas verticais de solo não-saturado. **Engenharia Agrícola**, Jaboticabal, v.25, n.3, p.677-685, 2005.
- NACHTIGALL, G.R.; VAN RAIJ, B. Análise e interpretação do potássio no solo. In: YAMADA, T.; ROBERTS, T.L. (Eds.). **Potássio na agricultura brasileira**. Piracicaba: **Associação Brasileira para Pesquisa da Potassa e do Fosfato**, 2005. cap.5, p.93-118.
- PEIXOTO, A. P. B. **Análise da dinâmica do potássio e nitrato em colunas de solo não saturado por meio de modelos não lineares e multiresposta**. Tese (Doutorado) - Escola

- Superior de Agricultura Luiz de Queiroz, Piracicaba, 2013.
- PINHERO, J.C.; BATES, D.M. **Mixed-Effects Models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528p.
- R Development Core Team (2012). R: A language and environment for statistical computing. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 jan. 2014.
- REICHERT, J.M; REINERT D.J.; BRAIDA, J.A. Manejo, qualidade do solo e sustentabilidade: condições físicas do solo agrícola. CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 29, 2003, Ribeirão Preto. Palestras. Ribeirão Preto: SBCS, 2003 (CD-ROM).
- RITZ, C.; STREIBIG, J.C. **Nonlinear regression with R**. New York: Springer, 2008. 148p.
- ROBERTSON, T.; WRIGHT, F. T.; DYKSTRA, R. L. **Order restricted statistical inference**. New York: J. Wiley, 1988. 521p.
- RODRIGUES, A. et al. Utilização da regressão isotônica em estudos de curvas de crescimento. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 4, p. 211-223, out./dez. 2010.
- SOUZA, G. S. **Introdução aos modelos de regressão linear e não linear**. Brasília: EMBRAPA-SPI/Embrapa-SEA, 1998. 489p.
- SOUZA, E. M. **Teoria da regressão isotônica e delineamentos sequenciais aplicados a ensaios de dose-resposta**, 2011. 411f. Tese (Doutorado em Estatística e Experimentação Agrônoma) - Universidade Federal de Lavras, Lavras, 2011.
- SOUZA, E. M.; CHAVES, L. M.; MUNIZ, J. A. Regressão isotônica aplicada à análise probit em ensaios de dose-resposta sequenciais. **Rev. Bras. Biom.**, São Paulo, v.29, n.1, p.122-146, 2011.
- SOUZA, E. M. de. CHAVES, L. M. MUNIZ, J. A. Aspectos geométricos da regressão isotônica e uma aplicação à estimação por máxima verossimilhança para a família exponencial. **Rev. Bras. Biom.** v.30, n.1, p.76-105. 2012.
- TÓTOLA, M.R.; CHAER, G.M. Microrganismos e processos microbiológicos como indicadores da qualidade dos solos. Tópicos em Ciências do Solo, **Sociedade Brasileira de Ciências do Solo**, Viçosa, v.2, p.195-276, 2002.
- WEISBERG, S. **Applied linear regression**. 3ed. Hoboken: J. Wiley, 2005. 352p.
- LEEuw, J.; HORNIK, K.; MAIR, P. Isotone optimization in R: Pool - adjacent - violators algorithm (pava) and active set methods. **J. Stat. Softw.**, Los Angeles, v.32, n.5, p.1-24, 2009.
- EEDEN, C. V. Estimation in restricted parameter spaces: some history and some recent developments. **Quarterly**, Amsterdam, v. 9, n. 1, p. 69-76, Mar. 1996.
- RATKOWSKY, D.A. **Nonlinear Regression Modeling: a Unified Practical Approach**. New York: Marcel Dekker, 1983.
- PEIXOTO, A. P. B. **Análise da dinâmica do potássio e nitrato em colunas de solo não saturado por meio de modelos não lineares e multiresposta**. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2013.
- HARTLEY, H.O.; BOOKER, A. Nonlinear least squares estimation. **The Annals of Mathematical Statistics**, Iowa, v.40, p.633-643, 1965.
- McKEOWN, J. J. Large residual nonlinear least-squares problems in nonlinear optimization theory and algorithms. **Birkhauser**, Boston, p.92-105, 1980.

# Apêndices

# APÊNDICE A – Rotina R utilizada para obtenção do ajuste do modelo logístico

```

### Ajuste do modelo logístico com 3 parâmetros

## Teor de Potássio - Solo LVA

rm(list=ls(all=TRUE))

# pacotes necessários para as análises -----
require(nlstools)
require(MASS)
require(moments)
require(car)
require(proto)
require(stats)
require(nlme)
require(nls2)
require(NRAIA)
library(qpcR)

# x=profundidade (m)
# y=teor de potássio (mg\L)

x=c(0.07,0.14,0.21,0.28,0.35,0.42,0.49,0.56,0.63,0.7,
0.07,0.14,0.21,0.28,0.35,0.42,0.49,0.56,0.63,0.7,
0.07,0.14,0.21,0.28,0.35,0.42,0.49,0.56,0.63,0.7)

y=c(1555.5,1530.4,1526.8,1286.1,667.0,193.97,78.4,129.7,123.7,
1647.6,1587.1,1484.1,1164.6,504.8,150.9,66.1,75,148.5,92.5,
1663.5,1641.4,1440.9,1227.5,665.7,232.5,80.9,76.6,101.8,117)

liber <- data.frame(x, y)

## Obtenção dos valores iniciais

getInitial(y ~ SSlogis(x,theta1,theta2, theta3), liber)

# procedimentos para o ajuste do modelo logistico =====
# calcular as derivadas -----
quoc.der <- deriv3(~theta1/(1 + exp((theta2 - x)/theta3)),
c("theta1", "theta2", "theta3"), function(theta1, theta2, theta3, x) NULL)
# ajuste do modelo -----
modelo <- nls(y~quoc.der(theta1, theta2, theta3, x),
data=liber, start=c(theta1=1620.7733, theta2=0.3289, theta3=-0.0470))
summary(modelo)

```

```
# Gráfico do modelo ajustado
plotfit(modelo, xlab = "Perfil (m)", ylab = "Concentração de potássio (mg/L)", main = "(c)")

# Critérios de informação AIC e BIC
AIC(modelo)
BIC(modelo)

# Coeficiente de determinação
Rsquared(modelo) # R2 pelo pacote qpcR
Rsquared.ad(modelo) # R2 ajustado pelo pacote qpcR

# teste das pressuposições da análise -----
shapiro.test(residuals(modelo))

qqmath(residuals(modelo), col=1,
prepanel=prepanel.qqmathline,
panel=function(x, ...){
panel.qqmathline(x, lty=2, ...)
panel.qqmath(x, ...)})

# matriz de covariância e correlação das estimativas -----
vc <- vcov(m0q);vc
cr <- solve(t(diag(sqrt(diag(vc)))))%*%vc%*%solve((diag(sqrt(diag(vc)))));cr

# intervalo de confiança assintótico -----
sm <- summary(modelo)$coef
cbind(sm[,1]-sm[,2]*qt(0.975, df=df.residual(modelo)),sm[,1]+sm[,2]*qt(0.975,
df=df.residual(modelo)))

## Análise de resíduo

## Parâmetros
theta1=coef(modelo)[1];theta1
theta2=coef(modelo)[2];theta2
theta3=coef(modelo)[3];theta3
p<-length(coef(modelo));p
n<-length(y);n

## Primeira e segunda derivada de mu em relação ao vetor de parâmetros Beta
g<- expression(theta1/(1 + exp((theta2 - x)/theta3))) # inserir o modelo
g<-deriv(g, c("theta1","theta2","theta3"), hessian = TRUE)
X<-attr(eval(g),"gradient") # X til
Xt=t(X)
XtX<- crossprod(X)
W<-attr(eval(g),"hessian") # p matrizes n x p
Wn<-matrix(W,n*p,p) # matriz W com dimensão (n*p) x p

## Matriz de projeção
H<-X%*%solve(XtX)%*%Xt
h<-diag(H)
r<-summary(m0q)$resid ;r ## Resíduo ordinário
s<-summary(m0q)$sigma;s ## Desvio padrão
```

```

t<-r/(s*sqrt(1-h));t      ## Resíduo estudentizado

## Resíduo projetado
I=diag(n)
xij<-matrix(0,n,1)
T<-matrix(0,n,p*p)
for (j in 1:(p*p))
{
l=(1+((j-1)*n)):(n+((j-1)*n))
xij<-W[l]
T[,j]= as.vector(xij)
}
(I-H)%*%T #Verificar quais são zero ou aprox. zero e retirar
Tf=cbind(T[,5],T[,6],T[,9]) # não-nulos
S<-(I-H)%*%Tf
St=t(S)
StS<- crossprod(S)
H1<-S%*%solve(StS)%*%St
V<-cbind(X,S)
Vt=t(V)
VtV<- crossprod(V)

H2<-V%*%solve(VtV)%*%Vt
trH2<-sum(diag(I-H2))
sigma2=(t(r)%*%(I-H2)%*%r)/trH2
sigma=sqrt(sigma2)
rp<-((I-H2)%*%r)/(sigma*diag(sqrt(I-H2)))
## Gráfico dos resíduos projetado estudentizados
plot(x, rp, pch=19, xlab="Perfil (m)", ylab="Resíduo projetado estudentizado",
ylim=c(-2.5,2.5),main='Solo LVA')
abline(-2,0,lty=2, col="red", lwd=2)
abline(2,0,lty=2, col="red", lwd=2)
identify(x, rp, cex=1, col="darkblue")

## Distância de Cook
D<-(t^2/p)*(h/(1-h))
## Gráfico da distância de Cook
plot(D, pch=19, xlab="Índice", ylab="Distância de Cook",main='Solo LVA')
identify(D, cex=1, col="darkblue")
## Gráfico dos elementos da diagonal da matriz de projeção H
plot(h, xlab="Índice", ylab="Diagonal de H", pch=19, ylim=c(0,0.5),main='Solo LVA')
abline((2*p)/n,0,lty = 2, lwd = 2, col="red")
identify(h, cex=1, col="darkblue")

##### Regressão Isotônica #####
# Pacotes
library(isotone)
library(graphics)
library("fdrtool")
library(Iso)

```

```
## Média das três repetições
x=c(0.07,0.14,0.21,0.28,0.35,0.42,0.49,0.56,0.63,0.7)
y=c(1622.20,1586.30,1483.93,1226.07,645.83,173.13,81.33,74.67,126.67,107.73)

#Ajuste da regressão antitônica
mr=monoreg(x, y, w=rep(1, length(x)), type=c("antitonic"));mr
plot(mr, ylim=c(0, 1600)) # this shows the averaged data points
points(x, y, pch=2) # add original data points
plot(mr, plot.type="row.wise")
```