

Algumas considerações em regressão não linear

Josmar Mazucheli¹ e Jorge Alberto Achcar²

¹Departamento de Estatística, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brasil.

²Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, São Paulo, Brasil. *Autor para correspondência.

RESUMO. Diferente dos modelos de regressão lineares, em que a qualidade e principalmente a validade do ajuste são simplesmente avaliadas por meio de diagnósticos de regressão, no caso não linear, além de diagnósticos usuais, outros procedimentos devem ser seguidos. Esses procedimentos, particulares dos modelos de regressão não lineares, são úteis na avaliação da extensão do comportamento não-linear. Modelos não-lineares com comportamento distante do comportamento linear podem ter seus resultados assintóticos invalidados, principalmente em situações em que pequenas amostras são disponíveis. Considerando a importância de se avaliar a extensão do comportamento não-linear de modelos de regressão não-lineares, neste artigo são apresentadas as principais medidas de não-linearidade discutidas na literatura. Em particular, consideram-se as Medidas de Curvatura de Bates e Watts, a Medida de Vício de Box, Estudo de Simulação e Medidas de Assimetria.

Palavras-chave: modelos de regressão não linear, medidas de curvatura, medida de vício, métodos de simulação, medidas de assimetria.

ABSTRACT. Considerations about nonlinear regression. Differently from linear regression models, where the quality and validity of a proposed model are checked through regression diagnostic methods, for nonlinear models, we should consider additional diagnostic methods. These special diagnostic methods for nonlinear regression models measure the nonlinearity amount of the proposed model. Nonlinear models with behavior too different from linear models could invalidate the asymptotical results, used to get interest inferences, especially in small-sized samples. In this paper, an overview of the main nonlinearity checking methods is offered. In particular, Bates and Watts curvature measures, Box bias measure, simulation methods use and some nonsymmetry measures are focused.

Key words: nonlinear regression models, curvature measures, bias measure, simulation methods, nonsymmetry measures.

Introdução

Avaliar a possível relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns em análise estatística. Pode-se atingir este objetivo por meio dos bem conhecidos modelos de regressão, os quais se dividem em duas classes distintas: os lineares e os não-lineares.

Dentre as muitas diferenças existentes entre essas duas classes de modelos, a principal está relacionada as suas formulações. No caso linear, a partir de um conjunto de observações, busca-se o modelo que melhor explique a relação, se existir alguma, entre as variáveis inerentes a um dado fenômeno. Por exemplo, se a resposta de interesse, usualmente

representada por y , depender de uma única variável independente, x , a partir da representação gráfica de x versus y , pode-se sugerir possíveis modelos. Na presença de várias variáveis independentes, uma alternativa para o ajuste de um possível modelo é partir, inicialmente, de um modelo completo e avaliar a qualidade do ajuste por meio de diagnósticos de regressão (Belsley *et al.*, 1980; Cook e Weisberg, 1982). A classe de modelos lineares é bastante flexível, uma vez que muitos modelos podem ser formulados (Draper e Smith, 1981; Seber, 1977). Já no caso não-linear, na maioria das vezes, as formulações de possíveis modelos são baseadas em considerações teóricas inerentes ao fenômeno que se tem interesse modelar. Modelos formulados desta forma são chamados de modelos

mecanísticos (Seber e Wild, 1989; Bates e Watts, 1988).

Neste artigo, são apresentados os conceitos fundamentais da classe dos modelos de regressão não-lineares. Os problemas relacionados à estimação, em particular no método de mínimos quadrados, propriedades assintóticas e conceitos de não-linearidade, também são tratados. Resultados de inferência estatística não são apresentados ou discutidos ao longo do artigo basicamente por dois motivos: primeiro por não fazer sentido falar em inferências utilizando modelos não-lineares, mesmo que as suposições básicas usuais estejam satisfeitas, antes da avaliação da extensão do comportamento não-linear, que pode invalidar as inferências assintóticas (previsões, intervalos de confiança etc.); segundo porque as inferências são baseadas nos mesmos princípios dos modelos lineares.

Formulação dos modelos de regressão não-lineares

Por definição, um modelo de regressão é não-linear se pelo menos um dos seus parâmetros aparecem de forma não-linear. Por exemplo, os modelos:

$$E(y) = \exp(\theta_1 + \theta_2 x), \quad (1)$$

$$E(y) = \theta_1 + \theta_2 \exp(-\theta_3 x), \quad (2)$$

$$E(y) = (\theta_1 + \theta_2 x)^{-1}, \quad (3)$$

$$E(y) = (\theta_1 - \theta_2)^{-1} [\exp(-\theta_1 x) + \exp(-\theta_2 x)] \quad (4)$$

são todos não-lineares e o operado $E(\cdot)$ denota a função esperança ou função de regressão. No modelo (1), os parâmetros θ_1 e θ_2 são não-lineares. No modelo (2), θ_1 e θ_2 são lineares enquanto que θ_3 é não-linear. Já nos modelos (3) e (4), ambos os parâmetros θ_1 e θ_2 são não-lineares.

Um modelo de regressão não-linear é considerado “intrinsecamente linear” se este pode ser reduzido a um modelo linear por meio de uma reparametrização apropriada. Pode-se, ainda, usar o termo “intrinsecamente linear” para referir-se a modelos que podem ser linearizados via alguma transformação. Por exemplo, aplicando-se logaritmo em ambos os membros da equação (1) pode-se reduzi-la à forma linear. Em geral, na prática, um modelo não-linear é linearizado para facilitar a obtenção das estimativas dos parâmetros. O inconveniente de uma transformação é que, além do parâmetro perder sua interpretação intrínseca, pode-se alterar a estrutura e distribuição do erro, ou seja, se os erros do modelo original satisfizerem as

suposições usuais de normalidade, independência e homogeneidade da variância, os erros do novo modelo, em geral, não satisfarão tais suposições (Khuri e Cornell, 1987). Caso não seja possível obter uma reparametrização ou uma transformação apropriada, que reduza o modelo a forma linear, tem-se os chamados modelos “intrinsecamente não-lineares”.

O método de estimação por mínimos quadrados

O método de estimação por mínimos quadrados é usado na análise de dados em que as observações são constituídas por variáveis resposta y_i obtidas em diferentes níveis da variável independente x_i , ($i=1, \dots, n$). Assume-se que a relação variável resposta/variável independente pode ser adequadamente representada por uma equação da forma:

$$y = f(\mathbf{x}; \boldsymbol{\theta}) + \boldsymbol{\varepsilon} \quad (5)$$

em que $\mathbf{y} = (y_1, \dots, y_n)^t$ e $\mathbf{x} = (x_1, \dots, x_n)^t$ são os vetores de variáveis resposta e variável explicativa, respectivamente, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^t$ é o vetor de parâmetros desconhecidos, $f(\mathbf{x}; \boldsymbol{\theta}) = (f(x_1; \boldsymbol{\theta}), \dots, f(x_n; \boldsymbol{\theta}))^t$ e uma função das variáveis regressoras e dos parâmetros chamada de função esperança ou função de regressão e $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^t$ é o vetor de erros aleatórios. Em geral, por motivos de inferência, assume-se que os erros são variáveis aleatórias independentes e identicamente distribuídas, normais com média $\mathbf{0}$ e variância constante $\sigma^2 \mathbf{I}_n$, em que \mathbf{I}_n representa a matriz identidade de ordem n .

Considerando que a função esperança é contínua e admite derivadas de primeira e segunda ordens com relação aos parâmetros, define-se a soma de quadrados dos erros de (5) por

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i; \boldsymbol{\theta})]^2. \quad (6)$$

Pode-se notar que após a realização de um dado experimento (x_i, y_i) , são observações fixas e conhecidas, portanto $S(\boldsymbol{\theta})$ 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 (6) com relação a cada θ_j ($j=1,\dots,p$).

Procedendo desta forma, determina-se p equações, chamadas de equações normais ou homogêneas, que são especificadas por:

$$\sum_{i=1}^n [y_i - f(x_i; \boldsymbol{\theta})] \left[\frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad (7)$$

e quando $\partial f(x_i; \boldsymbol{\theta})/\partial \theta_j$ não depender de $\boldsymbol{\theta}$, ou $\partial^2 f(x_i; \boldsymbol{\theta})/\partial \theta_j^2 = 0$, 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 $f(x_i; \boldsymbol{\theta})$. Em modelos multiparamétricos, as soluções das equações normais podem ser extremamente difíceis de serem obtidas e algum método iterativo de resolução de equações normais não-lineares deve ser utilizado na maioria dos casos (Bates e Watts, 1988; Ratkowsky, 1990).

A obtenção das estimativas de mínimos quadrados

Vários métodos iterativos são propostos na literatura para obtenção das estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não-linear. 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 e Watts, 1988). Esses métodos fazem uso das derivadas parciais da função esperança $f(x_i; \boldsymbol{\theta})$ com relação a cada parâmetro. Essa característica pode restringir suas aplicações, uma vez que, em geral, a função esperança é bastante complexa. Uma alternativa é calcular as derivadas numericamente, como, por exemplo, através de diferenças finitas (Hartmann, 1994) ou usar o método chamado D.U.D. (*doesn't use derivatives*), desenvolvido por Ralston e Jennrich (1978). O método D.U.D é bastante similar ao método de Gauss-Newton, exceto pelo fato de não exigir a especificação das derivadas parciais da função esperança. A grande maioria dos softwares estatísticos possuem rotinas com a implementação desses métodos, sendo que os mesmos produzem estimativas bastante similares e, em geral, são de rápida convergência (Bates e Watts, 1988). É evidente que a rapidez na convergência depende da complexidade do modelo em estudo e, principalmente, da qualidade dos valores iniciais, necessários em qualquer método iterativo. Ratkowsky (1983) discute procedimentos para obtenção de bons valores iniciais para algumas

classes de modelos (modelos de crescimento, modelos de regressão assintóticos etc.). Lawton e Silvestre (1971) introduziram um método através do qual valores iniciais são necessários apenas para os parâmetros que aparecem de forma não-linear. Esse procedimento restringe-se aos modelos "parcialmente não lineares". Outra alternativa para obtenção de bons valores iniciais é supor que os parâmetros estão compreendidos em um dado intervalo e a partir de várias combinações calcula-se a soma de quadrados dos erros. A combinação que resultar em menor soma de quadrados pode, então, ser usada como estimativas iniciais razoáveis. Um inconveniente nesse procedimento é que o tempo computacional pode ser bastante elevado, principalmente se o modelo for especificado a partir de muitos parâmetros.

Como observação, esses métodos de obtenção das estimativas de mínimos quadrados não-lineares estão disponíveis no procedimento NLP (NonLinear Programming) do software SAS (*Statistical Analysis System*) (Hartmann, 1994).

Algumas propriedades assintóticas dos estimadores de mínimos quadrados

Considerando como verdadeiras as suposições de que os erros, em (5), são variáveis aleatórias independentes e identicamente distribuídas, normais com média zero e variância constante σ^2 , pode-se fazer certas considerações a respeito dos estimadores de mínimos quadrados. A validade dessas suposições podem ser avaliadas através dos chamados diagnósticos de regressão.

Em modelos de regressão lineares, os estimadores de mínimos quadrados são não-viciados, normalmente distribuídos, e ainda possuem variância mínima possível entre qualquer outra classe de estimadores. Essas propriedades são aceitas como as melhores propriedades que uma classe de estimadores podem apresentar (Searle, 1971). Podem existir outros estimadores, também não viciados, mas esses são menos eficientes no sentido de suas variâncias excederem a variância dos estimadores de mínimos quadrados.

Já no caso não-linear, essas propriedades somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Em geral, em pequenas amostras, essas propriedades são desconhecidas (Jennrich, 1969).

Segundo Ratkowsky (1983), o tamanho da amostra necessário para os resultados assintóticos terem validade depende fundamentalmente do modelo em estudo. Pode existir modelos que, mesmo com amostras consideradas muito grandes,

os resultados assintóticos não são válidos. Entretanto, afirma-se que, à medida que o tamanho da amostra aumenta, os resultados assintóticos vão se tornando mais aplicáveis. Quando os estimadores de mínimos quadrados apresentarem um pequeno vício, distribuição próxima da normal e verdadeiras variâncias próximas daquelas dadas pela matriz de variâncias-covariâncias assintótica (Seber e Wild, 1989) pode-se afirmar que os estimadores de mínimos quadrados exibem um comportamento próximo do comportamento linear. Quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos e, conseqüentemente, mais confiáveis serão as inferências. Avalia-se a extensão do comportamento não-linear através das bem conhecidas medidas de não-linearidade.

A seguir, é apresentada uma aproximação linear da função esperança, na qual baseiam-se os resultados assintóticos de regressão não-linear. É apresentado, também, o algoritmo de Gauss-Newton, o qual é um dos métodos mais utilizados para obtenção de estimativas de mínimos quadrados não-lineares.

Aproximação linear

Para facilitar a compreensão dos resultados apresentados a seguir, considera-se a seguinte notação:

$$\mathbf{f}(\boldsymbol{\theta}) = (f_1(x_1; \boldsymbol{\theta}), \dots, f_n(x_n; \boldsymbol{\theta}))^t,$$

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^t} = \left[\left[\frac{\partial f_i(x_i; \boldsymbol{\theta})}{\partial \theta_j} \right] \right],$$

$$\mathbf{F} = \mathbf{F}(\boldsymbol{\theta}^*).$$

Como comentado anteriormente, os resultados em regressão não-linear somente são válidos assintoticamente, sendo esses baseados em uma aproximação linear, de primeira ordem, em série de Taylor da função esperança em torno de uma vizinhança de $\boldsymbol{\theta}^*$ (considere $\boldsymbol{\theta}^*$ como sendo o verdadeiro valor do parâmetro $\boldsymbol{\theta}$). Essa aproximação linear é dada por:

$$f(x_i; \boldsymbol{\theta}) \approx f(x_i; \boldsymbol{\theta}^*) + \sum_{j=1}^p \left[\frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_j - \theta_j^*), \quad (8)$$

ou, em notação matricial:

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \quad (9)$$

Usando (8), pode-se escrever a soma de quadrado dos erros, definida em (6), na forma:

$$\begin{aligned} S(\boldsymbol{\theta}) &= [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]^t [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})] \\ &= \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|^2 \\ &\approx \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^*) - \mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 \\ &= \|\mathbf{z} - \mathbf{F}(\boldsymbol{\theta})\|^2, \end{aligned} \quad (10)$$

em que, $\mathbf{z} = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^*) = \boldsymbol{\theta}$ e, $\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$. Utilizando propriedades de modelos lineares, tem-se que (10) será minimizada quando $\boldsymbol{\beta}$ for dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^t \mathbf{F})^{-1} \mathbf{F}^t \mathbf{z}. \quad (11)$$

Quando o tamanho da amostra n for suficientemente grande e sob certas condições de regularidade (Seber e Wild, 1989), $\hat{\boldsymbol{\theta}}$ estará quase certamente em uma vizinhança de $\boldsymbol{\theta}^*$. Assim, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \approx \hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \approx (\mathbf{F}^t \mathbf{F})^{-1} \mathbf{F}^t \boldsymbol{\varepsilon}$. Além disso, de (10) com $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, tem-se:

$$\begin{aligned} \mathbf{f}(\hat{\boldsymbol{\theta}}) - \mathbf{f}(\boldsymbol{\theta}^*) &\approx \mathbf{F}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\ &\approx \mathbf{F}(\mathbf{F}^t \mathbf{F})^{-1} \mathbf{F}^t \boldsymbol{\varepsilon} \\ &= \mathbf{P}_F \boldsymbol{\varepsilon} \end{aligned} \quad (12)$$

e

$$\begin{aligned} \mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}}) &\approx \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^*) - \mathbf{F}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\ &\approx \boldsymbol{\varepsilon} - \mathbf{P}_F \boldsymbol{\varepsilon} \\ &= (\mathbf{I}_n - \mathbf{P}_F) \boldsymbol{\varepsilon}, \end{aligned} \quad (13)$$

em que $\mathbf{P}_F = \mathbf{F}(\mathbf{F}^t \mathbf{F})^{-1} \mathbf{F}^t$ e $\mathbf{I}_n - \mathbf{P}_F$ é uma matriz simétrica e idempotente (Graybill, 1983). Conseqüentemente, usando (12) e (13) tem-se:

$$\begin{aligned} (n-p)s^2 &= S(\hat{\boldsymbol{\theta}}) \\ &= \|\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}})\|^2 \\ &\approx \|(\mathbf{I}_n - \mathbf{P}_F) \boldsymbol{\varepsilon}\|^2 \\ &= \boldsymbol{\varepsilon}^t (\mathbf{I}_n - \mathbf{P}_F) \boldsymbol{\varepsilon}, \end{aligned} \quad (14)$$

e,

$$\begin{aligned} \|\mathbf{f}(\hat{\boldsymbol{\theta}}) - \mathbf{f}(\boldsymbol{\theta}^*)\|^2 &\approx \|\mathbf{F}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|^2 \\ &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^t \mathbf{F}^t \mathbf{F} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\ &\approx \|\mathbf{P}_F \boldsymbol{\varepsilon}\|^2 \\ &= \boldsymbol{\varepsilon}^t \mathbf{P}_F \boldsymbol{\varepsilon}. \end{aligned} \quad (15)$$

Portanto, usando (14) e (15) obtém-se:

$$\begin{aligned} S(\boldsymbol{\theta}^*) - S(\hat{\boldsymbol{\theta}}) &\approx \boldsymbol{\theta}^t \boldsymbol{\theta} - \boldsymbol{\theta}^t (\mathbf{I}_n - \mathbf{P}_F) \boldsymbol{\theta} \\ &= \boldsymbol{\varepsilon}^t \mathbf{P}_F \boldsymbol{\varepsilon} \\ &\approx (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^t \mathbf{F}^t \mathbf{F} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*). \end{aligned}$$

Teorema 1. Dado que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, sob certas condições de regularidade (Seber e Wild, 1989) e n suficientemente grande, tem-se aproximadamente os seguintes resultados:

- (i) $\hat{\boldsymbol{\theta}}$ e $s^2 = S(\hat{\boldsymbol{\theta}})/(n-p)$ são estimadores consistentes de $\boldsymbol{\theta}^*$ e σ^2 respectivamente;
- (ii) $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \sim N_p(\mathbf{0}, \sigma^2 \mathbf{C}^{-1})$, em que $\mathbf{C} = \mathbf{F}^t \mathbf{F} = \mathbf{F}^t (\boldsymbol{\theta}^*) \mathbf{F} (\boldsymbol{\theta}^*)$;
- (iii) $(n-p) s^2 / \sigma^2 \approx \boldsymbol{\varepsilon}^t (\mathbf{I}_n - \mathbf{P}_F) \boldsymbol{\varepsilon} / \sigma^2 \sim \chi_{n-p}^2$;
- (iv) $\hat{\boldsymbol{\theta}}$ e s^2 são estatisticamente independentes.

O método de Gauss-Newton

Suponha que $\boldsymbol{\theta}^{(a)}$ é uma aproximação da estimativa de mínimos quadrados, $\hat{\boldsymbol{\theta}}$ do modelo (5). Para $\boldsymbol{\theta}$ próximo de $\boldsymbol{\theta}^{(a)}$, e considerando novamente uma expansão em série de Taylor, de primeira ordem, da função esperança como em (8), tem-se:

$$f(x_i; \boldsymbol{\theta}) \approx f(x_i; \boldsymbol{\theta}^{(a)}) + \sum_{j=1}^p \left[\frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(a)}} (\theta_j - \theta_j^{(a)}), \quad (16)$$

ou, em notação matricial $\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^{(a)}) + \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$.

Definindo $\mathbf{r}(\boldsymbol{\theta})$ como sendo um vetor de resíduos pode-se escrever:

$$\begin{aligned} \mathbf{r}(\boldsymbol{\theta}) &= \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}) \\ &\approx \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^{(a)}) - \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) \\ &= \mathbf{r}(\boldsymbol{\theta}^{(a)}) - \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}). \end{aligned}$$

Substituindo $\mathbf{r}^t(\boldsymbol{\theta})\mathbf{r}(\boldsymbol{\theta})$ em $S(\boldsymbol{\theta})$ obtém-se:

$$S(\boldsymbol{\theta}) \approx \mathbf{r}^t(\boldsymbol{\theta}^{(a)})\mathbf{r}(\boldsymbol{\theta}^{(a)}) - 2\mathbf{r}^t(\boldsymbol{\theta}^{(a)})\mathbf{F}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})^t \mathbf{F}^{t(a)} \mathbf{F}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$$

Portanto, $S(\boldsymbol{\theta})$ será minimizada quando:

$$\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)} = \left[\mathbf{F}^{t(a)} \mathbf{F}^{(a)} \right]^{-1} \mathbf{F}^{t(a)} \mathbf{r}(\boldsymbol{\theta}^{(a)}) \quad (17)$$

Assim, devido à aproximação $\boldsymbol{\theta}^{(a)}$, a próxima aproximação é dada por:

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + \left[\mathbf{F}^{t(a)} \mathbf{F}^{(a)} \right]^{-1} \mathbf{F}^{t(a)} \mathbf{r}(\boldsymbol{\theta}^{(a)}), \quad (18)$$

resultando no processo iterativo conhecido como Método de Gauss-Newton ou método da linearização.

Iniciando (18) com algum valor $\boldsymbol{\theta}^{(a)}$, o processo continua até a convergência, que ocorre quando $|\boldsymbol{\theta}^{(a+1)} - \boldsymbol{\theta}^{(a)}| < \delta$, em que δ é algum valor fixo, por exemplo, $\delta = 1e-10$.

Apesar do método de Gauss-Newton ser numericamente estável, este pode apresentar problemas, como, por exemplo:

- A convergência pode ser lenta se uma grande precisão for exigida;
- a matriz $\mathbf{F}^{(a)}$ pode ser singular ou tornar-se singular durante o processo iterativo;
- a convergência pode ser para um mínimo local e não para o mínimo global.

Uma característica interessante do método de Gauss-Newton ocorre quando a função esperança é linear. Nesse caso, o processo converge em uma única iteração, partindo-se de qualquer valor inicial (Ratkowsky, 1983).

Além do método Gauss-Newton, Marquardt e D.U.D., as estimativas de mínimos quadrados podem ser obtidas ainda por meio de outros métodos de otimização, como, por exemplo, o Método Simplex Nelder-Mead, Método de Otimização Quadrática, Método do Gradiente Conjugado, Método Newton-Raphson, Método Newton-Raphson Ridge. Entretanto, os métodos Gauss-Newton, Marquardt e D.U.D. são específicos para problemas de mínimos quadrados não-lineares (Hartmann, 1994).

Medidas de não-linearidade

Expressões usadas para avaliar a adequabilidade da aproximação linear e seus efeitos nas inferências são conhecidas na literatura como medidas de não-linearidade. Uma das primeiras tentativas relevantes no sentido de quantificar o comportamento não-linear de um modelo de regressão não-linear foi apresentada por Beale (1960), que propôs quatro medidas. De acordo com Guttman e Meeter (1965), essas medidas não devem ser usadas na prática, uma vez que elas tendem a subestimar a verdadeira não-linearidade (Bates e Watts, 1980).

Mais recentemente, Box (1971) propôs uma fórmula para estimar os vícios dos estimadores de mínimos quadrados de um modelo de regressão univariado. Gillis e Ratkowsky (1978), via simulação, concluíram que a medida de vício de Box não só estima o vício de maneira correta, mas também fornece uma boa indicação da extensão do comportamento não-linear do modelo. Bates e Watts (1980) apresentaram novas medidas de não-linearidade baseadas no conceito geométrico de curvatura. Eles provaram que a não-linearidade de

um modelo pode ser decomposta em duas componentes: a não linearidade intrínseca, associada à curvatura do espaço de estimação (Beale, 1960) e a não-linearidade devida ao efeito de parâmetros.

Na prática, a medida de vício de Box e as medidas de curvatura de Bates e Watts são as ferramentas mais utilizadas na avaliação da não-linearidade de um modelo de regressão não linear, podendo, também, ser usadas em problemas de discriminação, uma vez que o melhor modelo possível, dentre todos os propostos, pode ser considerado como aquele que apresenta o comportamento mais próximo do comportamento linear (Ratkowsky, 1983).

Dentre as muitas vantagens em se buscar modelos com comportamento próximo do comportamento linear, podem-se citar algumas:

- as estimativas de mínimos quadrados podem ser facilmente obtidas;
- os estimadores são aproximadamente não-viciados, normalmente distribuídos com variância mínima, mesmo em pequenas amostras;
- os valores de previsões são mais precisos;
- os métodos iterativos convergem mais rapidamente;
- os estimadores têm propriedades similares às propriedades ótimas de modelos lineares.

Segundo Ratkowsky (1990), o princípio conhecido como *Ochham's Razor* fornece uma base lógica para os modelos não-lineares uma vez que, em geral, modelos mais complexos provavelmente apresentarão maior extensão não-linear que modelos mais simples.

Medida de vício de Box

Box (1971) propôs uma estatística para avaliar o vício dos estimadores de mínimos quadrados dos parâmetros de um modelo de regressão não-linear univariado, dada por:

$$\text{Vício}(\hat{\theta}) = -\frac{\sigma^2}{2} \left[\sum_{i=1}^n \mathbf{F}(\theta) \mathbf{F}^t(\theta) \right]^{-1} \sum_{i=1}^n \mathbf{F}(\theta) \text{traço} \left[\left(\sum_{i=1}^n \mathbf{F}(\theta) \mathbf{F}^t(\theta) \right)^{-1} \mathbf{H}(\theta) \right], \quad (19)$$

em que $\mathbf{F}(\theta)$ é o vetor ($p \times 1$) de primeiras derivadas de $f(x_i; \theta)$, também chamado de vetor velocidade e $\mathbf{H}(\theta)$ é uma matriz ($p \times p$) de segundas derivadas com relação a cada elemento de θ .

Na prática, para o cálculo de (19), usa-se $\hat{\theta}$ e $\hat{\sigma}^2$ como sendo os verdadeiros valores de θ e σ^2 , respectivamente. Desta forma, o vetor ($p \times 1$), determinado em (21), representa a discrepância entre as estimativas dos parâmetros e seus verdadeiros valores. É comum expressar o valor da estimativa do vício em porcentagem, ou seja,

$$\% \text{Vício}(\hat{\theta}) = \frac{100 \text{Vício}(\hat{\theta})}{\hat{\theta}}. \quad (20)$$

Considera-se vícios acima de 1%, em valor absoluto, como um indicador do comportamento não-linear do modelo. A importância em se avaliar os vícios reside no fato deles indicarem qual ou quais parâmetros do modelo são os maiores responsáveis pelo comportamento distante do comportamento linear. Uma vez conhecidos esses parâmetros, pode-se buscar por uma reparametrização que possa reduzir a não-linearidade. O vício da nova parametrização pode ser avaliado por meio de outra estatística (Box, 1971).

Seja $\phi = g(\theta)$, isto é, o novo parâmetro ϕ é uma função de um ou mais elementos de θ_j , ($j = 1, \dots, p$). Box (1971) mostrou que:

$$\text{Vício}(\hat{\phi}) = \mathbf{G}^t(\hat{\theta}) \text{Vício}(\hat{\theta}) + \frac{1}{2} \text{traço} [\mathbf{M}(\hat{\theta}) \text{Cov}(\hat{\theta})], \quad (21)$$

em que $\mathbf{G}(\theta)$ é o vetor ($p \times 1$) de primeiras derivadas e $\mathbf{M}(\theta)$ é a matriz ($p \times p$) de segundas derivadas de $\mathbf{G}(\theta)$, com relação a cada θ_j , ($\mathbf{G}(\theta)$

e $\mathbf{M}(\theta)$ calculadas em $\hat{\theta}$), e $\text{Cov}(\hat{\theta})$ é a matriz de variâncias-covariâncias assintótica de $\hat{\theta}$ dada por $\hat{\sigma}^2 [\mathbf{F}^t(\hat{\theta}) \mathbf{F}(\hat{\theta})]^{-1}$. De maneira similar, pode-se obter a matriz de variâncias-covariâncias de $\hat{\phi}$ a partir da matriz de variâncias-covariâncias de $\hat{\theta}$, ficando esta representada pela seguinte expressão:

$$\text{Var}(\hat{\phi}) = \text{traço} [\mathbf{G}^t(\hat{\theta}) \mathbf{G}(\hat{\theta}) \text{Cov}(\hat{\theta})] \quad (22)$$

Além da medida de vício proposta por Box (1971), também são citados na literatura os trabalhos de Cox e Snell (1968), Clarke (1980), Amari (1982) e Hougaard (1985). Cook e Goldberg (1986), porém, mostraram que todas as medidas propostas para o cálculo de vício produzem basicamente resultados idênticos.

Medidas de curvatura de Bates e Watts

Usando conceitos de geometria diferencial, Bates e Watts (1980, 1981) estenderam as idéias de Beale (1960) e desenvolveram medidas de não-linearidade baseadas no conceito geométrico de curvatura. Essas medidas independem de mudanças na escala dos parâmetros ou dos dados, podendo ser usadas para comparar diferentes conjuntos de observações como também diferentes parametrizações (Seber e Wild, 1989).

Bates e Watts provaram que a não-linearidade de um modelo pode ser decomposta em duas componentes: a não-linearidade intrínseca (IN) e a não-linearidade devida ao efeito de parâmetros (PE). A não-linearidade intrínseca (IN) mede a curvatura do espaço de estimação no espaço amostral, em que o termo espaço de estimação se refere a todas as possíveis soluções do problema de mínimos quadrados. A solução de mínimos quadrados é o ponto no espaço de estimação que se encontra mais próximo do vetor de variáveis resposta. Em um modelo de regressão linear, a medida (IN) é nula, uma vez que o espaço de estimação é uma reta, um plano ou um hiperplano. Em um modelo não-linear o espaço de estimação é curvilíneo, e (IN) mede a extensão dessa curvatura. Bates e Watts (1980) e Ratkowsky (1983) concluíram que, na maioria dos modelos não-lineares de interesse prático, a medida (IN) geralmente é pequena. Isso significa que, se um modelo não-linear apresentar um comportamento distante do comportamento linear, essa não-linearidade é devida, principalmente, ao efeito de parâmetros. Então, uma reparametrização torna-se importante. Como a forma do espaço de estimação independe da parametrização, o processo de reparametrização não altera o valor de (IN). Segundo Ratkowsky (1983), modelos com curvatura intrínseca significativa possuem pouco interesse prático, sendo viável buscar-se modelos alternativos. Avalia-se a significância estatística de (IN) e (PE) comparando seus valores com $1/2\sqrt{F}$, em que $F = F_{p,n-p}^{\alpha}$ é o valor crítico obtido a partir de uma distribuição F com p e $(n-p)$ graus de liberdade.

A não-linearidade devida ao efeito de parâmetros é consequência da falta de uniformidade do sistema de coordenadas no espaço de estimação. No caso linear, as linhas paramétricas são retas paralelas. A medida (PE) é uma quantidade escalar que representa o máximo valor do efeito da parametrização, obtida a partir de um vetor tridimensional chamado vetor aceleração. Em um modelo linear, a matriz aceleração é formada de zeros, resultando, assim, em (PE) igual a zero.

Em um modelo não-linear com um dado valor de (IN), o valor de (PE) aumenta à medida que o seu comportamento se afasta do comportamento linear, uma vez que (PE) mede a extensão do comportamento não-linear causado pela parametrização.

Além das medidas de Bates e Watts, existem medidas de curvatura para subconjuntos de parâmetros, desenvolvidas por Cook e Goldberg (1986). Essas medidas são obtidas para todos os parâmetros do modelo e, desta forma, indicam qual ou quais parâmetros necessitam de uma reparametrização. Experiências práticas têm revelado que a obtenção dessas medidas requer um considerável esforço computacional (Ratkowsky, 1990), podendo ser substituídas pela medida de vício de Box, uma vez que esta também revela quais parâmetros são os maiores responsáveis pelo comportamento distante do comportamento linear. Uma outra maneira de se avaliar a extensão do comportamento não-linear é por meio dos chamados *profile t-plots*, *profile traces plots* e *profile pair sketches plots* (Bates e Watts, 1988). Segundo Ratkowsky (1990), os *profile t-plots* revelam graficamente a extensão da não-linearidade individualmente para cada parâmetro, e os *profile traces plots* fornecem informações de como os parâmetros interagem. O *Profile pair sketches plots* por sua vez, é um método computacionalmente simples para obtenção de regiões de verossimilhança (Bates e Watts, 1988). Outras medidas de não-linearidade também são discutidas na literatura, por exemplo, Linszen (1975) sugeriu modificações nas medidas propostas por Beale (1960) e Williams (1962) propôs uma medida de não-linearidade fundamentada na razão pela qual os algoritmos baseados em aproximações lineares convergem.

Estudo de simulação

As medidas de curvatura de Bates e Watts (1980) são as medidas usadas para avaliar a não-linearidade intrínseca e a não-linearidade devida ao efeito de parâmetros em modelos de regressão não-lineares. Se a não-linearidade intrínseca (IN) for pequena, não significativa, pode ser que a não-linearidade, devido ao efeito de parâmetros (PE) seja alta, significativa, então, uma reparametrização muitas vezes pode reverter esse fato.

A medida de efeito de parâmetros (PE), apesar de ser uma medida de não-linearidade, não oferece nenhuma estratégia para possíveis reparametrizações, uma vez que, em situações multiparamétricas, esta não indica qual ou quais parâmetros são os maiores responsáveis pelo comportamento não-linear do

modelo. Em tais situações, a medida de vício de Box (1971) é bastante importante. No entanto, se o tamanho da amostra não for suficientemente grande, o vício de θ_j pode não ser uma boa medida e, possivelmente, $\hat{\sigma}^2[\mathbf{F}^t(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}$ subestima a verdadeira variância.

Como alternativa, Ratkowsky (1983) propõe um método baseado em simulações de n observações experimentais M vezes. Esse procedimento pode ser extremamente útil na verificação das propriedades assintóticas dos estimadores de mínimos quadrados, sem a necessidade do cálculo das medidas de curvatura de Bates e Watts ou do vício de Box.

Para investigar as propriedades dos estimadores de um particular modelo deve-se, inicialmente, obter as estimativas dos parâmetros e a estimativa da variância residual. Considerando essas estimativas como sendo os verdadeiros valores dos parâmetros, 500 ou preferivelmente 1000 conjuntos de observações de tamanho n são gerados a partir do modelo

$$y_i = f(\mathbf{x}_i; \hat{\theta}) + \delta_i, \quad (23)$$

em que, δ_i são variáveis aleatórias independentes e identicamente distribuídas $N(\mathbf{0}, \hat{\sigma}^2)$.

Para cada amostra simulada, deve-se obter as estimativas dos parâmetros $\tilde{\theta}_m$ ($m = 1, \dots, M$) para o modelo $y_i = f(\mathbf{x}_i; \theta) + \varepsilon_i$.

Desta forma, a distribuição dos estimadores pode ser examinada usando-se testes formais (Seber e Wild, 1989 ou Ratkowsky, 1983), ou por meio de histogramas de frequência para cada parâmetro. Histogramas revelam facilmente o comportamento não normal dos estimadores, uma vez que um histograma assimétrico é uma forte evidência da falta de normalidade.

Alguns testes usados para examinar as distribuições marginais dos parâmetros - uma vez que se analisar a distribuição conjunta - pode apresentar alguns problemas mesmo para casos mais simples, os quais são apresentados em Ratkowsky (1983).

Medidas de assimetria

Utilizando as idéias propostas por Lowry e Morton (1983), Morton (1987), Ratkowsky (1983) discute outras medidas de vício e de não-linearidade, as quais também são baseadas em simulações. Segundo ele, essas novas medidas têm se mostrado ferramentas bastante úteis na avaliação do vício dos estimadores de mínimos quadrados de um modelo

de regressão não linear univariado. Essas medidas são apresentadas a seguir.

Considera-se um modelo de regressão usualmente escrito na forma:

$$y_i = f(\mathbf{x}_i; \theta) + \varepsilon_i, \quad (24)$$

em que $f(\mathbf{x}_i; \theta)$ pode ser linear ou não-linear nos parâmetros. Assumindo que θ é conhecido, pode-se simular uma série de ε_i independentes e identicamente distribuídos, normais, com média zero e variância constante σ^2 . Uma vez que os erros são simétricos e em torno de zero, existem, na verdade, dois modelos que podem ser simulados, ou seja,

$$y_i^+ = f(\mathbf{x}_i; \hat{\theta}) + \varepsilon_i \quad \text{e} \quad y_i^- = f(\mathbf{x}_i; \hat{\theta}) - \varepsilon_i,$$

em que y_i^+ e y_i^- são as i -ésimas respostas obtidas somando-se e subtraindo-se ε_i , respectivamente, cujas estimativas de mínimos quadrados serão denotadas por $\hat{\theta}^+$ e $\hat{\theta}^-$, respectivamente.

No caso linear, segundo Ratkowsky (1983), para cada parâmetro θ_j tem-se que:

$$(\hat{\theta}_j^+ - \theta_j) = (\hat{\theta}_j^- - \theta_j). \quad (25)$$

Geometricamente, as estimativas $\hat{\theta}_j^+$ e $\hat{\theta}_j^-$ são simétricas em torno do verdadeiro valor do parâmetro. No caso não-linear esse fato, em geral, não é verdade e, desta forma, a estatística:

$$\Psi_j(\hat{\theta}_j^+, \hat{\theta}_j^-) = \frac{(\hat{\theta}_j^+ - \theta_j) + (\hat{\theta}_j^- - \theta_j)}{2}, \quad (26)$$

pode ser usada como uma medida do comportamento não-linear dos estimadores dos parâmetros. Uma vez que a distribuição dos $\hat{\theta}_j^+$ e $\hat{\theta}_j^-$ são idênticas, segue que:

$$E(\psi_j) = E(\hat{\theta}_j^+ - \theta_j) = E(\hat{\theta}_j^- - \theta_j) = \text{Vício}(\hat{\theta}_j). \quad (27)$$

Substituindo (26) em (27), tem-se que em um modelo linear a esperança e a variância são iguais a zero. Para o caso não-linear, a variância de Ψ_j é dada por:

$$\text{Var}(\psi_j) = \frac{1}{4} \text{var}(\hat{\theta}_j^+) + \frac{1}{4} \text{var}(\hat{\theta}_j^-) + \frac{1}{2} \text{cov}(\hat{\theta}_j^+, \hat{\theta}_j^-), \quad (28)$$

e o coeficiente de correlação entre $\hat{\theta}_j^+$ e $\hat{\theta}_j^-$ é definido por:

$$\text{Corr}(\hat{\theta}_j^+, \hat{\theta}_j^-) = \frac{\text{cov}(\hat{\theta}_j^+, \hat{\theta}_j^-)}{\sqrt{\text{var}(\hat{\theta}_j^+)}\sqrt{\text{var}(\hat{\theta}_j^-)}}. \quad (29)$$

No caso de modelos lineares, $\text{Corr}(\hat{\theta}_j^+, \hat{\theta}_j^-)$ é igual a -1 e, no caso não-linear, $-1 \leq \text{Corr}(\hat{\theta}_j^+, \hat{\theta}_j^-) \leq 1$, sendo que, para $\text{Corr}(\hat{\theta}_j^+, \hat{\theta}_j^-)$ próximo de -1 , tem-se modelos com comportamento próximo do comportamento linear.

Discussão

Diferente da classe dos modelos de regressão lineares, em que a validade das inferências são avaliadas principalmente por meio de diagnósticos de regressão, no caso não-linear, além de diagnósticos usuais, deve-se, também, avaliar a extensão do comportamento não-linear do modelo adotado. Diagnósticos usuais de regressão são ferramentas bastante conhecidas na literatura e utilizadas por estatísticos ou usuários que utilizam da estatística para melhor entender seus problemas. Por outro lado, medidas de não linearidade, em princípio, não são tão conhecidas, e requerem algum esforço extra nas suas obtenções. Neste artigo, considerando a importância de se avaliar a extensão do comportamento não-linear de modelos de regressão não-lineares, procurou-se apresentar, sem a pretensão de ser exaustivo, as principais ferramentas discutidas na literatura de regressão não-linear. Dentre algumas medidas apresentadas, citam-se as de curvatura de Bates e Watts, a medida de vício proposta por Box (Box, 1971), métodos de simulação e medidas de assimetria. Apesar da ausência de alguma aplicação prática, os conceitos apresentados neste artigo podem ser de grande interesse, principalmente para pesquisadores e usuários de métodos estatísticos. Mazucheli (1995), Mazucheli e Achcar (1997), sob o ponto de vista clássico e bayesiano, apresentam uma série de aplicações para alguns modelos não-lineares de crescimento sigmóide.

Referências

- AMARI, S.I. Differential geometry of curved exponential families-curvatures and information loss. *Ann. Stat.*, v.10, p.375-385, 1982.
- BATES, D.M.; WATTS, D.G. Relative curvature measures of nonlinearity. *J. R. Stat. Soc., B*, London, v.42, p.1-25, 1980.
- BATES, D.M.; WATTS, D.G. Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Stat.*, Hayward, v.9, p.1152-1167, 1981.
- BATES, D.M.; WATTS, D.G. *Nonlinear regression analysis and its applications*. New York: John Wiley and Sons, 1988.

- BEALE, E.M.L. Confidence regions in nonlinear estimation. *J. R. Statist. Soc., B*, London, v.22, p.41-76, 1960.
- BELSLEY, D.A. *et al. Regression diagnostic - identifying influential data and sources of variation*. New York: John Wiley and Sons, 1980.
- BOX, M.J. Bias in nonlinear estimation. *J. R. Stat. Soc., B*, v.33, p.171-201, 1971.
- CLARKE, G.P.Y. Moments of the least squares estimators in a nonlinear regression model. *J. R. Stat. Soc. Ser. B*, London, v.42, p.227-237, 1980.
- COOK, R.D.; GOLDBERG, M.L. Curvatures for parameter subsets in nonlinear regression model. *Ann. Stat.* v.14, p.1399-1418, 1986.
- COOK, R.D.; WEISBERG, S. *Residuals and influence in regression*. London: Chapman and Hall, 1982.
- COX, D.R.; SNELL, E.J. A general definition of residuals. *J. R. Stat. Soc., B*, v.30, p.248-275, 1968.
- DRAPER, N.R.; SMITH, H. *Applied regression analysis*. 2. ed. New York: John Wiley and Sons, 1981.
- GILLIS, P.R.; RATKOWSKY, D.A. The behavior of estimators of the parameters of various yield-density relationships. *Biometrics*, Washington, DC., v.34, p.191-198, 1978.
- GRAYBILL, F.A. *Matrices with applications in statistics*. 2. ed. Ed. Belmont, CA: Wadsworth Publishing Co., 1983.
- GUTTMAN, I.; MEETER, D.A. On the use of measures of non-linearity. *Technometrics*, Washington, DC., v.7, p.623-637, 1965.
- HARTMANN, W.M. Internal draft document: nonlinear optimization in I.M.L. SAS Institute, 1994.
- HOUGAARD, P. The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. R. Stat. Soc., B*, London, v.47, p.103-114, 1985.
- JENNRICH, R.I. Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Stat.*, Beochwood, v.40, p.633-643, 1969.
- KHURI, A.I.; CORNELL, J.A. *Response surface: designs and analyses*. New York: Marcel Dekker, Inc., 1987.
- LAWTON, W.H.; SYLVESTRE, E.A. Estimation of linear parameters in nonlinear regression. *Technometrics*, v.13, p.461-467, 1971.
- LINSSEN, H.N. Nonlinearity measures: a case study. *Stat. Neerland.*, Barendrecht, v.29, p.93-99, 1975.
- LOWRY, R.K.; MORTON, R. An asymmetry measure for estimators in non-linear regression models. *Proc. 44th Session Inst., Madrid, Contributed Papers*, v.1, p.351-354, 1983.
- MAZUCHELI, J. *Análise bayesiana e discriminação de modelos não lineares*. 1995. Dissertação (Mestrado) - Universidade de São Paulo, São Carlos, 1995.
- MAZUCHELI, J.; ACHCAR, J.A. Análise bayesiana para modelos de crescimento. *Revista Brasileira de Estatística*, Rio de Janeiro, v.58, p.77-94, 1997.
- MORTON, R. Asymmetry of estimators in nonlinear regression. *Biometrika*, London, v.74, p.678-685, 1987.

- RALSTON, M.L.; JENNRICH, R.I. Duda, a derivative-free algorithm for nonlinear least squares. *Technometrics*, v.20, p.7-14, 1978.
- RATKOWSKY, D.A. *Nonlinear regression models*. New York: Marcel Dekker, Inc., 1983.
- RATKOWSKY, D.A. *Handbook of nonlinear regression models*. New York: Marcel Dekker, Inc., 1990.
- SEARLE, S.R. *Linear models*. New York: John Wiley and Sons, 1971.
- SEBER, G.A.F. *Linear regression*. New York: John Wiley and Sons, 1977.
- SEBER, G.A.F.; WILD, C. J. *Nonlinear regression*. New York: John Wiley and Sons, 1989.
- WILLIAMS, E.J. Exact fiducial limits in non-linear estimation. *J. R. Stat. Soc. B*, London, v.24, p.125-139, 1962.

Received on September 12, 2002.

Accepted on November 06, 2002.