

Método de Newton e Gauss-Newton na estimação dos parâmetros de modelo de regressão não linear

Edilson Marcelino Silva ¹, Ariana Campos Frühauf ², Felipe Augusto Fernandes ³, Gustavo Sérgio de Paula ⁴, Joel Augusto Muniz ⁵, Tales Jesus Fernandes ⁶

1 Introdução

Em estatística, os modelos de regressão consistem em ajustar uma função a um conjunto de n observações. A regressão estuda a relação entre duas ou mais variáveis e incorpora os erros experimentais sendo expressa por:

$$\mathbf{y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

em que, \mathbf{y} é o vetor com os valores observados (variável dependente); \mathbf{X} matriz com as covariáveis (variável independente); f é a função que associa as variáveis; $\boldsymbol{\beta}$ é o vetor de parâmetros e $\boldsymbol{\varepsilon}$ é o vetor de erros experimentais.

Os modelos de regressão são classificados basicamente em modelos lineares e não lineares em relação aos seus parâmetros. Considera-se o modelo de regressão como linear quando as derivadas parciais em relação a cada parâmetro do modelo não dependem de nenhum parâmetro do modelo. Por outro lado, se pelo menos uma derivada parcial depender de algum parâmetro este é classificado como não linear. Não há forma fechada para a solução do sistema de equações normais de um modelo não linear, sendo necessário utilizar métodos numéricos iterativos (DRAPER e SMITH, 2014).

Para a solução aproximada de sistema de equações foi criado por Newton (1643-1727) o denominado “Método de Newton”, um processo que é aplicado iterativamente para obter aproximação da solução, tão precisa quanto se queira. O método de Newton aproxima $f(\boldsymbol{\beta})$ em $\boldsymbol{\beta}^n$ por sua aproximação em série de Talyor até o termo quadrático, assim:

$$f(\boldsymbol{\beta}) \approx f(\boldsymbol{\beta}^n) + \mathbf{J}(\boldsymbol{\beta}^n)'(\boldsymbol{\beta} - \boldsymbol{\beta}^n) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^n)'\mathbf{H}(\boldsymbol{\beta}^n)(\boldsymbol{\beta} - \boldsymbol{\beta}^n)$$

em que \mathbf{J} é a matriz de derivadas parciais de $f(\boldsymbol{\beta})$ em relação a $\boldsymbol{\beta}$ denominada matriz Jacobiana e \mathbf{H} é matriz de segundas derivadas denominada de matriz Hessiana.

Derivando a função quadrática em relação a $\boldsymbol{\beta}$ igualando a zero é obtida a condição para ter o mínimo:

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *edilsonmg3@hotmail.com*

²Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *arianafruhauf@gmail.com*

³Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *fernandesfelipepest@gmail.com*

⁴Iniciação científica, Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *gustavo028@hotmail.com*

⁵Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *joamuniz@ufla.br*

⁶Departamento de Estatística (DES), Universidade Federal de Lavras (UFLA). e-mail: *tales.jfernandes@ufla.br*

$$\mathbf{J}^n + \mathbf{H}^n(\boldsymbol{\beta} - \boldsymbol{\beta}^n) = 0$$

ou

$$\boldsymbol{\beta} = \boldsymbol{\beta}^n - (\mathbf{H}^n)^{-1} \mathbf{J}^n.$$

Assim a solução da equação $f(\boldsymbol{\beta})$ é aproximada pelo método iterativo de Newton como apresentado abaixo:

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - (\mathbf{H}^n)^{-1} \mathbf{J}^n$$

(BARD, 1974).

Entre as descobertas mais importantes de Gauss (1777-1855) pode-se destacar o método dos mínimos quadrados. O diretor de um observatório na Itália tinha algumas observações de um asteroide, nomeado Ceres, o qual não foi visto depois de algumas semanas. O problema despertou o interesse de Gauss, pois a partir de um determinado número de observações queria descobrir a órbita do planeta. Assim, descobriria para onde os observadores deveriam apontar o telescópio. Para determinar a órbita de planetas, Gauss criou o método de mínimos quadrados que levou a resultados surpreendentes. O asteroide Ceres foi visualizado depois em posição muito próxima a indicada pelos cálculos realizados por Gauss (BOYER, 2010; EVES, 2004).

O objetivo do trabalho foi comparar os métodos iterativos de Newton e Gauss-Newton no ajuste do modelo Stanford e Smith (1972) aos dados de mineralização de carbono de palha de aveia no solo.

2 Material e métodos

Os dados utilizados para o ajuste do modelo foram extraídos de Giacomini et al. (2008). O experimento foi realizado em laboratório. Um argissolo vermelho distrófico arênico, da camada de 0-10cm de uma área que vinha sendo manejada em sistema plantio direto foi avaliado. A coleta da aveia foi feita no estádio de maturação fisiológica, submetida à secagem ao ar e armazenada em lugar seco até o momento da incubação. Neste trabalho foi avaliado um tratamento com palha de aveia incorporada no solo. A mineralização do C foi avaliada por meio da emissão de CO₂, durante a incubação, medindo-se a porcentagem de carbono mineralizado sempre nas mesmas unidades experimentais aos 3, 5, 9, 14, 20, 25, 30, 35, 45, 55, 65 e 80 dias do início da incubação.

Avaliou-se o modelo Stanford e Smith:

$$C_i = C_0(1 - \exp(-kt_i)) + \varepsilon_i$$

No modelo, C_i corresponde a porcentagem do carbono adicionado mineralizado até o tempo t_i (em dias), $i = 1, 2, \dots, n$; C_0 é o carbono potencialmente mineralizável; k é a taxa de mineralização e ε_i é o erro experimental (SILVA, et al., 2019a; SILVA, et al., 2019b).

O método de mínimos quadrados protosto por Gauss consiste em minimizar a soma de quadrados do erro. Para os modelos de regressão o erro é obtido por: $\boldsymbol{\varepsilon} = \mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})$ e a soma de quadrados do erro é dada por:

$$\begin{aligned} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta}))'(\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})) \\ &= \mathbf{y}'\mathbf{y} - 2f'(\mathbf{X}, \boldsymbol{\beta})\mathbf{y} + f'(\mathbf{X}, \boldsymbol{\beta})f(\mathbf{X}, \boldsymbol{\beta}). \end{aligned}$$

Observa-se que o ponto que minimiza a soma de quadrados do erro $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ é obtido derivando-se a função em relação a $\boldsymbol{\beta}$ e igualando a zero, chega-se ao sistema de equações normais:

$$\mathbf{X}'f(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}'\mathbf{y}$$

em que $\mathbf{x} = \frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. Logo, \mathbf{X} é a matriz de derivadas parciais do modelo, e para modelo não linear a matriz \mathbf{X} tem parâmetro pela definição de que um modelo é classificado como não linear se pelo menos uma derivada parcial da função depender de parâmetro. Portanto, $f(\mathbf{X}, \boldsymbol{\beta})$ e \mathbf{X} dependem de $\boldsymbol{\beta}$. Deste modo, não é possível obter forma fechada para a solução $\hat{\boldsymbol{\beta}}$, sendo necessário o uso de método iterativo para aproximação da solução (DRAPER e SMITH, 2014).

Assim, precisa-se determinar \mathbf{J} e \mathbf{H} da função $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ para usar o método de Newton para ter uma solução aproximada dos $\boldsymbol{\beta}$ que minimizam a soma de quadrados do erro:

$$\mathbf{J} = \frac{\partial \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = -2 \frac{\partial f'(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathbf{y} + 2 \frac{\partial f'(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} f(\mathbf{X}, \boldsymbol{\beta}) = -2 \frac{\partial f'(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} (\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})) = -2\mathbf{X}'\boldsymbol{\varepsilon}$$

e

$$\mathbf{H} = \frac{\partial^2 \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \frac{\partial \mathbf{J}}{\partial \boldsymbol{\beta}} = -2 \left[\mathbf{X}'\mathbf{X} - \sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i \right]$$

Pelo método de Newton:

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - (\mathbf{H}^n)^{-1} \mathbf{J}^n$$

Substituindo-se \mathbf{J} e \mathbf{H} na expressão acima, tem-se:

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - \left(-\frac{1}{2} \right) \left[\mathbf{X}'\mathbf{X} - \sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i \right]^{-1} (-2\mathbf{X}'\boldsymbol{\varepsilon})$$

ou

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - \left[\mathbf{X}'\mathbf{X} - \sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i \right]^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$$

Para o seu problema de minimizar a soma de quadrados do erro $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, Gauss desconsiderou o segundo termo da matriz Hessiana, pelo fato do termo $\sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i$ ser muito pequeno em relação a $\mathbf{X}'\mathbf{X}$ a medida que se aproxima da solução do sistema. Com essa otimização, criou o denominado método de Gauss-Newton que ficou da seguinte forma:

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$$

(BARD, 1974; JUDGE, et al., 1985; NOCEDAL e WRIGHT, 2006).

O método de Gauss-Newton também é denominado “método da linearização”, pois desconsiderar o termo do erro na matriz Hessiana equivale a considerar a equação $f(\mathbf{X}, \boldsymbol{\beta})$ a série de Taylor até o termo de primeira ordem (aproximação linear) e aplicar o método de mínimos quadrados (DRAPER e SMITH, 2014; MAZUCHELI e ACHCAR, 2002).

Considerou-se que o modelo ajustado convergiu se a diferença da soma de quadrados do erro (SQE) das iterações $i - 1$ e i foi menor que 10^{-4} . Além disso, foi contado o número de iterações que o método precisou para convergir e calculou-se a SQE. Os valores iniciais para os parâmetros foram $\boldsymbol{\beta}^0 = (C_0^0 \ k^0)' = (70 \ 0,03)'$ As análises foram realizadas utilizando-se o software R (R Core Team, 2017).

3 Resultados e discussão

Apresenta-se abaixo, com base no modelo Stanford e Smith, a matriz \mathbf{X} usada na implementação dos algoritmos de Newton e Gauss-Newton e a matriz $\frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ usada na implementação do método iterativo de Newton. Uma das vantagens do método iterativo de Gauss-Newton é não precisar calcular a matriz $\frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ de segundas derivadas (NOCEDAL e WRIGHT, 2006).

$$\mathbf{X} = \begin{bmatrix} 1 - e^{-kt_1} & t_1 C_0 e^{-kt_1} \\ \vdots & \vdots \\ 1 - e^{-kt_n} & t_n C_0 e^{-kt_n} \end{bmatrix} \text{ e } \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \begin{bmatrix} 0 & t_i e^{-kt_i} \\ t_i e^{-kt_i} & -t_i^2 C_0 e^{-kt_i} \end{bmatrix}$$

Na Tabela 1 são apresentadas as estimativas dos parâmetros do modelo Stanford e Smith aos dados de mineralização do carbono de palha de aveia no solo considerando os métodos iterativos de Newton e Gauss-Newton. Observa-se que a diferença na estimativa do parâmetro C_0 foi na terceira casa decimal e do parâmetro k na sexta casa decimal, o que na prática é desprezível, como ilustrado na Figura 1. Logo, ambos os métodos iterativos foram eficientes na estimação dos parâmetros do modelo. Segundo Mazucheli e Achcar (2002) a escolha de bons valores iniciais são fundamentais para a convergência de qualquer método iterativo.

Tabela 1: Estimativas dos parâmetros do modelo Stanford e Smith considerando os métodos iterativos (Fonte: próprio autor).

Parâmetro	Método de Newton	Método Gauss-Newton
C_0	67,70202200	67,70131516
k	0,02414186	0,02414234

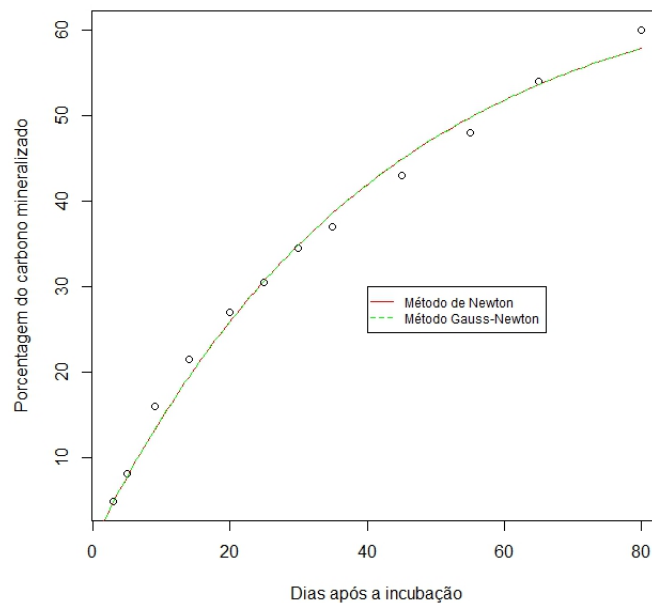


Figura 1: Ajuste do modelo Stanford e Smith considerando os métodos iterativos (Fonte: próprio autor).

Na Tabela 2 são apresentadas a análise da convergência considerando os métodos iterativos de Newton e Gauss-Newton, que precisou de apenas 6 e 4 iterações até atingir a convergência ($SQE_{i-1} - SQE_i < 10^{-4}$), respectivamente, e a SQE apresentou diferença depois da quinta casa decimal (Tabela 2). O método Gauss-Newton é específico para estimar parâmetros de modelos de regressão não linear (MAZUCHELI e ACHCAR, 2002) e pela eficiência está implementado nos softwares até hoje (R Core Team, 2017).

Tabela 2: Análise da convergência considerando os métodos iterativos (Fonte: próprio autor).

	Método de Newton	Método Gauss-Newton
Número de iterações	6	4
SQE	27,35236	27,35236
$SQE_{i-1} - SQE_i$	0,00006	0,00001

SQE - soma de quadrados dos erros.

Para ilustrar os passos dos algoritmos, a Figura 2 foi feita com 6 iterações de cada método numérico. Percebe-se que a partir da quarta iteração o tamanho do passo foi muito pequeno.

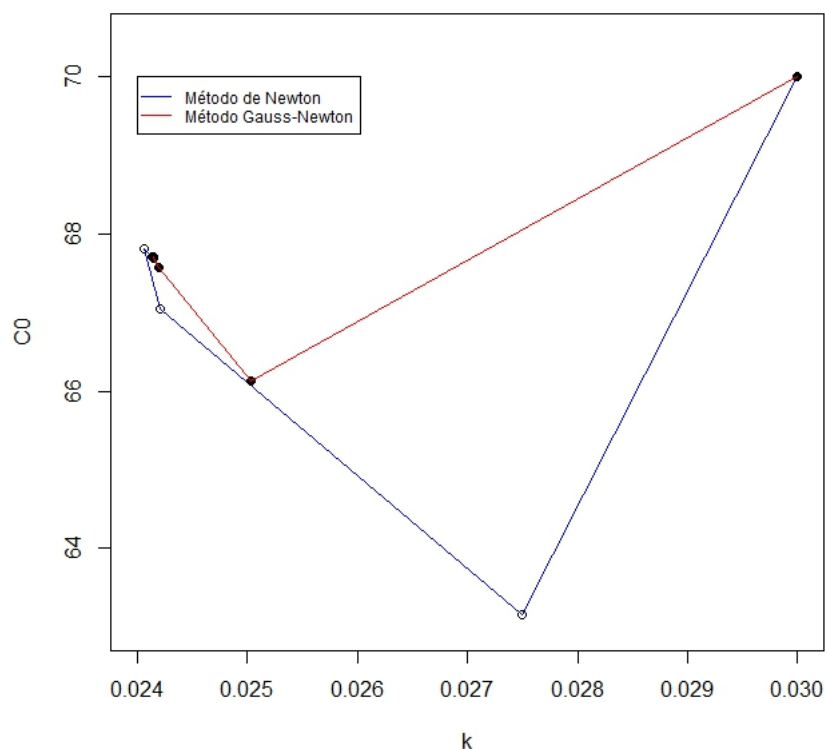


Figura 2: Passos dos métodos iterativos no ajuste do modelo Stanford e Smith aos dados de mineralização de palha de aveia (Fonte: próprio autor).

Apresenta-se abaixo, com base no modelo Stanford e Smith, a matriz $\mathbf{X}'\mathbf{X}$ e a matriz $\sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i$ na última iteração do método de Newton. Observa-se que os termos da matriz $\sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i$ são irrelevantes em relação aos termos da matrix $\mathbf{X}'\mathbf{X}$ (NOCEDAL e WRIGHT, 2006).

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3,4183 & 4967,5200 \\ 4967,5200 & 7869902,7300 \end{bmatrix} \text{ e } \sum_{i=1}^n \frac{\partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \boldsymbol{\varepsilon}_i = \begin{bmatrix} 0 & 0,0017 \\ 0,0017 & 47857,4700 \end{bmatrix}$$

4 Conclusão

O método dos mínimos quadrados proposto por Gauss tem grande aplicação na Estatística na estimação de parâmetros. O método Gauss-Newton, específico para estimar parâmetros de modelos não lineares, foi mais eficiente que o método de Newton na estimação dos parâmetros do modelo Stanford e Smith ajustado a dados de mineralização de carbono. O método Gauss-Newton é implementado nos softwares para estimação de parâmetros de modelos de regressão não linear.

Agradecimentos

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio nesta pesquisa.

Referencias Bibliográficas

- BARD, Y. *Nonlinear parameter estimation*. 1a ed. Academic Press, 1974.
- BOYER, C. B. *História da Matemática*, 3 edição. São Paulo: Blücher, 2010.
- DRAPER, N. R.; SMITH, H. *Applied regression analysis*. 3a ed, reprint. John Wiley, 2014.
- EVES, H. W. *Introdução a história da matemática*. São Paulo: UNICAMP, 2004.
- GIACOMINI, S. J.; AITA, C.; MIOLA, E. C. C.; RECOUS, S. Mineralização do carbono da palha de aveia e dejetos de suínos aplicados na superfície ou incorporados ao solo. *Revista Brasileira de Ciência do Solo*, v. 32, p. 2661-2668, 2008.
- JUDGE, G. G.; GRIFFITHS, W. E.; HILL, R. C.; LÜTKEPOHL, H.; LEE, T. *The theory and practice of econometrics*. 2a ed. John Wiley, 1985.
- MAZUCHELI, J.; ACHCAR, J. A. Algumas considerações em regressão não-linear. *Acta Scientiarum*, v. 24, p. 1761-1770, 2002.
- NOCEDAL, J. WRIGHT, S. *Numerical Optimization*. 2a ed. Springer, 2006.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2017. URL <http://www.R-project.org/>.

SILVA, E. M.; RIBEIRO, T. D.; FERNANDES, J. G.; MUNIZ, J. A. Descrição da mineralização do carbono de dejetos de suínos e palha de aveia no solo por modelos não lineares. *Revista Agrogeoambiental*, v. 11, p. (no prelo), 2019a.

SILVA, E. M.; SILVEIRA, S. C.; RIBEIRO, T. D.; MUNIZ, J. A. Ajuste da decomposição do lodo de esgoto e palha de aveia por modelos não lineares. *Revista Agrogeoambiental*, v. 11, p. (no prelo), 2019b.

STANFORD, G.; SMITH, S. J. Nitrogen mineralization potentials of soil. *Soil Science Society of America Journal*, v. 36, p. 465-471, 1972.