Revista Ceres, 50(287):9-26,2003

TESTE PARA VERIFICAR A IGUALDADE DE PARÂMETROS E A IDENTIDADE DE MODELOS DE REGRESSÃO NÃO-LINEAR[1]

Adair José Regazzi[2]

RESUMO

Neste trabalho, foi considerado o ajustamento de g equações de regressão não-lincar (g grupos), com o objctivo dc aprcscntar um mctodo para tcstar as scguintcs hipóteses: (a) H_0: as g equações são idênticas, isto é, uma equação comum pode ser usada como estimativa das g equações envolvidas; e (b) H_0: um determinado subconjunto de parâmetros é igual nos g grupos. Como ilustração, considerou-se o modelo de crescimento logístico. A identidade de modelos de regressão não-linear e a igualdade de qualquer subconjunto de parâmetros podem ser verificadas por meio do teste da razão de verossimilhança. A metodologia apresentada é geral e pode ser usada em qualquer modelo de regressão não-linear.

Palavras-chave: curvas de crescimento, verossimilhança, teste de hipótese.

ABSTRACT

TEST FOR PARAMETER EQUALITY IN NONLINEAR REGRESSION MODELS

This paper is concerned with the adjustment of g nonlinear regression equations. A framework is presented for testing the following null hypothesis: (a) H_0 : the g nonlinear equations are all identical, i.e., one single equation can be adjusted; and (b) H_0 : a subset of the parameters from the g equations is equal. The proposed framework is illustrated

---

[1] Aceito para publicação em 30.10.2002.

[2] Departamento de Informática, Universidade Federal de Viçosa 36571-000 Viçosa, MG. E-mail: adairreg@mail.ufv.br (Bolsista do CNPq).

with the logistic growth model. The likelihood ratio test is used and the methodology is very general in the sense that it can be applied to any nonlinear regression model.

Key words: growth curves, likelihood, hypothesis test.

INTRODUÇÃO

Na análise estatística de dados de uma variável Y obtidos de experimentos em que os tratamentos são níveis crescentes de um fator quantitativo X, como níveis de adubação, tempo, temperatura ou idade, o efeito de tratamento deve ser avaliado, em geral, por meio de uma análise de regressão, pois neste caso o uso de procedimentos para comparações múltiplas (testes de médias) não é indicado. A análise de regressão, com modelo linear ou não-linear, é uma técnica potencialmente útil na análise de dados, tendo grande aplicação nas mais diversas áreas do conhecimento.

Com muita freqüência, estuda-se a relação funcional entre uma variável dependente Y e uma ou mais variáveis independentes X. Nestes casos, estudos são realizados em diferentes tratamentos ou fatores, e para cada situação a análise de regressão é aplicada separadamente, obtendo-se tantas equações quanto o número de situações distintas. Um problema que tem aplicação importante é determinar se um conjunto de curvas é idêntico. Graybill (*5*) apresentou um método geral para testar a hipótese de igualdade de um conjunto de modelos lineares, empregando o teste F. Como exemplo, citou o uso de fertilizantes em determinada cultura, em que se usa certo número de variedades e, para cada uma, obtém-se a relação entre a produção e a quantidade de fertilizante aplicada, mediante equações de regressão. Steel e Torrie (*17*) apresentaram testes para verificar a igualdade entre dois coeficientes de regressão e, também, entre mais de dois coeficientes de regressão linear simples. Neter et al. (*9*) testaram se duas equações de regressão linear simples eram idênticas, utilizando o teste F. Mostraram também a utilização de variáveis indicadoras em modelos de regressão, com o intuito de comparar valores de parâmetros em modelos de regressão com variáveis independentes qualitativas.

Regazzi (*13*) considerou o ajustamento de H equações de regressão polinomial de grau k, mediante o emprego da técnica dos polinômios ortogonais, em que apresentou, em detalhes, um método para testar as seguintes hipóteses: (a) H_0: as H equações são idênticas; b) H_0: as H equações têm uma constante de regressão comum; e c) H_0: as H equações têm um ou mais coeficientes de regressão iguais. Ele concluiu que o método apresentado é geral e pode ser usado em modelos polinomiais de qualquer grau, ortogonais ou não, e também em modelos de regressão múltipla. Regazzi (*14*) apresentou, em detalhes, um método para testar as mesmas

hipóteses citadas anteriormente, considerando o caso de dados provenientes de delineamentos experimentais (com repetições).

A regressão linear é amplamente utilizada para a representação dos fenômenos biológicos como o crescimento de organismos vivos na sua fase inicial. No entanto, esses fenômenos, quando estudados durante um tempo maior de desenvolvimento do organismo, não são bem representados por uma função linear. O processo de desenvolvimento de organismos vivos é caracterizado por uma fase de rápido crescimento que vai se atenuando com o passar do tempo até a idade adulta, quando o processo tende a se estabilizar. Este processo pode ser bem representado por funções curvilíneas assintóticas, como: Função de Spillman, de Mitscherlich, de Gompertz, de Richards e Logística, dentre outras. Modelos de regressão linear têm aplicações nas mais diversas áreas do conhecimento. Entretanto, existem muitas situações em que modelos lineares podem não ser apropriados. Em aplicações mais realistas, especialmente nos casos de crescimento biológico, pode ser necessário ajustar funções não-lineares para melhor explicar o processo de crescimento. Neste caso, são referências úteis: Draper e Smith (*3*), Ratkowsky (*11*), Gallant (*4*), Bates e Watts (*1*), Cordeiro e Paula (*2*), Myers (*8*), Souza (*18*), Khattree e Naik (*6*), dentre outros.

Certamente um modelo de regressão não-linear é escolhido com base em considerações teóricas de um especialista na matéria. Assim, conhecimentos específicos de Química, Física ou Biologia podem conduzir automaticamente a um modelo para a função resposta. Muitos modelos de regressão não-linear pertencem a categorias delineadas para situações específicas ou ambientais. Talvez a categoria mais amplamente utilizada e conhecida de modelos não-lineares seja a dos modelos de crescimento. Estes descrevem o crescimento com mudanças na variável regressora. Freqüentemente a variável regressora é a idade ou o tempo. Aplicações típicas são em Biologia, quando plantas e organismos crescem com o tempo ou a idade, mas há também muitas aplicações em Economia e Engenharia. Por exemplo, o crescimento de uma planta, após sua emergência, em função do tempo, pode freqüentemente ser descrito por um modelo de regressão não-linear.

Fenômenos produzindo curvas sigmoidais na forma de S são freqüentemente encontrados na Agricultura, em Biologia, Ecologia, Engenharia e Economia. Essas curvas começam em algum ponto fixo e crescem monotonicamente até um ponto de inflexão, a partir daí a taxa de crescimento começa a diminuir até a curva se aproximar de um valor final, chamado de assíntota. No Quadro 1 são relacionados alguns modelos usuais com essa forma, conforme parametrização apresentada por Ratkowsky

(*11*). A denominação componente sistemático do modelo refere-se à parte fixa do modelo estatístico, isto é, sem o erro aleatório.

QUADRO 1 – Exemplos de alguns modelos do tipo sigmoidal com o correspondente componente sistemático

Modelo	Componente sistemático
Gompertz	$\alpha \exp\{-\exp(\beta - \gamma x)\}$
Logístico	$\alpha/\{1+\exp(\beta - \gamma x)\}$
Richards	$\alpha/\{[1+\exp(\beta - \gamma x)]^{1/\delta}\}$
Morgan-Mercer-Flodin (MMF)	$(\beta\gamma + \alpha x^{\delta})/(\gamma + x^{\delta})$
Weibull	$\alpha - \beta \exp(-\gamma x^{\delta})$

Fonte: Ratkowsky (*11*)

Nesses modelos o parâmetro α é o valor máximo esperado para a resposta, ou assíntota. O parâmetro β está relacionado com o intercepto, isto é, com o valor de E(y) correspondente a x=0. Em todos os modelos do Quadro 1 esse parâmetro pelo menos determina o intercepto. O parâmetro γ está relacionado com a taxa média de crescimento da curva e, finalmente, o parâmetro δ, que aparece em alguns modelos, é utilizado para aumentar a flexibilidade destes no ajuste aos dados.

Em algumas aplicações a resposta esperada E(y) é dada pela solução de um conjunto de equações diferenciais lineares. Estes modelos são freqüentemente chamados de modelos de compartimento, e uma vez que reações químicas às vezes podem ser descritas por um sistema linear de equações diferenciais de primeira ordem, eles têm aplicações freqüentes em Química, Engenharia Química e Farmacocinética. Em outras situações específicas, a função resposta obtida com a solução de equação diferencial não-linear ou para equação integral não tem solução analítica.

Quanto aos métodos de estimação de parâmetros, existem vários procedimentos numéricos para resolver problemas de mínimos quadrados não-lineares. Um método amplamente usado em algoritmos computacionais na regressão não-linear é o iterativo de Gauss-Newton, que se baseia numa aproximação por uma série de Taylor de primeira ordem para produzir uma linearização da função não-linear. O procedimento básico de Gauss-Newton no caso não-linear pode convergir muito lentamente em alguns casos, exigindo muitas iterações. Em outras situações ele pode mover na direção contrária, aumentando a soma de quadrado residual, ou pode não convergir. Várias modificações no algoritmo básico de Gauss-Newton têm sido propostas para melhorar sua performance. O método de Gauss-Newton

modificado, que é um procedimento disponível no PROC NLIN do SAS (*15*), encontra-se descrito por Souza (*18*). Outra modificação do algoritmo básico de Gauss-Newton foi desenvolvida por Marquardt (*7*). Este procedimento é freqüentemente chamado de Compromisso de Marquardt ("Marquardt Compromisse"), porque o vetor de incrementos produzido pelo seu método está entre o vetor de Gauss-Newton e o do método do gradiente ("steepest descent").

Neste trabalho, foi considerado o ajustamento de g equações de regressão não-linear (g grupos), com o objetivo de apresentar uma metodologia adequada para testar as seguintes hipóteses: (a) H_0: as g equações são idênticas, isto é, a equação comum pode ser usada como uma estimativa das g equações envolvidas; e (b) H_0 : um determinado subconjunto de parâmetros é igual nos g grupos.

METODOLOGIA E RESULTADOS

Embora haja vários modelos não-lineares para curvas de crescimento, considere-se, inicialmente, o ajustamento dos dados de observação relativos a g equações de regressão não-linear (g grupos), supondo o modelo logístico.

Seja o modelo de crescimento logístico com erro aditivo e a seguinte parametrização:

$$y_{ij} = \frac{a_i}{1 + b_i \exp(-c_i x_{ij})} + \varepsilon_{ij} , \quad \text{com} \quad j=1, \ldots, n_i , \; i=1, \ldots, g \qquad (1)$$

$$a_i , b_i , c_i > 0$$

em que

y_{ij} = valor observado na j-ésima unidade experimental do i-ésimo grupo;

x_{ij} = valor da variável independente ou covariável (por exemplo: idade, tempo etc.) associado a y_{ij};

a_i = para cada grupo i, é o valor máximo esperado para a resposta (assíntota), ou seja, a_i é o limite da esperança de y_{ij} quando a covariável tende a infinito;

b_i = parâmetro para o i-ésimo grupo que está relacionado com o valor de $E(y_{ij})$ correspondente a $x_{ij} = 0$, isto é, $E(y_{ij} / x_{ij} = 0) = a_i / (1 + b_i)$;

c_i = parâmetro para o i-ésimo grupo que está relacionado com a taxa média de crescimento da curva; e

ε_{ij} = erro aleatório com as pressuposições usuais, $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$.

$\sum_{i=1}^{g} n_i = n$ é o número total de observações.

As hipóteses que serão consideradas são as seguintes:

1. $H_o^{(1)} : a_1 = \ldots = a_g (= a)$ vs. $H_a^{(1)}$: nem todos a_i são iguais.
2. $H_o^{(2)} : b_1 = \ldots = b_g (= b)$ vs. $H_a^{(2)}$: nem todos b_i são iguais.
3. $H_o^{(3)} : c_1 = \ldots = c_g (= c)$ vs. $H_a^{(3)}$: nem todos c_i são iguais.
4. $H_o^{(4)} : a_1 = \ldots = a_g (= a)$ e $c_1 = \ldots = c_g (= c)$ vs. $H_a^{(4)}$: pelo menos uma igualdade é uma desigualdade.
5. $H_o^{(5)} : a_1 = \ldots = a_g (= a)$, $b_1 = \ldots = b_g (= b)$ e $c_1 = \ldots = c_g (= c)$ vs. $H_a^{(5)}$: pelo menos uma igualdade é uma desigualdade.

Sejam as variáveis "dummy"

$$D_i = \begin{cases} 1 \text{ se a observação } y_{ij} \text{ pertence ao grupo } i, \\ 0 \text{ em caso contrário;} \end{cases} \qquad i = 1, \ldots, g.$$

Então o modelo da equação (1) pode ser escrito como:

$$y_{ij} = \sum_{u=1}^{g} D_u \left[\frac{a_u}{1 + b_u \exp(-c_u x_{ij})} \right] + \varepsilon_{ij}, \; j = 1, \ldots, n_i, i = 1, \ldots, g.$$

O problema é comparar os vários grupos. O teste da razão de verossimilhança pode ser usado para testar as hipóteses formuladas inicialmente. Estimativas de máxima verossimilhança dos parâmetros desconhecidos podem ser obtidas usando o PROC NLIN do SAS (*15*).

Seja o modelo da equação (1) escrito como

$$y_{ij} = \mu(a_i, b_i, c_i, x_{ij}) + \varepsilon_{ij},$$

em que

$$\mu(a_i, b_i, c_i, x_{ij}) = \frac{a_i}{1 + b_i \exp(-c_i x_{ij})}$$

. Então, a função de verossimilhança dos parâmetros, dadas as observações y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, g$, é

$$f(\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[y_{ij} - \mu(a_i, b_i, c_i, x_{ij})\right]^2\right\} \qquad (2)$$

Aqui, θ representa todos os parâmetros $a_1,\ldots,a_g,b_1,\ldots,b_g,c_1,\ldots,c_g$ e σ^2. Para σ^2 fixo, maximizar $f(\theta)$ na equação (2) com respeito aos parâmetros $a_i, b_i, c_i, i = 1,\ldots,g$ é o mesmo que minimizar

$$S(a_1,\ldots,a_g,b_1,\ldots,b_g,c_1,\ldots,c_g) = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[y_{ij} - \mu(a_i, b_i, c_i, x_{ij})\right]^2$$

com respeito aos parâmetros correspondentes, e esta minimização é feita iterativamente. Uma vez obtidas estas estimativas, a estimativa de máxima verossimilhança de σ^2 é dada por

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{a}_1,\ldots,\hat{a}_g,\hat{b}_1,\ldots,\hat{b}_g,\hat{c}_1,\ldots,\hat{c}_g),$$

na qual $\hat{a}_1,\ldots,\hat{a}_g,\hat{b}_1,\ldots,\hat{b}_g,\hat{c}_1,\ldots,\hat{c}_g$ são as estimativas obtidas pelo método dos mínimos quadrados não-linear a partir da minimização descrita anteriormente.

Considere o problema geral de testar a hipótese de nulidade: $H_o : \theta \in w$ versus $H_a : \theta \in w^c$, em que w é um subconjunto do espaço paramétrico Ω e w^c é o complemento de w, com $\Omega = w \cup w^c$. A estatística do teste da razão de verossimilhança para este problema é

$$L = \left(\frac{\hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{w}}\right)^{n/2},$$

sendo $\hat{\sigma}^2_{\Omega}$ a estimativa de máxima verossimilhança de σ^2 quando nenhuma restrição no espaço paramétrico é feita e $\hat{\sigma}^2_{w}$ é a estimativa de máxima verossimilhança de σ^2 quando as restrições lineares colocadas em H_o são impostas no espaço paramétrico Ω. Em grandes amostras de tamanho n, a distribuição de $-2 \ln L$ é aproximadamente qui-quadrado com ν graus de liberdade, em que ν é o número de parâmetros estimados em Ω menos o número de parâmetros estimados em w, conforme Rao (*10*).

$$\text{Assim, } -2\ln L = -n \ln\left(\frac{\hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{w}}\right) \xrightarrow[n\to\infty]{d} \chi^2_{\nu}.$$

Para aplicação do teste de uma forma ainda mais clara, pode-se seguir os seguintes passos:

1: Ajustar o modelo completo Ω e obter $\hat{\sigma}^2_{\Omega} = \frac{SQR_{\Omega}}{n}$

SQR_{Ω} é a soma de quadrados residual para o modelo completo Ω;

e

p_{Ω} é o número de parâmetros estimados em Ω.

2: Ajustar o modelo reduzido w (modelo sob a restrição dada por H_0) e obter $\hat{\sigma}^2_{w} = \frac{SQR_{w}}{n}$

SQR_{w} é a soma de quadrados residual para o modelo reduzido w; e

p_{w} é o número de parâmetros estimados em w

3: Obter a estatística do teste

$$\chi^2_{calculado} = -n \ln \left(\frac{\hat{\sigma}^2_{\Omega}}{\hat{\sigma}^2_{w}} \right),$$

ou ainda,

$$\chi^2_{calculado} = -n \ln \left(\frac{SQR_{\Omega}/n}{SQR_{w}/n} \right) = -n \ln \left(\frac{SQR_{\Omega}}{SQR_{w}} \right)$$

4: Regra de decisão

Se $\chi^2_{calculado} \geq \chi^2_{tabelado} \Rightarrow$ Rejeita-se H_o. Caso contrário, não se rejeita H_o.

O valor tabelado é função do nível de significância α e do número de graus de liberdade $\nu = p_{\Omega} - p_{w}$.

Quanto ao teste de $H_o^{(1)}, H_o^{(2)}, H_o^{(3)}, H_o^{(4)}$ e $H_0^{(5)}$ contra as hipóteses alternativas correspondentes, as estatísticas do teste da razão de verossimilhança que têm aproximadamente distribuição de qui-quadrado são:

$$-n \ln \left(\hat{\sigma}^2_{\Omega} / \hat{\sigma}^2_{w_i} \right) = -n \ln \left(\frac{SQR_{\Omega}}{SQR_{w_i}} \right), i = 1, 2, 3, 4, 5.$$

Cada uma destas estatísticas tem aproximadamente distribuição de qui-quadrado com $(g-1)$, $(g-1)$, $(g-1)$, $2(g-1)$ e $3(g-1)$ graus de liberdade, respectivamente. Os w_i são os respectivos subconjuntos do

espaço paramétrico Ω definidos pelas hipóteses de nulidade $H_o^{(i)}, i = 1,...,5$. Dadas as estimativas iniciais $\theta = \theta_0$ e a forma da função $\mu(\theta, x)$, o PROC NLIN ajusta um modelo do tipo $y = \mu(\theta, x) + \varepsilon$ a partir dos dados $(y_i, x_i), i = 1,...,n$ e fornece as estimativas de mínimos quadrados de θ, por métodos iterativos.

ILUSTRAÇÃO DO MÉTODO

Julgou-se adequada a ilustração dos resultados apresentados neste estudo. Assim, com base nos dados do Quadro 2, foram efetuados os cálculos, ilustrando os procedimentos apresentados.

Dado que $g = 2$, introduziram-se mais duas variáveis independentes D_1 e D_2, para identificar os grupos 1 e 2, respectivamente. O modelo completo adotado para analisar estes dados foi

$$y_{ij} = D_1\left[\frac{a_1}{1 + b_1 exp(-c_1 x_{ij})}\right] + D_2\left[\frac{a_2}{1 + b_2 exp(-c_2 x_{ij})}\right] + \varepsilon_{ij} \qquad (3)$$

QUADRO 2 – Matéria seca total $(g \cdot m^{-2})$ de uma cultura de milho, em períodos de 15 a 135 dias após a emergência, em duas condições (grupos 1 e 2)

Dias após emergência	Matéria seca total $(g \cdot m^{-2})$	
	Grupo 1	Grupo 2
15	41,4	80,5
30	161,7	200,6
45	564,5	580,0
60	1288,6	1250,4
75	1430,1	1500,2
90	1752,6	1920,3
105	2354,2	2380,0
120	2453,4	2450,8
135	2469,6	2520,7

O objetivo deste estudo foi ajustar o modelo logístico aos dados de matéria seca total de uma cultura de milho dos grupos 1 e 2, separadamente. O interesse foi também determinar se uma única equação para os dois grupos poderia ser considerada adequada e, caso contrário, se

havia certos parâmetros no modelo que poderiam ser considerados o mesmo nos dois grupos.

A codificação em SAS para calcular as estimativas de máxima verossimilhança dos parâmetros sob Ω (nenhuma restrição no espaço paramétrico), w_1 (espaço paramétrico restrito por: $a_1 = a_2 = a$), w_2 (espaço paramétrico restrito por: $b_1 = b_2 = b$), w_3 (espaço paramétrico restrito por: $c_1 = c_2 = c$), w_4 (espaço paramétrico restrito por $a_1 = a_2 = a$ e $c_1 = c_2 = c$) e w_5 (espaço paramétrico restrito pelas restrições em w_1, w_2 e w_3 conjuntamente) é dada no programa SAS. Em cada caso, o PROC NLIN requer os valores iniciais das estimativas dos parâmetros com a opção PARMS. No modelo em questão, os valores iniciais $\hat{a}$, $\hat{b}$ e $\hat{c}$, em cada grupo, podem ser obtidos como a seguir.

1° passo: Obtenção de $\hat{a}_1$

Sejam x_A, x_B e x_C três valores eqüidistantes de x, isto é, $x_C - x_B = x_B - x_A = \Delta x$, e y_A, y_B e y_C, os correspondentes valores de y. Tais valores podem pertencer ou não à amostra disponível. Caso não pertençam, marcam-se em um gráfico todos os pontos correspondentes às observações da amostra, traça-se a curva logística "a olho" e lêem-se neste gráfico as coordenadas dos três pontos escolhidos, cujas posições no eixo das abscissas sejam equidistantes. No grupo 1, utilizando-se os períodos de crescimento $x_A = 15$, $x_B = 75$ e $x_C = 135$ dias após a emergência e as correspondentes produções de matéria seca total $y_A = 41{,}4$, $y_B = 1430{,}1$ e $y_C = 2469{,}6$, calculou-se a estimativa inicial $\hat{a}_1$ (estimativa assintótica da produção de matéria seca máxima), que é dada por:

$$\hat{a}_1 = \frac{y_B(y_A y_B + y_B y_C - 2y_A y_C)}{y_B^2 - y_A y_C} = 2492{,}6244 .$$

Uma alternativa simples para o valor inicial $\hat{a}_1$ consiste em tomar um valor um pouco maior do que o maior valor observado de y.

2° passo: Obtenção de $\hat{b}_1$ e $\hat{c}_1$

Note que a função resposta pode ser linearizada facilmente. Então, segue que:

$$y_i = \frac{\hat{a}_1}{1 + b_1 e^{-c_1 x_i}}$$

$$\ln\left(\frac{\hat{a}_1 - y_i}{y_i}\right) = \ln b_1 - c_1 x_i \quad , \quad y_i > 0 \ \text{ e } \ \hat{a}_1 - y_i > 0$$

para $i = 1, \ldots, n_1$.

Fazendo $z_i = \ln\left(\frac{\hat{a}_1 - y_i}{y_i}\right)$ e ajustando-se o modelo de regressão linear simples

$z_i = A_1 + B_1 x_i + e_i$, obtêm-se os valores iniciais do seguinte modo:

Equação ajustada: $\hat{z}_i = \hat{A}_1 + \hat{B}_1 x_i$. Assim, $\hat{b}_1 = e^{\hat{A}_1}$ e $\hat{c}_1 = -\hat{B}_1$.

No exemplo, obteve-se $\hat{b}_1 = e^{4,81921} = 123,8672$ e $\hat{c}_1 = 0,07151$.

No grupo 2, tomando-se $x_A = 15$, $x_B = 75$ e $x_C = 135$, com $y_A = 80,5$, $y_B = 1500,2$ e $y_C = 2520,7$ foram obtidas as estimativas iniciais $\hat{a}_2$, b_2 e $\hat{c}_2$ de modo análogo ao grupo 1. Os resultados obtidos foram: $\hat{a}_2 = 2561,6411$, $\hat{b}_2 = e^{4,18415} = 65,6377$ e $\hat{c}_2 = 0,06181$.

As estimativas iniciais dos modelos restritos são obtidas por uma simples inspeção apropriada. Por exemplo, em w_1 a média das estimativas iniciais de a_1 e a_2 é usada como estimativa inicial de a, uma vez que em $H_o^{(1)}$ tem-se a restrição $a_1 = a_2 = a$.

Na sentença MODEL do PROC NLIN, a forma explícita da equação 3 (suprimindo o termo erro) é colocada. Há vários métodos iterativos disponíveis no procedimento NLIN para o ajuste do modelo, alguns dos quais são baseados nas derivadas e exigem especificação explícita das derivadas parciais. Se nenhuma das derivadas é colocada, o procedimento usa o "default" DUD ("Secant Method"), em que as derivadas são estimadas pelo programa. O número máximo de iterações é especificado em MAXITER = opção.

Executando-se o programa, têm-se as estimativas de cada um dos parâmetros e um resumo da análise de variância da regressão. A estimativa de máxima verossimilhança de σ^2, isto é, $\hat{\sigma}^2$, é obtida pela soma de quadrados residual dividida por n. Assim, executando-se o procedimento NLIN com nenhuma restrição e com várias restrições especificadas por w_1

a w_5, obtêm-se $\hat{\sigma}^2_{\Omega}$, $\hat{\sigma}^2_{w_1}$, ..., $\hat{\sigma}^2_{w_5}$. Estas estimativas são usadas para obter o teste estatístico das várias hipóteses descritas.

Programa SAS

```
title ' Função Logística: Y = a/(1+b*exp(-c*X)) ';
data milho;
input grupo X Y d1 d2;
cards;
1   15     41.4  1  0
1   30    161.7  1  0
1   45    564.5  1  0
1   60   1288.6  1  0
1   75   1430.1  1  0
1   90   1752.6  1  0
1  105   2354.2  1  0
1  120   2453.4  1  0
1  135   2469.6  1  0
2   15     80.5  0  1
2   30    200.6  0  1
2   45    580.0  0  1
2   60   1250.4  0  1
2   75   1500.2  0  1
2   90   1920.3  0  1
2  105   2380.0  0  1
2  120   2450.8  0  1
2  135   2520.7  0  1
;
/*  Programa para o modelo completo omega  */
/*  Estimativas iniciais foram tomadas para cada grupo separadamente  */
proc nlin data=milho maxiter=100;
parms a1=2492.6244  b1=123.8672  c1=0.07151
      a2=2561.6411  b2= 65.6377  c2=0.06181;
model Y = d1*(a1/(1+b1*exp(-c1*X)))+d2*(a2/(1+b2*exp(-c2*X)));
run;

/*  Modelo com restrição w1:a1=a2=a  */
/*  Estimativa inicial de a foi tomada como a média de a1 e a2  */
proc nlin data=milho maxiter=100;
```

```
parms a=2527.13275  b1=123.8672  c1=0.07151  b2= 65.6377
c2=0.06181;
model Y = d1*(a/(1+b1*exp(-c1*X)))+d2*(a/(1+b2*exp(-c2*X)));
run;

/*  Modelo com restrição w2:b1=b2=b  */
/*  Estimativa inicial de b foi tomada como a média de b1 e b2  */
proc nlin data=milho maxiter=100;
parms a1=2492.6244  b=94.75245  c1=0.07151  a2=2561.6411
c2=0.06181;
model Y = d1*(a1/(1+b*exp(-c1*X)))+d2*(a2/(1+b*exp(-c2*X)));
run;

/*  Modelo com restrição w3:c1=c2=c  */
/*  Estimativa inicial de c foi tomada como a média de c1 e c2  */
proc nlin data=milho maxiter=100;
parms a1=2492.6244  b1=123.8672  c=0.06666  a2=2561.6411
b2=65.6377;
model Y = d1*(a1/(1+b1*exp(-c*X)))+d2*(a2/(1+b2*exp(-c*X)));
run;

/*  Modelo com restrição w4:a1=a2 e c1=c2 */
/*  Estimativa inicial de a e c foi tomada como as médias dos ai e ci,
 respectivamente  */
proc nlin data=milho maxiter=100;
parms a=2527.13275  b1=123.8672  b2=65.6377  c=0.06666;
model Y = d1*(a/(1+b1*exp(-c*X)))+d2*(a/(1+b2*exp(-c*X)));
run;

/*  Modelo com restrição w5:a1=a2, b1=b2 e c1=c2  */
/*  Estimativa inicial de a, b e c foi tomada como as médias dos ai, bi e ci,
respectivamente  */
proc nlin data=milho maxiter=100;
parms a=2527.13275  b=94.75245  c=0.06666;
model Y = d1*(a/(1+b*exp(-c*X)))+d2*(a/(1+b*exp(-c*X)));
run;
quit;
```

No Quadro 3 estão apresentadas as estimativas dos parâmetros do modelo com nenhuma restrição no espaço paramétrico (Ω) e com várias restrições especificadas por w_1 a w_5.

No Quadro 4 estão apresentados os resultados do teste para as cinco hipóteses formuladas. Uma vez que o valor–p foi grande em todas as hipóteses, neste caso não se rejeita nenhuma delas. Como o teste de H_o [5] foi não-significativo, pode-se concluir que as duas equações não diferem estatisticamente. Assim, a equação comum, cujas estimativas estão apresentadas no Quadro 3 (modelo w_5), pode ser usada como uma estimativa das duas equações envolvidas, obtendo-se assim uma única curva de crescimento para os dois grupos.

Quando se utiliza a análise de regressão para definir a relação funcional entre variáveis, defronta-se com o problema da especificação, ou seja, a determinação da forma matemática da função que será ajustada, que pode ser feita utilizando-se o conhecimento que se tem *a priori* sobre o fenômeno e o conhecimento adquirido pela inspeção dos dados numéricos disponíveis. Freqüentemente ajusta-se mais de um modelo e, com base nos resultados e testes estatísticos, escolhe-se aquele que melhor se ajusta aos dados e melhor representa o fenômeno que estiver sendo estudado.

Em modelos de regressão linear que incluem o termo constante (intercepto), o coeficiente de determinação R^2 representa a proporção da variação explicada pelo modelo. Neste caso, o quadrado do coeficiente de correlação entre os valores observados e preditos é exatamente o R^2. Se o modelo é linear e o termo constante não está presente (sem intercepto), o R^2 é redefinido conforme Searle (*16*), e muito cuidado deve ser tomado na sua interpretação, pois ele não é mais igual ao quadrado do coeficiente de correlação entre os valores observados e preditos. Pode ocorrer que o valor do coeficiente de determinação, na versão sem intercepto, domine em muito o valor correspondente ao caso com intercepto, em modelos equivalentes.

Segundo Souza (*18*), no caso do modelo de regressão não-linear, a adequacidade do ajustamento pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos. Diz ainda que esta medida pode ser calculada com a utilização da fórmula $R^2 = [1 - (SQR/SQTotal_c)]$, na qual SQR é soma dos quadrados residuais e $SQTotal_c$, a soma de quadrados total corrigida pela média. O que se observa, na prática, é que, em muitos trabalhos de pesquisa, no caso não-linear o cálculo do R^2 não é feito de uma única maneira. Alguns utilizam a fórmula apresentada por Souza (*18*), outros empregam a fórmula $R^2 = [1 - (SQR/SQTotal_{nc})]$, na qual SQR é a soma

QUADRO 3 – Estimativas dos parâmetros do modelo irrestrito (Ω) e modelos restritos $(w_1$ a $w_5)$ e respectivas somas de quadrados residuais

Parâmetros	Estimativas dos parâmetros dos modelos Ω, w_1, w_2, w_3, w_4 e w_5					
	Ω	w_1	w_2	w_3	w_4	w_5
a_1	2562,1	-	2555,0	2545,1	-	-
b_1	36,6609	36,3323	-	39,1989	40,0903	-
c_1	0,0529	0,0527	0,0537	-	-	-
a_2	2573,5	-	2580,4	2589,3	-	-
b_2	40,2158	40,5678	-	37,4823	36,4386	-
c_2	0,0556	0,0558	0,0548	-	-	-
a	-	2568,4	-	-	2568,9	2566,9
b	-	-	38,3302	-	-	38,3269
c	-	-	-	0,0542	0,0541	0,0542
* $n\hat{\sigma}^2$	243478	243545	243901	244559	246576	252640

* Soma dos quadrados residuais, n=18 e $\hat{\sigma}^2 = \frac{SQR}{n} = \frac{n-p}{n} \cdot QMR$, sendo p o número de parâmetros estimados e QMR o quadrado médio do resíduo da ANOVA fornecida pelo SAS.

QUADRO 4 – Hipóteses avaliadas, valores da estatística do teste qui-quadrado, número de graus de liberdade e nível descritivo do teste (valor-p)

Hipóteses	$\chi^2_{calculado}$ *	Número de graus de liberdade (ν)	Valor-p $P(\chi^2_\nu > \chi^2_{calculado})$
$H_o^{(1)} : a_1 = a_2 = a$	0,0049	1	0,9442
$H_o^{(2)} : b_1 = b_2 = b$	0,0312	1	0,8598
$H_o^{(3)} : c_1 = c_2 = c$	0,0797	1	0,7777
$H_o^{(4)} : a_1 = a_2 = a$ e $c_1 = c_2 = c$	0,2276	2	0,8924
$H_o^{(5)} : a_1 = a_2 = a$, $b_1 = b_2 = b$ e $c_1 = c_2 = c$	0,6649	3	0,8814

* $-n\ln\left(\frac{n\hat{\sigma}^2_\Omega}{n\hat{\sigma}^2_{w_i}}\right)$ do Quadro 3.

dos quadrados residuais e $SQTotal_{nc}$, a soma de quadrados total não-corrigida pela média. Com estes cálculos, às vezes se obtêm valores extremamente altos, por exemplo $R^2 = 99\%$, mesmo havendo enorme discrepância entre os valores observados e preditos. O fato é que, independentemente de haver ou não um termo constante no modelo, o R^2 não tem nenhum significado óbvio no caso de modelos de regressão não-linear e, segundo Ratkowsky (*12*), ele nunca precisa ser calculado.

Para decidir se um modelo de regressão não-linear se ajustou bem aos dados é interessante atentar para os valores observados e preditos para verificar se não há muita discrepância, fazer uma análise gráfica dos resíduos, tendo os devidos cuidados, pois esta não é uma simples extensão do caso linear, e verificar a magnitude da variância residual obtida por máxima verossimilhança, para decidir se ela é suficientemente pequena. Uma vez atendidas as pressuposições usuais da análise para julgar a qualidade do modelo ajustado, pode-se usar a raiz quadrada da estimativa de máxima verossimilhança da variância residual, isto é, a raiz do quadrado médio do resíduo (RQMR), dada por:

$$RQMR = \left[\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 / n\right]^{1/2},$$

em que Y_i e $\hat{Y}_i$ são os valores observados e estimados, respectivamente. Evidentemente, quanto menor for o valor de RQMR melhor será o ajuste.

CONCLUSÕES

1) A identidade de modelos de regressão não-linear e a igualdade de qualquer subconjunto de parâmetros podem ser verificadas por meio do teste da razão de verossimilhança.

2) A metodologia apresentada é geral e pode ser usada em qualquer modelo de regressão não-linear.

REFERÊNCIAS

1. BATES, D. M. & WATTS, D.G. Nonlinear regression analysis and its applications. New York, John Wiley & Sons, 1988. 365p.
2. CORDEIRO, G. M. & PAULA, G.A. Modelos de regressão para análise de dados univariados. Rio de Janeiro, Instituto de Matemática Pura e Aplicada do CNPq, 1989. 353p.
3. DRAPER, N. R. & SMITH, H. Applied regression analysis. 2ª ed. New York, John Wiley e Sons, 1981. 709p.
4. GALLANT, A. R. Nonlinear statistical models. New York, John Wiley & Sons, 1987. 611p.

5. GRAYBILL, F.A. Theory and application of the linear model. Belmont, Duxbury Press, 1976. 704p.
6. KHATTREE, R. & NAIK, D. N. Applied multivariate statistical with SAS software. 2ª ed. Cary, NC, SAS Institute Inc., 1999. 338p.
7. MARQUARDT, D. W. An algorithm for least squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Maths., 2 : 431-41, 1963.
8. MYERS, R. H. Classical and modern regression with applications. 2ª ed. Boston, PWS-KENT Publishing Company, 1990. 488p.
9. NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J. & WASSERMAN, W. Applied linear statistical models. 4ª ed. USA, Richard D. Irwin, 1996. 1408 p.
10. RAO, C. R. Linear statistical inference and its applications. New York, John Wiley & Sons, 1973. 522p.
11. RATKOWSKY, D.A. Nonlinear regression modeling - A unified practical approach. New York and Basel, Marcel Dekker, 1983. 276p.
12. RATKOWSKY, D.A. Handbook of nonlinear regression models. New York and Basel, Marcel Dekker, 1990. 241p.
13. REGAZZI, A.J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. Revista Ceres, 40: 176-95, 1993.
14. REGAZZI, A.J. Teste para verificar a identidade de modelos de regressão e a igualdade de parâmetros no caso de dados de delineamentos experimentais. Revista Ccrcs, 46: 383-409, 1999.
15. SAS INSTITUTE INC. SAS/STAT User's guide. Version 6, 4ª ed., v. 2. Cary, NC, SAS Institute, 1990. 796p.
16. SEARLE, S. R. Linear models. New York, John Wiley & Sons, 1971. 532p.
17. STEEL, R.G.D. & TORRIE, J.H. Principles and procedures of statistics. New York, McGraw-Hill Book Company, 1980. 633p.
18. SOUZA, G.S. Introdução aos modelos de regressão linear e não-linear. Brasília, Embrapa - Serviço de Produção de Informação, 1998. 505p.