

# TESTE PARA VERIFICAR A IGUALDADE DE PARÂMETROS E A IDENTIDADE DE MODELOS DE REGRESSÃO NÃO-LINEAR.

## I. DADOS NO DELINEAMENTO INTEIRAMENTE CASUALIZADO

Adair José REGAZZI<sup>1</sup>  
Carlos Henrique Osório SILVA<sup>1</sup>

- RESUMO: Considerou-se o ajustamento de  $g$  equações de regressão não-linear e o teste da razão de verossimilhança, com aproximação dada pela estatística qui-quadrado, para se testar a identidade dos modelos e comparar, por meio de simulação de dados, as aproximações dadas pelas estatísticas  $\chi^2$  e  $F$ . Para um número total de observações suficientemente grande ( $N \geq 120$ ) as duas aproximações são praticamente equivalentes. Para amostras menores, a aproximação dada pela estatística  $F$  deve ser a preferida, uma vez que a taxa de erro tipo I foi sempre menor, independentemente do valor de  $N$ .
- PALAVRAS-CHAVE: Curvas de crescimento; verossimilhança; teste de hipótese.

### 1 Introdução

Em análise de regressão, com muita frequência deseja-se saber se um conjunto de  $g$  equações ajustadas são idênticas, ou seja, se o fenômeno em estudo pode ser representado por uma única equação. Vários autores apresentaram métodos para testar hipóteses relativas à identidade de modelos lineares, como por exemplo, Graybill (1976), Steel e Torrie (1980), Neter et al. (1996), Regazzi (1993, 1999), entre outros. Modelos de regressão linear têm aplicações nas mais diversas áreas do conhecimento. Muitas vezes, um modelo linear é utilizado apenas pela facilidade em descrever o relacionamento aproximado. Entretanto, o verdadeiro relacionamento entre uma variável dependente e uma ou mais variáveis independentes pode ser descrito por um modelo não-linear, determinado com base em conhecimentos teóricos do problema tratado. Portanto, existem muitas situações em que modelos lineares podem não ser apropriados. Um exemplo típico na área de Ciências Biológicas é a modelagem de crescimento, em que pode ser necessário ajustar funções não-lineares para melhor explicar o processo de crescimento. Nesse caso, são referências úteis: Draper e Smith (1981), Ratkowsky (1983), Gallant (1987), Bates e Watts (1988), Cordeiro e Paula (1989), Myers (1990), Souza (1998), Khattree e Naik (1999), dentre outros.

Bates e Watts (1988), para comparar modelos de regressão não-linear, apresentaram um teste assintótico baseado na razão de verossimilhança como no caso de modelos

---

<sup>1</sup> Departamento de Informática, Universidade Federal de Viçosa - UFV, CEP 36570-000, Viçosa, MG, Brasil.  
E-mail: [adairreg@ufv.br](mailto:adairreg@ufv.br) / [chos@dpi.ufv.br](mailto:chos@dpi.ufv.br).

lineares, com aproximação dada pela estatística  $F$ , conforme sumarizado na Tabela 1. A estatística do teste é então:

$$F_0 = \frac{(S_p - S_f) / v_e}{S_f / v_f} \sim F(v_e; v_f).$$

*Cabe ressaltar que para modelos de regressão não-linear esta análise é apenas aproximada, porque a razão entre os quadrados médios não tem uma distribuição F exata (Souza, 1998). O teste a um nível de significância  $\alpha$ , da hipótese  $H_0$  de que o modelo reduzido é adequado, consiste em rejeitar  $H_0$  se  $F_0 \geq F_{\alpha}(v_e; v_f)$ .*

Tabela 1 - Esquema da análise para execução do teste proposto por Bates e Watts (1988) sobre a igualdade de parâmetros em modelos de regressão não-linear<sup>1</sup>

Causas de variação	GL	SQ	QM	F <sub>0</sub>
Parâmetros extras	$v_e = P_f - P_p$	$S_e = S_p - S_f$	$s_e^2 = S_e / v_e$	$s_e^2 / s_f^2$
Modelo completo	$v_f = N - P_f$	$S_f$	$s_f^2 = S_f / v_f$	
Modelo reduzido	$v_p = N - P_p$	$S_p$		

<sup>1</sup>S= Soma de quadrados residuais, v = Número de graus de liberdade, P= Número de parâmetros e N = Número total de observações. Subscritos: f = modelo completo, p = modelo reduzido, e = extra.

Regazzi (2003), utilizando o teste da razão de verossimilhança conforme Rao (1973), apresentou em detalhes a aplicação desse teste em modelos de regressão não-linear, com aproximação dada pela estatística qui-quadrado ( $\chi^2$ ). Cabe ressaltar que Regazzi (2003) admitiu apenas um valor observado de Y para cada um de X.

Considerando agora o caso de dados com repetições, o objetivo do presente trabalho é apresentar o teste da razão de verossimilhança para testar a identidade de modelos de regressão não-linear, com aproximação dada pela estatística  $\chi^2$  e comparar as estatísticas F (Bates e Watts, 1988) e  $\chi^2$  (Rao, 1973) por simulação de dados.

## 2 Metodologia e resultados

### 2.1 Dados provenientes de um delineamento inteiramente casualizado

#### 2.1.1 Preliminares

Considere inicialmente um experimento com I tratamentos, sendo I níveis de um fator quantitativo, cada um com  $r_i$  ( $i=1, \dots, I$ ) repetições, no delineamento com um fator inteiramente casualizado.

Do ponto de vista da estatística experimental, e considerando-se um modelo de regressão não-linear com p parâmetros, o esquema da análise da variância para testar a falta de ajustamento do modelo é como apresentado na Tabela 2.

Tabela 2 - Esquema da análise da variância

Causas de variação	GL <sup>1</sup>	QM
Parâmetros da regressão	$p$	
Falta de ajustamento	$I - p$	QMFA
Parâmetros do delineamento	$(I)$	
Resíduo (Erro puro)	$n - I$	QMR
Total não corrigido	$n$	

$$^1 n = \sum_{i=1}^n r_i$$

É fato que, no delineamento com um fator inteiramente casualizado, SQParâmetros da regressão + SQFalta de ajustamento = SQParâmetros do delineamento. Se a matriz do modelo linear aproximante (via matriz Jacobiana) tiver uma coluna de “uns”, então, SQRegressão = SQParâmetros da regressão – C, onde

$$C = \frac{(\sum_{i=1}^n y_i)^2}{n}.$$

Neste caso, SQRegressão + SQFalta de ajustamento = SQTratamentos, pois SQtratamentos = SQParâmetros do delineamento – C. Assim teríamos,  $p-1$ ,  $I-p$  e  $I-1$  graus de liberdade associados às causas de variação devido à regressão, falta de ajustamento e tratamentos, respectivamente. No esquema da Tabela 2, o resíduo da regressão que tem  $n-p$  graus de liberdade foi decomposto em falta de ajustamento mais o erro puro. O teste  $F$  aproximado (assintótico) para a falta de ajustamento do modelo é dado por:

$$F = \frac{QMFA}{QMR} \sim F(I - p; n - I).$$

É sempre desejável que o teste para falta de ajustamento do modelo dê um resultado não-significativo. Nos casos em que a falta de ajustamento é significativa, pode-se concluir que o modelo utilizado não é apropriado. Assim, deve-se modificar a função esperança a fim de obter-se uma função mais apropriada.

Considerando-se a análise de variância com um enfoque em Estatística Experimental, em modelos de regressão linear, o teste  $F$  para falta de ajustamento e para regressão, tem como denominador o quadrado médio do resíduo ou erro puro ( $QMR$ , Tabela 2). Quando a falta de ajustamento é não-significativa, pode-se optar em combinar a falta de ajustamento e o erro puro, obtendo-se o quadrado médio do resíduo da regressão ( $QMRR$ ) a ser utilizado como denominador no teste  $F$  para regressão. Para modelos lineares, os testes de Wald e da razão de verossimilhança são equivalentes teórica e numericamente. Para modelos de regressão não-linear, teoricamente os testes de Wald e da razão de verossimilhança são assintoticamente equivalentes, desde que certas condições, comumente denominadas “de regularidade”, sejam atendidas. Mas em pequenas amostras, os resultados assintóticos aproximam de modo distinto cada procedimento (Souza, 1998).

### 2.1.2 Modelo estatístico

Considerou-se, inicialmente o ajustamento dos dados de observação relativos a  $g$  grupos ( $g$  diferentes condições experimentais ou  $g$  variedades etc.). Para cada um deles, foi ajustado o modelo Michaelis-Menten com erro aditivo dado por:

$$y_{ij} = \frac{a_i x_{ij}}{b_i + x_{ij}} + \varepsilon_{ij}, \quad \text{com } j=1, \dots, n_i, \quad i=1, \dots, g \quad (1)$$

em que

- $y_{ij}$  é o valor observado na  $j$ -ésima unidade experimental do  $i$ -ésimo grupo;
- $x_{ij}$  é o valor da variável independente associado a  $y_{ij}$
- $\varepsilon_{ij}$  é o erro aleatório não observável do modelo.

O parâmetro  $a_i$  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  $x_{ij}$  tende a infinito;  $b_i$  é o parâmetro para o  $i$ -ésimo grupo, tal que, quando  $x_{ij} = b_i$ ,  $y_{ij}$  é a metade do valor máximo esperado.

A metodologia que será apresentada pressupõe que  $\varepsilon_{ij} \sim NID(0, \sigma^2)$ . Temos ainda que

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

Como foram admitidos  $g$  grupos e se pressupõe homocedasticidade, sua verificação poderá ser feita mediante um teste de homogeneidade de variâncias apropriado, como por exemplo, o teste de Bartlett, citado por Li (1964). Essa homogeneidade de variâncias é sobre os resíduos da regressão para os  $g$  grupos. Conforme se verifica em Pimentel-Gomes (1990), quando se tem uma relação de variâncias (max.  $QMRR$  / min.  $QMRR$ ) menor que sete, é quase sempre possível combinar as variâncias residuais, obtendo-se uma estimativa comum. Quando esse quociente for além de sete, convirá considerar separadamente subgrupos de modelos, em que se tenha, dentro de cada subgrupo, uma razoável homogeneidade de variâncias.

As hipóteses que serão consideradas são as seguintes:

- $H_o^{(1)} : b_1 = \dots = b_g (= b)$  vs.  $H_a^{(1)} : \text{nem todos } b_i \text{ são iguais.}$
- $H_o^{(2)} : a_1 = \dots = a_g (= a)$  vs.  $H_a^{(2)} : \text{nem todos } a_i \text{ são iguais.}$
- $H_o^{(3)} : b_1 = \dots = b_g (= b) \text{ e } a_1 = \dots = a_g (= a)$  vs.  $H_a^{(3)} : \text{pelo menos uma igualdade é uma desigualdade.}$

Para  $i = 1, \dots, g$ , sejam as variáveis “dummy”  $D_i$  tais que:

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

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

$$y_{ij} = \sum_{u=1}^g D_u \left[ \frac{a_u x_{ij}}{b_u + x_{ij}} \right] + \varepsilon_{ij}, \quad j=1, \dots, n_i, i=1, \dots, g$$

O problema é comparar as equações ajustadas para os vários grupos. Para testar as hipóteses formuladas, a metodologia apresentada se baseia no teste da razão de verossimilhança, com aproximação dada pela estatística  $\chi^2$ . No presente trabalho as estimativas de máxima verossimilhança dos parâmetros foram obtidas utilizando-se o procedimento PROC NLIN do SAS\* (1990). Maiores detalhes podem ser vistos em Regazzi (2003).

Considere o problema geral de testar a hipótese de nulidade:  $H_0: \theta \in \omega$  versus  $H_a: \theta \in \omega^c$ , onde  $\omega$  é um subconjunto do espaço paramétrico  $\Omega$  e  $\omega^c$  é o complemento de  $\omega$ , com  $\Omega = \omega \cup \omega^c$ . A estatística do teste da razão de verossimilhança para este problema é

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

sendo  $\hat{\sigma}_{\Omega}^2$  a estimativa de máxima verossimilhança de  $\sigma^2$  com nenhuma restrição no espaço paramétrico e  $\hat{\sigma}_{\omega}^2$  a estimativa de máxima verossimilhança de  $\sigma^2$  com as restrições lineares colocadas em  $H_0$ . Para grandes amostras de tamanho  $N$ , conforme Rao (1973), tem-se que

$$-2 \ln L = -N \ln \left( \frac{\hat{\sigma}_{\Omega}^2}{\hat{\sigma}_{\omega}^2} \right) \xrightarrow[N \rightarrow \infty]{d} \chi_v^2.$$

Designaremos de completo o modelo ajustado sob nenhuma restrição em  $\Omega$  e de reduzido o modelo ajustado sob as restrições definidas em  $H_0$ .

Para o teste de  $H_0$  a estatística do teste da razão de verossimilhança é:

$$\chi_{Calculado}^2 = -N \ln \left( \hat{\sigma}_{\Omega}^2 / \hat{\sigma}_{\omega}^2 \right) = -N \ln \left( \frac{SQRR_{\Omega}}{SQRR_{\omega}} \right).$$

A regra de decisão consiste em rejeitar  $H_0$  a um nível de significância  $\alpha$  se  $\chi_{Calculado}^2 \geq \chi_{\alpha}^2(v)$ , em que  $v = p_{\Omega} - p_{\omega}$  é o número de graus de liberdade, sendo  $p_{\Omega}$  e  $p_{\omega}$  o número de parâmetros estimados nos modelos completo e reduzido, respectivamente, e  $SQRR$  é a soma de quadrados do resíduo da regressão. Investigamos o

---

\* SAS versão 8.2 licenciado para a UFV, 2003.

ajuste de modelos para dados com repetições, portanto a  $SQRR$  envolve o erro puro mais a falta de ajustamento.

Para o teste de  $H_0^{(1)}$ ,  $H_0^{(2)}$  e  $H_0^{(3)}$  as estatísticas do teste da razão de verossimilhança são:

$$-N \ln \left( \hat{\sigma}_\Omega^2 / \hat{\sigma}_{\omega_i}^2 \right) = -N \ln \left( \frac{SQRR_\Omega / N}{SQRR_{\omega_i} / N} \right) = -N \ln \left( \frac{SQRR_\Omega}{SQRR_{\omega_i}} \right)$$

com, respectivamente,  $(g-1)$ ,  $(g-1)$  e  $2(g-1)$  graus de liberdade para  $i = 1, 2, 3$ . A seguir ilustraremos a metodologia com o auxílio do procedimento PROC NLIN do pacote estatístico SAS. Dadas as estimativas iniciais  $\theta_0$  e a forma da função  $\mu(\theta, x)$ , o PROC NLIN ajusta o modelo  $y = \mu(\theta, x) + \varepsilon$  para os dados  $(y_i, x_i)$ ,  $i = 1, \dots, N$  e fornece as estimativas de mínimos quadrados de  $\theta$ , por métodos iterativos, tais como Newton-Raphson, Gauss-Newton, Steepest Descent etc. Uma introdução aos métodos iterativos aplicada a mínimos quadrados em modelos não-lineares é apresentada em Johnson e Frasier (1985).

## 2.2 Ilustração do método

Para ilustrar a metodologia apresentada neste estudo, utilizaram-se os dados da Tabela 3, em que  $g = 2$ . Inicialmente, aplicou-se um teste de homogeneidade de variâncias (Tabela 4) e resultou que a hipótese de homogeneidade não foi rejeitada ( $P > 0,05$ ).

Com base na Tabela 2, aplicou-se o teste para falta de ajustamento do modelo para cada grupo obtendo-se os seguintes resultados:

Grupo 1:

$$F = \frac{QMFA}{QMR} = \frac{124,4750}{116,2500} = 1,07, \text{ com 4 e 6 graus de liberdade } (P=0,4472).$$

Grupo 2:

$$F = \frac{QMFA}{QMR} = \frac{115,65}{79,40} = 1,46, \text{ com 4 e 5 graus de liberdade } (P=0,3391).$$

Em ambos os grupos considerou-se o resultado não-significativo ( $P > 0,05$ ) indicando que o modelo não-linear adotado pode ser considerado adequado e obteve-se o resíduo da regressão conforme a Tabela 4.

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

$$y_{ij} = D_1 \left[ \frac{a_1 x_{ij}}{b_1 + x_{ij}} \right] + D_2 \left[ \frac{a_2 x_{ij}}{b_2 + x_{ij}} \right] + \varepsilon_{ij}$$

em que  $i = 1, 2$ ;  $j = 1, \dots, n_i$  com  $n_1 = 12$  e  $n_2 = 11$ , portanto  $N = 23$ .

Tabela 3 - Velocidade da reação *versus* concentração de substrato de um experimento com Puromycin<sup>1</sup>

Concentração de substrato (ppm)	Velocidade (contagem/min <sup>2</sup> )	
	Grupo 1	Grupo 2
0,02	76	67
	47	51
0,06	97	84
	107	86
0,11	123	98
	139	115
0,22	159	131
	152	124
0,56	191	144
	201	158
1,10	207	160
	200	-

<sup>1</sup> Fonte: Bates e Watts (1988). Grupo 1: Enzima tratada com Puromycin; Grupo 2: Enzima não tratada.

Tabela 4 - Resumo da análise da variância da regressão e do teste de homogeneidade de variâncias

Causas de Variação	Grupo 1		Grupo 2	
	GL	SQ	GL	SQ
Parâmetros da regr.	2	270213,6	2	147348,4
Resíduo da regressão	10	1195,4	9	859,6
$H_0 : \sigma_1^2 = \sigma_2^2$	$\chi_{Obs.}^2 = 0,113$ n.s. $\chi_{Tab.}^2 (5\%; 1) = 3,841$			

n.s. Não-significativo (P>0,05), pelo teste de Bartlett

O objetivo deste exemplo é ajustar o modelo Michaelis-Menten e determinar se uma única equação para os dois grupos pode ser considerada adequada e, caso contrário, se certos valores dos parâmetros no modelo podem ser considerados como iguais nos dois grupos.

Adiante no texto, fornecemos uma alternativa simples de codificação em SAS para calcular as estimativas de máxima verossimilhança dos parâmetros sob  $\Omega$  (nenhuma restrição no espaço paramétrico),  $\omega_1$  (espaço paramétrico restrito por:  $b_1 = b_2 = b$ ),  $\omega_2$  (espaço paramétrico restrito por:  $a_1 = a_2 = a$ ) e  $\omega_3$  (espaço paramétrico restrito por  $b_1 = b_2 = b$  e  $a_1 = a_2 = a$ ). Em cada caso, o PROC NLIN requer os valores iniciais das estimativas dos parâmetros com a opção PARMS.

Executando-se o procedimento NLIN com nenhuma restrição e com várias restrições especificadas por  $\omega_1, \omega_2$  e  $\omega_3$ , obtém-se as estimativas  $\hat{\sigma}_\Omega^2, \hat{\sigma}_{\omega_1}^2, \hat{\sigma}_{\omega_2}^2$  e  $\hat{\sigma}_{\omega_3}^2$  necessárias para se executar o teste estatístico das várias hipóteses descritas anteriormente.

### 2.3 Programa SAS

```
title ' Função Michaelis-Menten y = ax/(b+x) ';
data batesex1;
input grupo trat rep x y d1 d2;
cards;
1 1 1 0.02 76 1 0
1 1 2 0.02 47 1 0
1 2 1 0.06 97 1 0
1 2 2 0.06 107 1 0
1 3 1 0.11 123 1 0
1 3 2 0.11 139 1 0
1 4 1 0.22 159 1 0
1 4 2 0.22 152 1 0
1 5 1 0.56 191 1 0
1 5 2 0.56 201 1 0
1 6 1 1.10 207 1 0
1 6 2 1.10 200 1 0
2 1 1 0.02 67 0 1
2 1 2 0.02 51 0 1
2 2 1 0.06 84 0 1
2 2 2 0.06 86 0 1
2 3 1 0.11 98 0 1
2 3 2 0.11 115 0 1
2 4 1 0.22 131 0 1
2 4 2 0.22 124 0 1
2 5 1 0.56 144 0 1
2 5 2 0.56 158 0 1
2 6 1 1.10 160 0 1
;
/* programa para o modelo completo omega (Ω) */
/* Estimativas iniciais foram tomadas para cada grupo separadamente */
proc nlin data=batesex1 maxiter=100;
parms a1=205 b1=0.08 a2=165 b2=0.06;
model y=d1*((a1*x)/(b1+x)) + d2*((a2*x)/(b2+x));
run;

/* programa para o modelo com restrição ω1:b1=b2=b */
/* Estimativa inicial de b foi tomada como a média de b1 e b2 */
proc nlin data=batesex1 maxiter=100;
parms a1=205 b=0.07 a2=165;
model y=d1*((a1*x)/(b+x)) + d2*((a2*x)/(b+x));
run;

/* programa para o modelo com restrição ω2:a1=a2=a */
/* Estimativa inicial de a foi tomada como a média de a1 e a2 */
proc nlin data=batesex1 maxiter=100;
```



```

parms a=185 b1=0.08 b2=0.06;
model y=d1*((a*x)/(b1+x)) + d2*((a*x)/(b2+x)) ;
run;

/* programa para o modelo com restrição ω3:a1=a2=a e b1=b2=b */
/* Estimativa inicial de a e b foi tomada como médias de ai e bi,
respectivamente */
proc nlin data=batesex1 maxiter=100;
parms a=185 b=0.07;
model y=d1*((a*x)/(b+x)) + d2*((a*x)/(b+x)) ;
run;
quit;

```

Na Tabela 5 estão apresentadas as estimativas dos parâmetros do modelo com nenhuma restrição no espaço paramétrico ( $\Omega$ ) e com as restrições especificadas por  $\omega_1, \omega_2$  e  $\omega_3$ .

Tabela 5 - Estimativas dos parâmetros do modelo completo ( $\Omega$ ) e modelos reduzidos ( $\omega_1, \omega_2$  e  $\omega_3$ ) e respectivas somas de quadrados do resíduo da regressão ( $SQRR$ ) e número de graus de liberdade do resíduo da regressão ( $GLRR$ )

Parâmetros	Estimativas dos parâmetros dos modelos			
	$\Omega$	$\omega_1$	$\omega_2$	$\omega_3$
$a_1$	212,7	208,6	-	-
$b_1$	0,0641	-	0,0480	-
$a_2$	160,3	166,6	-	-
$b_2$	0,0477	-	0,0887	-
$a$	-	-	194,1	190,8
$b$	-	0,0580	-	0,0604
$SQRR$	2055,1	2240,9	4815,4	7276,5
$GLRR$	19	20	20	21

Na Tabela 6 estão apresentados os resultados do teste para as três hipóteses formuladas. Como o teste de  $H_0^{(3)}$  foi significativo, pode-se concluir que as duas equações diferem, estatisticamente. Assim, a equação comum, cujas estimativas estão apresentadas na Tabela 5 (modelo  $\omega_3$ ), não pode ser usada para os dois grupos. Entretanto, a hipótese de que o valor do parâmetro  $b$  é igual no modelo dos dois grupos não foi rejeitada ( $P > 0,05$ ). Assim, o modelo escolhido é aquele com três parâmetros ( $a_1, a_2$  e  $b$ ).

Um teste para falta de ajustamento para o modelo escolhido está apresentado na Tabela 7. Note que,  $SQ_{\text{Falta de ajustamento}} = SQ_{\text{Resíduo da regressão}} - SQ_{\text{Erro puro}}$ . Uma vez que o teste para falta de ajustamento foi não-significativo ( $P > 0,05$ ), pode-se concluir que o modelo é adequado para descrever os dados. As equações ajustadas são:

$$\hat{y} = d_1 \frac{208,6x}{0,0580+x} + d_2 \frac{166,6x}{0,0580+x}$$

Tabela 6 - Hipóteses avaliadas, valores da estatística do teste qui-quadrado, número de graus de liberdade e nível descritivo do teste (Valor-P)

Hipóteses	$\chi^2_{Calc.}$	Graus de liberdade (v)	$P(\chi^2_v > \chi^2_{Calc.})$
$H_0^{(1)} : b_1 = b_2 = b$	1,991	1	0,1582
$H_0^{(2)} : a_1 = a_2 = a$	19,584	1	$9,6272 \times 10^{-6}$
$H_0^{(3)} : a_1 = a_2 = a \text{ e } b_1 = b_2 = b$	29,079	2	$4,8481 \times 10^{-7}$

Tabela 7 - Análise com o teste para falta de ajustamento para o modelo escolhido

Causas de Variação	G.L.	SQ	QM	F	Valor-P
Falta de ajustamento	9	1146,40	127,38	1,28	0,3441
Resíduo (Erro puro)	11	1094,50	99,50		
Resíduo da regressão	20	2240,90			

$$\text{Grupo 1 } (d_1=1 \text{ e } d_2=0) \Rightarrow \hat{y} = \frac{208,6x}{0,0580+x}$$

$$\text{Grupo 2 } (d_1=0 \text{ e } d_2=1) \Rightarrow \hat{y} = \frac{166,6x}{0,0580+x}$$

O resultado do teste da razão de verossimilhança, com aproximação dada pela estatística  $F$  (Bates e Watts, 1988), para as três hipóteses sobre igualdade de parâmetros foram:  $F(H_0^{(1)}) = 1,72$  com 1 e 19 graus de liberdade ( $P = 0,2053$ ),  $F(H_0^{(2)}) = 25,52$  com 1 e 19 graus de liberdade ( $P = 70,8 \times 10^{-6}$ ) e  $F(H_0^{(3)}) = 24,14$  com 2 e 19 graus de liberdade ( $P = 60,708 \times 10^{-7}$ ). Assim, comparando-se os resultados obtidos com base na estatística  $\chi^2$  apresentada neste trabalho e a estatística  $F$  apresentada por Bates e Watts (1988), com base nos valores-P (Tabela 8), as conclusões são as mesmas. Entretanto, os níveis descritivos dos testes com base na estatística  $F$  foram sempre superiores aos obtidos com a estatística  $\chi^2$ .

Tabela 8 - Níveis descritivos para os testes do exemplo

Hipóteses	$P(\chi^2_v > \chi^2_{Calculado})$	$P(F > F_{Calculado})^a$
$H_0^{(1)}$	0,1582	0,2053
$H_0^{(2)}$	$9,6272 \times 10^{-6}$	$70,8 \times 10^{-6}$
$H_0^{(3)}$	$4,8481 \times 10^{-7}$	$60,708 \times 10^{-7}$

<sup>a</sup> Bates e Watts (1988)

### 3 Simulações

Com o objetivo de comparar as duas aproximações, isto é, as estatísticas  $F$  e  $\chi^2$  por simulação de dados, estabeleceu-se trabalhar com a seguinte hipótese de identidade de modelos:  $H_0: b_1 = b_2 (=b)$  e  $a_1 = a_2 (=a)$  vs.  $H_a$ : pelo menos uma igualdade é uma desigualdade.

$M$  conjuntos de dados foram simulados sob  $H_0$  para o modelo

$$y_i = \frac{a x_i}{b + x_i} + \varepsilon_i.$$

Os valores  $y_i$  foram simulados adotando-se como parâmetros do modelo os valores  $a = 185$  e  $b = 0,07$ . Os seguintes valores e critérios foram estabelecidos:  $M = 1.000$  conjunto de dados para cada valor  $r$ ;  $g = 2$  (somente dois grupos);  $r = 2, 5, 10$  e  $50$  (4 simulações com dados balanceados) e os seis valores  $x$  conforme especificados no exemplo descrito na Tabela 3. Para  $\varepsilon_i$ , o erro aleatório com as pressuposições  $\varepsilon_i \sim NID(0, \sigma^2)$  adotou-se  $\sigma^2 = 15^2 = 225$ . Portanto obteve-se  $N$  igual a 24, 60, 120 e 600 respectivamente, para  $r = 2, 5, 10$  e  $50$ . O valor  $M = 1.000$  foi estabelecido por tentativas. Testou-se inicialmente  $M = 100, 200$  e  $500$  e verificou-se uma estabilização razoável dos resultados para  $M = 1.000$  ou superior. Obteve-se, para cada cenário de simulação e para cada uma das estatísticas, a porcentagem de decisão incorreta (erro tipo I)  $\hat{\alpha} = c/M$ , em que  $c$  é o número de vezes que a estatística ( $F$  ou  $\chi^2$ ) levou à rejeição da hipótese de identidade dos modelos, ao nível de significância nominal de 5%.

Ajustou-se o modelo (1), completo e reduzido, para cada conjunto de dados simulados, interativamente com o procedimento PROC NLIN do SAS (1990).

Um resumo dos resultados do estudo por simulação de dados estão apresentados na Tabela 9.

Tabela 9 - Valores da porcentagem de decisão incorreta (erro tipo I), ao nível de significância nominal de 5%, para os  $M=1.000$  conjuntos de dados simulados, em cada uma das quatro simulações

Estatísticas	Simulações*			
	$r = 2$	$r = 5$	$r = 10$	$r = 50$
	$n = 12$ $N = 24$	$n = 30$ $N = 60$	$n = 60$ $N = 120$	$n = 300$ $N = 600$
$F$	6,1	5,6	5,2	5,2
$\chi^2$	10,0	6,7	5,9	5,5

\*  $r$  é o número de repetições para cada um dos seis valores de  $x$  em cada um dos dois grupos,  $n$  é o número de observações de cada grupo e  $N$  é o número total de observações.

Em todos os casos avaliados, a aproximação  $\chi^2$  foi a que proporcionou maiores porcentagens de erros tipo I. Para  $N=24$  a taxa de erro tipo I com a aproximação  $\chi^2$  foi exatamente duas vezes o nível de significância nominal, enquanto na aproximação  $F$  essa taxa foi de apenas 1,1% superior ao nível nominal 5%. De acordo com os resultados

apresentados, pode-se concluir que para um número total de observações suficientemente grande ( $N \geq 120$ ) as duas aproximações são praticamente equivalentes. Para amostras menores, a aproximação dada pela estatística  $F$  deve ser a preferida, uma vez que a taxa de erro tipo I foi sempre menor, independentemente do valor de  $N$ .

### Conclusões

- (i) A hipótese de identidade de modelos de regressão não-linear pode ser testada pelo teste da razão de verossimilhança utilizando-se as aproximações  $F$  ou qui-quadrado;
- (ii) Para um número total de observações suficientemente grande, as duas aproximações são praticamente equivalentes;
- (iii) Para amostras menores, a aproximação dada pela estatística  $F$  deve ser a preferida, uma vez que a taxa de erro tipo I é sempre menor, independentemente do número total de observações  $N$ .

### Agradecimentos

Os autores são gratos à Joy Smith – SAS consultant, Department of Statistics, North Carolina State University, pela valiosa colaboração com os programas utilizados nas simulações.

REGAZZI, A. J.; SILVA, C. H. O. Test for parameters equality in nonlinear regression models. I. Data in the randomized complete design. *Rev. Mat. Est.*, São Paulo, v.22, n.3, p.33-45, 2004.

- **ABSTRACT:** *We consider the adjustment of  $g$  nonlinear regression equations and the likelihood ratio test with  $\chi^2$  approximation in order to test model identity, and we also compare the approximations given by  $F$  and  $\chi^2$  test statistics in a simulation study. For sufficiently large data sets ( $N \geq 120$ ) we found the two approximations to perform roughly the same. For smaller sample sizes the  $F$  statistic approximation should be preferred since it yielded smaller type I error rates for  $N$  values investigated.*
- **KEYWORDS:** *Growth curve; likelihood; hypothesis test.*

### Referências

- BATES, D. M.; WATTS, D. G. *Nonlinear regression analysis and its applications*. New York: John Wiley, 1988. 365p.
- 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.
- DRAPER, N. R.; SMITH, H. *Applied regression analysis*. 2<sup>nd</sup>. ed. New York: John Wiley, 1981. 709p.
- GALLANT, A. R. *Nonlinear statistical models*. New York: John Wiley, 1987. 611p.
- GRAYBILL, F.A. *Theory and application of the linear model*. Belmont: Duxbury Press, 1976. 704p.
- JOHNSON, M. L.; FRASIER, S. G. Nonlinear least-square analysis. *Methods Enzymol.*, San Diego, v.117, p.301-342, 1985.

- KHATTREE, R.; NAIK, D. N. *Applied multivariate statistical with SAS software*. 2<sup>nd</sup>. ed. Cary: SAS Institute, 1999. 338p.
- LI, J. C. R. *Statistical inference*. Ann Arbor: Edwards Brothers, 1964. v.1, 658p.
- MYERS, R. H. *Classical and modern regression with applications*. 2<sup>nd</sup>. ed. Boston: PWS-Kent, 1990. 488p.
- NETER, J. et al. *Applied linear statistical models*. 4<sup>th</sup>. ed. Chicago: Richard D. Irwin, 1996. 1408 p.
- PIMENTEL GOMES, F. *Curso de estatística experimental*. 13. ed. Piracicaba: Nobel, 1990. 468p.
- RAO, C. R. *Linear statistical inference and its applications*. New York: John Wiley, 1973. 522p.
- RATKOWSKY, D. A. *Nonlinear regression modeling: a unified practical approach*. New York : Marcel Dekker, 1983. 276p.
- REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. *Rev. Ceres*, Viçosa, v.40, n.228, p.176-195, 1993.
- REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de parâmetros no caso de dados de delineamentos experimentais. *Rev. Ceres*, Viçosa, v.46, n.266, p.383-409, 1999.
- REGAZZI, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear. *Rev. Ceres*, Viçosa, v.50, n.287, p.9-26, 2003.
- SAS INSTITUTE. *SAS/STAT User's guide*. Version 6. 4<sup>th</sup>. ed. Cary, 1990. v.2, 796p.
- SEARLE, S. R. *Linear models*. New York: John Wiley, 1971. 532p.
- STEEL, R. G. D.; TORRIE, J. H. *Principles and procedures of statistics*. New York: McGraw-Hill Book, 1980. 633p.
- SOUZA, G. S. *Introdução aos modelos de regressão linear e não-linear*. Brasília: Serviço de Produção de Informação, 1998. 505p.

Recebido em 09.02.2004.

Aprovado após revisão em 19.12.2004.