

Ajuste de curvas de crescimento usando NLGAMLSS

Aline Minniti de Campos¹, Marinho Gomes de Andrade Filho¹

¹ Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, Caixa Postal 668, 13560-970, São Carlos - SP

Resumo: Neste trabalho, utilizamos o modelo aditivo generalizado não linear para locação, escala e forma (NLGAMLSS) para dados de peso-idade de bovinos da raça Canchim, do nascimento até 40 meses de idade. Supomos que a variável resposta (peso) assume as seguintes distribuições: normal, t-Student, normal generalizada, skew normal e skew t. O parâmetro de locação foi modelado como função não linear da idade. Foram utilizados, para tais funções, os modelos de crescimento de Brody, Gompertz, logístico e von Bertalanffy. A qualidade do ajuste dos modelos foi avaliada usando-se os critérios AIC e BIC. Concluímos que a distribuição skew t com o modelo de Brody resultou no ajuste para descrever o crescimento dos bovinos. O método bootstrap foi utilizado para obter intervalos de confiança mais precisos para os parâmetros.

Palavras-chave: NLGAMLSS, curvas de crescimento, bootstrap.

1 Introdução

Os modelos aditivos generalizados para locação, escala e forma (GAMLSS) foram introduzidos por Rigby *et al.* (2005) e Akantziliotou *et al.* (2002) como uma maneira de superar algumas limitações associadas com os modelos lineares generalizados (GLM) e os modelos aditivos generalizados (GAM).

No GAMLSS a suposição da distribuição da família exponencial para a variável resposta é substituída por uma família de distribuições geral, incluindo distribuições com assimetria e/ou curtose.

A parte sistemática do modelo é estendida para permitir a modelação não só da média (ou locação), mas também dos outros parâmetros da distribuição de variável resposta como funções paramétricas lineares e/ou não paramétricas aditivas da variáveis explicativas. A estimação de máxima verossimilhança é usada para ajustar os modelos.

Neste trabalho, consideramos distribuições para modelar assimetria, como as distribuições skew normal e skew t. A distribuição skew normal é uma extensão da distribuição normal (Gaussiana), considerando a presença de assimetria. Similarmente, a distribuição skew t permite regular assimetria e curtose de um modelo ajustado. Essas distribuições são obtidas introduzindo o parâmetro de assimetria nas distribuições normal e t-Student. A distribuição normal generalizada, destacada aqui, é utilizada na modelação quando a concentração de valores em torno da média e o comportamento das caudas são de interesse.

¹Agradecimento à CAPES pelo apoio financeiro. Contato: alinemc@icmc.usp.br, marinho@icmc.usp.br

O ajuste de funções de crescimento peso-idade é uma ferramenta fundamental no planejamento da produção animal, pois permite obter informações precisas sobre o padrão de crescimento dos animais estudados e permite fazer prognósticos futuros para animais do mesmo grupo racial sob a mesma situação ambiental (ver Freitas (2007)). Na literatura, são propostos vários modelos não lineares para descrever curvas de crescimento em várias espécies de animais, e dentre estes, os mais citados, segundo Fitzhugh Jr. (1976), são: Brody, von Bertalanffy, logístico e Gompertz.

Neste trabalho consideramos a extensão dos modelos GAMLSS para funções não lineares que denotamos aqui por NLGAMLSS.

2 Metodologia

Um GAMLSS assume observações independentes y_i para $i = 1, 2, \dots, n$ com função densidade de probabilidade $f(y_i|\boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor de parâmetros que está relacionado às variáveis explicativas. Denotaremos esse vetor de parâmetros como (μ, σ, ν, τ) . Os dois primeiros parâmetros μ e σ são usualmente caracterizados como parâmetros de locação e escala, enquanto os parâmetros restantes, se existirem, são caracterizados como parâmetros de forma.

A seguir, apresentaremos as distribuições utilizadas que supomos para a variável resposta (peso), bem como os modelos de crescimento não lineares usados para modelar o parâmetro de locação μ , denotado na seção 2.2 por $\mu(t, a, b, k, m)$.

2.1 Distribuições

2.1.1 Distribuição normal

A função densidade de probabilidade da distribuição normal, denotada por $NO(\mu, \sigma)$, é dada por

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right]$$

para $-\infty < y < \infty$, em que $-\infty < \mu < \infty$ e $\sigma > 0$.

2.1.2 Distribuição t-Student

A distribuição da família t é adequada para modelar dados com curtose positiva. A função densidade de probabilidade da distribuição t-Student, denotada aqui por $TF(\mu, \sigma, \nu)$, é dada por

$$f_Y(y|\mu, \sigma, \nu) = \frac{1}{\sigma B\left(\frac{1}{2}, \frac{\nu}{2}\right) \nu^{1/2}} \left[1 + \frac{(y - \mu)^2}{2\sigma^2\nu}\right]^{-\frac{\nu+1}{2}}$$

para $-\infty < y < \infty$, em que $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$ e $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ é a função Beta.

2.1.3 Distribuição normal generalizada

A distribuição normal generalizada, denotada por $NG(\mu, \sigma, \nu)$, é adequada para dados com curtose positiva e negativa. Sua função densidade de probabilidade é dada por

$$f_Y(y|\mu, \sigma, \nu) = \frac{\nu \exp[-|z|^\nu]}{2\sigma\Gamma(\frac{1}{\nu})}$$

para $-\infty < y < \infty$, em que $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$ e $z = (y - \mu)/\sigma$.

2.1.4 Distribuição skew normal

A função densidade de probabilidade da distribuição skew normal, denotada por $SN(\mu, \sigma, \nu)$, é dada por

$$f_Y(y|\mu, \sigma, \nu) = \frac{2}{\sigma} f_{Z_1}(z) F_{Z_1}(\nu z)$$

para $-\infty < y < \infty$, em que $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $z = (y - \mu)/\sigma$ e f_{Z_1} e F_{Z_1} são a função densidade de probabilidade e a função de distribuição acumulada de $Z_1 \sim NG(0, \sqrt{2}, 2)$. Esta distribuição foi introduzida por Azzalini (1986).

2.1.5 Distribuição skew t

A função densidade de probabilidade da distribuição skew t, denotada por $ST(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{2}{\sigma} f_{Z_1}(z) F_{Z_1}(\nu z)$$

para $-\infty < y < \infty$, em que $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$, $z = (y - \mu)/\sigma$ e f_{Z_1} e F_{Z_1} são a função densidade de probabilidade e a função de distribuição acumulada de $Z_1 \sim TF(0, 1, \tau)$, uma distribuição t com $\tau > 0$ graus de liberdade, com τ tratado como parâmetro contínuo. Esta distribuição foi proposta por Azzalini (1986).

2.2 Modelos de crescimento não lineares

As funções de crescimento consideradas neste trabalho para o parâmetro de locação $\mu(t, a, c, k, m)$ do NLGAMLSS são apresentadas na Tabela 1.

Tabela 1: Modelos de crescimento não lineares.

Modelos	$y(t)$	$\mu(t, a, c, k, m)$
Brody	$a(1 - ce^{-kt})$	$\ln(a) + \ln(1 - ce^{-kt})$
Gompertz	$ae^{-ce^{(-kt)}}$	$\ln(a) - ce^{(-kt)}$
Logístico	$a(1 + e^{-kt})^{-m}$	$\ln(a) - m \ln(1 + e^{-kt})$
von Bertalanffy	$a(1 - ce^{-kt})^3$	$\ln(a) + 3 \ln(1 - ce^{-kt})$

em que $y(t)$ é o peso do animal (em kg), $\mu(t, a, c, k, m) = E(\ln y(t))$ e t é a idade em meses. O parâmetro a representa o peso assintótico ou peso do animal adulto, c é uma constante que está relacionada ao peso inicial do animal, indicando a proporção do peso assintótico a

ser ganho depois do nascimento. O parâmetro k é a taxa de crescimento do peso, indica a velocidade com que o animal se aproxima da idade adulta, representa o índice e maturidade do animal e determina a eficiência do crescimento. O parâmetro m define a forma da curva nos modelos logísticos e determina o ponto em que o animal passa a fase de crescimento inibitório, determinando o ponto em que ele cresce com menor eficiência.

Vários critérios para escolha do modelo de crescimento que melhor se ajusta a um conjunto de dados podem ser adotados. Neste trabalho, consideramos o critério de Akaike (AIC) e o critério de informação bayesiano (BIC).

As estimativas dos parâmetros e os critérios de seleção citados acima foram calculados usando a função `nlglmss` do software estatístico R Development Core Team (2009), sendo que o desvio padrão e o intervalo de confiança foram calculados utilizando o método bootstrap paramétrico, devido à incoerência de valores gerados pela matriz de covariâncias desta função.

3 Resultados e discussão

Foram utilizados dados provenientes de um experimento de avaliação de sistema de cruzamento entre raças bovinas de corte executado na Embrapa Pecuária Sudeste, São Carlos-SP. Os dados correspondem a 40 medidas de peso de bovinos machos da raça Canchim, sendo que a pesagem foi realizada mensalmente durante 40 meses iniciando-se a pesagem no nascimento do animal. O objetivo é comparar os modelos de crescimento utilizando o NLGAMLSS considerando as distribuições citadas na seção anterior para a variável resposta (peso).

Os valores dos critérios de seleção calculados para os modelos analisados, são apresentados na Tabela 2.

Tabela 2: Critérios de seleção AIC e BIC para os modelos considerados.

Distribuições	Modelos							
	Brody		Gompertz		Logístico		von Bertalanffy	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Normal	-314.5	-298.5	-251.5	-235.4	-221.7	-205.6	-285.9	-269.9
t-Student	-325.9	-305.8	-270.3	-250.2	-243.4	-223.4	-303.4	-283.3
Normal generalizada	-330.1	-310.1	-286.6	-266.5	-262.9	-242.8	-316.8	-296.8
Skew normal	-276.2	-256.1	-248.2	-228.1	-209.5	-189.4	-292.2	-272.2
Skew t	-393.6	-369.5	-283.3	-259.2	-253.9	-229.8	-323.1	-299.1

Ambos critérios determinaram a distribuição skew t com o modelo de Brody como o melhor modelo, para representar a relação peso \times idade da amostra considerada.

A Tabela 3 a seguir apresenta as estimas dos parâmetros, o desvio padrão e o intervalo de confiança (com nível de confiança de 95%) do modelo selecionado.

Tabela 3: Sumário do modelo ajustado.

Parâmetros	a	c	k	σ	ν	τ
Estimavas	1204.37	0.96	0.02	0.21	-3.47	7.79
Desvio Padrão	70.96	0.002	0.002	0.13	0.85	0.6
IC (95%)	(1062.45, 1346.28)	(0.953, 0.962)	(0.016, 0.024)	(0.16, 0.27)	(-5.16, -1.77)	(2.35, 25.88)

Como ilustração, apresentamos na Figura 1 o gráfico dos percentis, em que verificamos que o modelo se ajustou bem aos dados utilizados.

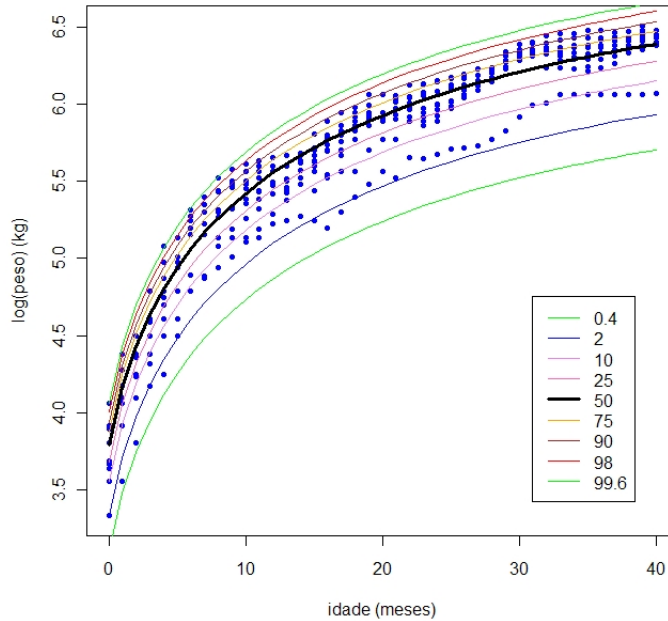


Figura 1: Gráfico dos percentis.

A Figura 2 apresenta os gráficos dos resíduos \times valor ajustado, resíduos \times índice (variável explicativa), a densidade estimada dos resíduos e a Normal Q-Q Plot.

Com estes gráficos, verificamos a homocedasticidade, a não dependência e a suposição de normalidade dos resíduos.

4 Conclusão

Os resultados obtidos neste trabalho nos leva a concluir que o modelo não linear de Brody, juntamente com a suposição da distribuição skew t para a variável resposta, foi mais versátil que os demais modelos analisados para ajustar os dados de crescimento de bovinos macho da raça Canchim até os 40 meses de idade.

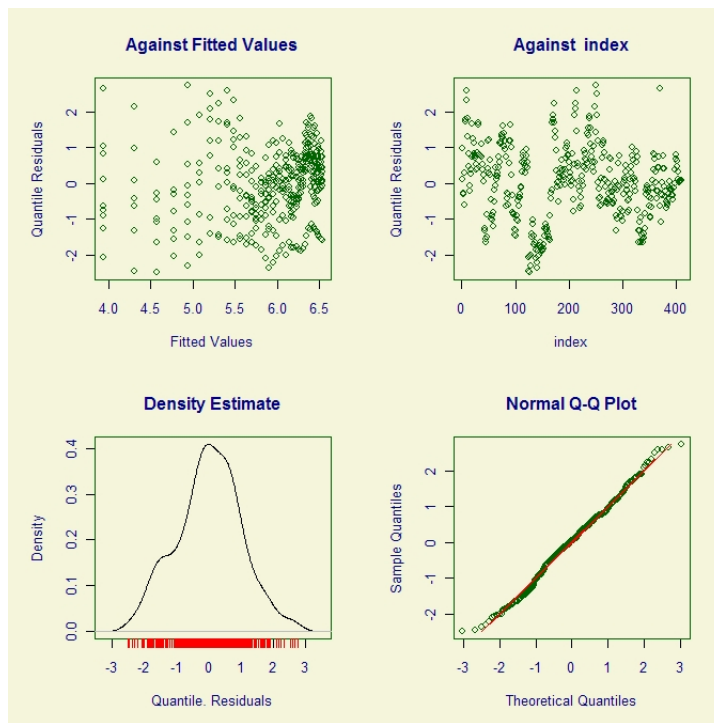


Figura 2: Gráfico dos quantis residuais.

Referências

- Akantziliotou, K., Rigby, R. A. & Stasinopoulos, D. M. (2002). The R implementation of generalized additive models for location, scale and shape. *Statistical modelling in Society: Proceedings of the 17th International Workshop on statistical modelling*, page 75–83.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46**(3), 199–208.
- Fitzhugh Jr., H. A. (1976). Analysis of growth curves and strategies for altering their shape. *Journal of animal science*, **42**(4), 1036–1051.
- Freitas, A. R. (2007). Estimativas de curvas de crescimento na produção animal. *Embrapa Pecuária Sudeste*, **Documento 68**, 1–30.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rigby, R. A., Stasinopoulos, D. M. & Lane, P. W. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, **54**(3), 507–554.