

Modelagem Bayesiana para curvas de crescimentos de codornas assumindo assimetria nos erros

Bayesian modeling growth curves for quail assuming skewness in errors

Robson Marcelo Rossi^{1*}; Lucimary Afonso dos Santos²

Resumo

Assumir distribuições como a normal nas análises de dados é comum em diferentes áreas do conhecimento. Entretanto, pode-se fazer uso de outras que possuem capacidade de modelar também o parâmetro de assimetria, para as situações em que são necessários modelar dados com caudas mais pesadas que a normal. Este trabalho pretende apresentar alternativas à suposição de normalidade nos erros, dispondo também de distribuições assimétricas. Propõe-se uma abordagem Bayesiana para ajuste de modelos não-lineares quando os erros não são normais. Assim, adotam-se as distribuições t , skew-normal e skew- t . A metodologia visa aplicação em diferentes curvas de crescimento para dados de pesos de codornas. Verificou-se que os modelos de Gompertz com erros skew-normal e skew- t , respectivamente, para machos e fêmeas, são os que melhor se ajustam aos dados.

Palavras-chave: Distribuições com erros assimétricos, inferência Bayesiana, MCMC, modelos de crescimento

Abstract

Bayesian modeling growth curves for quail assuming skewness in errors - To assume normal distributions in the data analysis is common in different areas of the knowledge. However we can make use of the other distributions that are capable to model the skewness parameter in the situations that is needed to model data with tails heavier than the normal. This article intend to present alternatives to the assumption of the normality in the errors, adding asymmetric distributions. A Bayesian approach is proposed to fit nonlinear models when the errors are not normal, thus, the distributions t , skew-normal and skew- t are adopted. The methodology is intended to apply to different growth curves to the quail body weights. It was found that the Gompertz model assuming skew-normal errors and skew- t errors, respectively for male and female, were the best fitted to the data.

Key words: Skewness errors distribution, bayesian inference, MCMC, growth models

¹ Prof. Dr. Adjunto do Dept^o de Estatística, Universidade Estadual de Maringá, UEM, Maringá, PR. E-mail: rmrossi@uem.br

² Prof^a Dr^a Adjunto do colegiado de Matemática, Universidade Estadual do Paraná, UNESPAR, Paranavaí, PR. E-mail: lucimaryafonso@hotmail.com

* Autor para correspondência

Introdução

Modelos de regressão não-lineares com erros normais são usualmente aplicados em muitas áreas de estudo, porém em muitas situações, devido à falta de simetria, a suposição de normalidade não se aplica. Segundo Campos (2011), propostas são discutidas na literatura para que seja substituída a suposição de normalidade por uma classe de distribuições que seja mais flexível.

A distribuição skew normal foi introduzida formalmente por Azzalini (1985), que estudou as propriedades desta distribuição e verificou que a mesma apresentava problemas na estimação do parâmetro que controla a assimetria. O estudo foi realizado por meio do método dos momentos e máxima verossimilhança (CAMPOS, 2011).

Sahu, Dey e Branco (2003) desenvolveram uma nova classe de distribuições por meio da introdução de assimetria em distribuições multivariadas elipticamente simétricas. A classe foi obtida usando condicionamento e transformação, contendo muitas famílias padrão, incluindo a skew-normal multivariada e skew- t . Eles obtiveram formas analíticas das densidades e estudaram suas propriedades. Ainda, deram exemplos de aplicação prática em modelos de regressão Bayesiana. Segundo esses autores, essa nova classe de distribuições é muito geral, bastante flexível e amplamente aplicável, sendo, uma alternativa às distribuições simétricas, frequentemente utilizadas em regressão. Constataram não ser fácil trabalhar com suas densidades, porém os modelos podem facilmente ser ajustados usando MCMC.

Utilizando modelo aditivo generalizado não-linear para locação, escala e forma (NLGAMLSS), Campos e Andrade Filho (2010) supuseram que a variável resposta em estudo assumia as distribuições: normal, t -Student, normal generalizada, skew normal e skew- t e ao considerarem alguns modelos de crescimento concluíram que a distribuição skew- t com o modelo de Brody foi o que apresentou melhor ajuste para

dados de peso-idade de bonivos da raça Canchim.

Os modelos não-lineares de crescimento logístico, Gompertz e de Von Bertalanffy assumindo erros skew-normais foram estudados, sob enfoque Bayesiano, por Cancho, Lachos e Ortega (2005). Os autores concluem que o método representa uma alternativa importante à análise frequentista, ao analisarem dados de quantidade de água presente em células de raiz de feijão, obtida em diferentes pontos de raízes descritos em Ratkowsky (1983).

De la Cruz e Branco (2009) modelaram dados, tomados ao longo do tempo, de um grupo de mulheres grávidas, considerando aproximação Bayesiana para modelos de regressão não-linear, substituindo a distribuição normal nos termos do erro por alguma distribuição assimétrica, que leva em consideração apenas a assimetria ou assimetria e caudas pesadas. Concluíram que o melhor modelo adotado para os erros foi o que considerou a distribuição skew- t .

Cancho, Lachos e Ortega (2010) discutiram aspectos inferenciais de modelos de regressão não-lineares com erros skew-normais por meio de aproximação frequentista e Bayesiana como extensão de modelo de regressão não-linear normal usual. Usaram a distribuição introduzida por Sahu, Dey e Branco (2003), a qual julgaram ser mais atrativa do que a proposta por Azzalini (1985) por apresentar menor grau de dificuldade na estimativa dos parâmetros de assimetria.

A importância dos modelos de crescimento utilizada na modelagem de dados de crescimento animal está em geral direcionada a auxiliar na maximização dos sistemas de produção. A busca de uma metodologia que ao mesmo tempo seja flexível, robusta e com precisão nas suas estimativas nos leva a procedimentos elegantes como, por exemplo, os Bayesianos devido a grandes vantagens amplamente divulgadas na literatura (ROSSI, 2011).

Campos e Andrade Filho (2010) e Campos (2011) propuseram uma abordagem Bayesiana, para ajustar modelos não-lineares quando os erros são não normais e/ou apresentam heteroscedasticidade,

adotando distribuições skew-normal e skew- t para situações que envolvem modelagem de dados com caudas mais pesadas os mais leves que a normal. Constataram que os modelos com erros seguindo distribuição skew-normal resultaram em melhor ajuste se comparados aos modelos simétricos.

Mazucheli, Souza e Philippsen (2011) utilizaram dados de peso-idade em codornas provenientes de observações longitudinais para conduzir uma análise de regressão considerando o modelo de Gompertz sob suposições de homoscedasticidade e heteroscedasticidade multiplicativa das variâncias. Obtiveram as estimativas dos parâmetros dos modelos por meio de maximização do logaritmo da função de log-verossimilhança via método MCMC (Cadeias de Markov em processos de Monte Carlo), sob o enfoque Bayesiano. O grupo constatou que sob suposição de homogeneidade de variâncias parece superestimar os parâmetros o verdadeiro peso médio assintótico das aves, o que não ocorre sob a suposição contrária, isto é, um modelo com erros heterocedásticos.

Este trabalho teve como objetivo avaliar o ajuste de diferentes modelos não-lineares a dados de peso corporal de codornas assumindo diferentes distribuições para o erro sob o ponto de vista Bayesiano.

Material e Métodos

O experimento foi realizado no setor de Coturnicultura da Fazenda Experimental de Iguatemi, pertencente à Universidade Estadual de Maringá - PR - Brasil, no período de agosto de 2006 a janeiro de 2007, com a utilização de 1.831 codornas de corte (*Coturnix coturnix japonica*) e fazem parte da dissertação de Paiva (2007).

O plantel foi constituído de 903 fêmeas e 928 machos. Os animais foram pesados semanalmente, formando um banco de dados para o peso corporal (em gramas) ao nascimento, 7, 14, 21, 28, e 35 dias de idade.

Utilizou-se o procedimento Bayesiano com objetivo de se comparar os pesos médios entre gêneros, em cada idade observada. Para tal, foi considerado que a resposta (Y_{ij}) segue distribuição de Normal, isto é, $y_{ij} | \mu_j, \sigma_j^2 \sim N(\mu_j, \sigma_j^2)$, $i = 1, 2, \dots, n_j$ para o j -gênero, $j = 1, 2$. Para cada μ_j e σ_j^2 foram consideradas, *a priori*, distribuições normal e gama não-informativas, respectivamente, $\mu_j \sim N(0, 10^{-6})$ e $\frac{1}{\sigma_j^2} \sim \text{Gamma}(10^{-3}, 10^{-3})$ (parametrização OpenBugs e/ou WinBugs (SPIEGELHALTER et al., 2003); ROSSI, 2011). Foram tomadas as diferenças entre distribuições *a posteriori* das médias dos respectivos gêneros e consideraram-se como diferentes, em nível de 5% de significância, os gêneros cujos intervalos de credibilidade para as diferenças médias não contemplaram o valor zero. Como valores iniciais para os parâmetros, foram tomados as médias das amostras e a obtenção das distribuições marginais *a posteriori* para todos os parâmetros foi por meio do pacote *BRugs* do programa R (R DEVELOPMENT CORE TEAM, 2012). Para tal, foram gerados 11.000 de valores em um processo MCMC, considerando um período de descarte amostral de 1.000 valores iniciais, assim a amostra final contém 10.000 valores gerados. A convergência das cadeias de todos os parâmetros analisados foi confirmada por meio do pacote CODA do programa R, pelos critérios de Heidelberger e Welch (1983) e de Geweke (1992).

Procedimento amostral: Em um cálculo preliminar, supondo normalidade nos dados, foi determinado um tamanho amostral suficientemente grande e igual a 100 aves para cada gênero, considerando um nível de significância de 5%, uma variância amostral piloto de 144 g² (o desvio-padrão médio, considerando sexo e idade foi de 12 g) e um erro máximo de estimação de 2,2 g. Segundo Magalhães (2006), o Teorema do Limite Central postula que amostras suficientemente grandes, mesmo para populações não normais, a média das amostras segue, assintoticamente, a distribuição normal. Sendo assim, foram tomados

10 estratos amostrais baseados nos decis e, por meio de amostragem aleatória simples, respectivamente, em cada estrato populacional, a amostra final foi formada. Tal procedimento foi utilizado com o propósito de minimizar o tempo computacional nas obtenções das estimativas dos parâmetros dos modelos envolvidos. Cuidados, como testes de aderência de Kolmogorov-Smirnov (MARSAGLIA; TSANG; WANG, 2003) e verificação de medidas de posição, dispersão e assimetria nas amostras com valores próximos às populacionais, foram adotados para que houvesse representabilidade da amostra, em nível de 5% de significância.

Modelagem: Existem diferentes modelos e parametrizações descritos na literatura para ajustar curvas de crescimentos. Em especial os modelos aqui apresentados possibilitam uma explicação biológica para seus parâmetros. Uma descrição mais detalhada pode ser encontrada em Ratkowsky (1983), Amaral (2008), Campos (2011) e Freitas (2005).

Em geral, assumir que os erros seguem distribuição normal pode ser inadequado, mesmo que a característica da variável de interesse seja contínua como, por exemplo, o peso corporal de aves. Há situações em que a variável de interesse apresenta formas assimétricas à direita ou à esquerda ou valores extremos em ambas. Para tais situações, são sugestivas distribuições alternativas à normal, que podem proporcionar um melhor ajuste aos dados como, por exemplo, a distribuição *t*, a skew-

normal e a skew-*t* (AZZALINI, 1985; SAHU; DEY; BRANCO, 2003; DE LA CRUZ; BRANCO, 2009; CANCHO; LACHOS; ORTEGA, 2010).

Desta forma, optou-se para o presente estudo um procedimento de análise em duas etapas: a primeira consistiu em considerar alguns modelos não-lineares mais utilizados para descrever curvas de crescimento de aves em diferentes formas parametrizadas (Equações 1 até 7), assumindo-se distribuição normal para os erros e, selecionar aquele que melhor se ajusta aos dados por meio da inferência Bayesiana. Para tal utilizou-se o critério DIC (*Deviance Information Criterion*; SPIEGELHALTER et al., 2002), como principal, e as magnitudes do parâmetro σ . O DIC pode ser utilizado tanto para a comparação, quanto para seleção de (co)variáveis em modelos. Segundo os autores, quanto menor o DIC, melhor é o ajuste e, sugerem utilizar o seguinte critério para o módulo da diferença entre os valores de DIC de dois modelos, A e B analisados, isto é: $D = |DIC_A - DIC_B|$. Assim, se $D < 5$, se conclui que a diferença não é significativa, se $5 \leq D \leq 10$, que a diferença é significativa e se $D > 10$, que a diferença é altamente significativa.

Na segunda etapa, distribuições alternativas à normal também foram consideradas para o erro, a saber, a distribuição *t*, a skew-normal e skew-*t* seguindo os procedimentos descritos por De la Cruz e Branco (2009) e a skew-normal por Cancho, Lachos e Ortega (2010).

Curvas de crescimento:

Brody (1945): $f(x) = \beta_1 (1 - \beta_2 e^{-\beta_3 x})$ [1]

Gompertz (1825): $f_1(x) = \beta_1 e^{-\beta_2 e^{-\beta_3 x}}$ [2]

$f_2(x) = \beta_1 e^{-e^{(\beta_2 - \beta_3 x)}}$ [3]

Logística (Verhulst, 1838): $f_1(x) = \frac{\beta_1}{1 + e^{(\beta_2 - \beta_3 x)}}$ [4]

$f_2(x) = \frac{\beta_1}{1 + \beta_2 e^{-\beta_3 x}}$ [5]

$f_3(x) = \frac{\beta_1}{1 + e^{\left(\frac{\beta_2 - x}{\beta_3}\right)}}$ [6]

Von Bertalanffy (1957): $f(x) = \beta_1 (1 - \beta_2 e^{-\beta_3 x})^3$ [7]

Segundo Campos (2011), para os modelos de crescimento aqui apresentados, o parâmetro β_1 representa o valor assintótico da variável resposta; β_2 é uma constante que está relacionada ao valor observado inicial, indicando a proporção do valor assintótico a ser ganho; β_3 é a taxa de crescimento da variável resposta que indica a velocidade com que o valor observado se aproxima do máximo observado, isto é, determina a eficiência do crescimento.

Modelagem para o erro: A metodologia Bayesiana, por meio de simulações de Cadeias de Markov em processos de Monte Carlo (MCMC) permite com o uso das distribuições *a priori* não

$$y_{ij} = f(x_{ij}, \boldsymbol{\beta}) + \varepsilon_{ij}; i = 1, 2, \dots, m \text{ e } j = 1, 2, \dots, n_i \quad [8]$$

em que $f(\cdot)$ é uma função não-linear que depende da covariável x_i e do vetor de parâmetros, $\boldsymbol{\beta}$, no modelo investigado (curva de crescimento) tal que ε_{ij} é o erro associado ao modelo, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, no i -ésimo animal na j -ésima ocasião de observação. Para a implementação via MCMC, a seguinte estrutura hierárquica foi utilizada nos modelos: $y_{ij} | x_{ij}, \boldsymbol{\beta}, \sigma^2 \sim N(f(x_{ij}, \boldsymbol{\beta}), \sigma^2)$, com distribuições *a priori* para os parâmetros $\boldsymbol{\beta}$ e σ^2 , respectivamente, normal e gama, isto é, $\boldsymbol{\beta} \sim N(0, 10^{-6})$ e $\frac{1}{\sigma^2} \sim \text{Gamma}(10^{-3}, 10^{-3})$ (parametrização OpenBugs).

Entretanto, é sugestivo assumir distribuições de probabilidade para os erros, de tal forma a

informativas para os parâmetros, obter estimativas idênticas às frequentistas para todos os parâmetros *a posteriori* dos modelos analisados. Assumindo independência *a priori*, a distribuição *a posteriori* conjunta é dada pelo produto das distribuições *a priori* pela função de verossimilhança. Dentre as muitas vantagens do procedimento, é possível incorporar diferentes estruturas na modelagem, isto é, diferentes distribuições para os erros, heteroscedasticidade, efeitos aleatórios, estruturas de (co)variância etc. Na modelagem de curvas de crescimento por meio de modelos não-lineares é comum assumir a normalidade nos erros (resíduos), isto é:

adequar caudas mais pesadas e/ou assimétricas e pode ser mais indicado em determinadas situações como, por exemplo, quando se verifica que os resíduos proporcionados pelo ajuste, assumindo distribuição normal aos erros, se distribuem de forma assimétrica. Há também os casos em que se desconsideram efeitos fixos e/ou aleatórios no modelo de crescimento, visto que o resíduo é inflado por tais desconsiderações.

De la Cruz e Branco (2009) e Cancho, Lachos e Ortega (2010) apresentam formas distintas de obtenção dos parâmetros em modelos de regressão não-lineares Bayesianos assumindo assimetria nos erros. A seguir uma breve descrição de ambos:

Procedimento 1 (P₁) – De la Cruz e Branco (2009) [9]

Para o caso de independência nos erros, os autores apresentam uma modelagem por meio da distribuição skew-*t* e skew-normal, como um caso especial da primeira, isto é, $\varepsilon_{ij} \sim ST(-\mu_{\varepsilon_{ij}}, \sigma^2, \lambda, \nu)$, tal que $\mu_{\varepsilon_{ij}} = \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \lambda$, onde λ é o onde λ é o parâmetro de assimetria. Desta forma segue que: $E(y_{ij}) = f(x_{ij}, \boldsymbol{\beta})$,

$$\text{Var}(y_{ij}) = \left(\frac{\nu}{\nu-2}\right)(\sigma^2 + \lambda^2) - \frac{\nu}{\pi} \left[\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right]^2 \lambda^2$$

$\nu > 2$. Observe que se $\lambda=0$, os erros assumem a distribuição *t*-Student e para $\nu \rightarrow \infty$, a distribuição normal.

Para a implementação via MCMC, usam a seguinte estrutura hierárquica nos modelos:

$y_{ij} | x_{ij}, \beta, \sigma^2, \lambda, \nu, \delta_{ij}, z_{ij} \sim N\left(f(x_{ij}, \beta) - \mu_{\varepsilon_{ij}} + \lambda z_{ij}, \delta_{ij}^{-1} \sigma^2\right)$, sendo $Z_{ij} \sim N(0, \delta_{ij}^{-1}) I_{(z_{ij} > 0)}$ e $\delta_{ij} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, e assumindo $\beta \sim t(0, 10^{-6}, 2)$, $\sigma \sim U(0, 10^2)$, $\lambda \sim t(0, 10^{-2}, 2)$ e $\nu \sim \text{Exp}(10^{-1}) I_{(2.5, \infty)}$ como distribuições *a priori*. Para $\delta_{ij} = 1$, e excluindo a distribuição *a priori* para ν , os erros assumem

Procedimento 2 (P₂) – Cancho, Lachos e Ortega (2010)

[10]

Os autores apresentam uma modelagem por meio da distribuição skew-normal para os erros, isto é, $\varepsilon_{ij} \sim SN\left(-\mu_{\varepsilon_{ij}}, \sigma^2, \lambda\right)$, tal que $E(y_{ij}) = f(x_{ij}, \beta)$ e $\text{Var}(y_{ij}) = \sigma^2 + \left(1 - \frac{2}{\pi}\right) \lambda^2$. Desta forma tem-se que $y_{ij} | x_{ij}, \beta, \sigma^2, \lambda \sim SN\left(f(x_{ij}, \beta) - \mu_{\varepsilon_{ij}}, \sigma^2, \lambda\right)$, em que $\mu_{\varepsilon_{ij}} = \sqrt{\frac{2}{\pi}} \lambda$.

Para a implementação via MCMC, os autores usam a seguinte estrutura hierárquica nos modelos reparametrizando a distribuição skew-normal por meio da representação estocástica de Henze (1986), dada por $Z_{ij} = f(x_{ij}, \beta) - \mu_{\varepsilon_{ij}} + \sigma \left(\delta |T_0| + (1 - \delta^2)^{1/2} T_1\right)$, com $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$, T_0 e T_1 variáveis independentes com distribuição normal padrão. Desta forma, tem-se que $E(Z) = f(x_{ij}, \beta) - \mu_{\varepsilon_{ij}} + \sqrt{\frac{2}{\pi}} \sigma \delta$ e $\text{Var}(Z) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right)$. Finalmente

$y_{ij} | T_i = t_i, x_{ij}, \hat{\alpha}, \sigma^2, \lambda, z_{ij} \sim N\left(f(x_{ij}, \beta) + \sigma \delta t_i, \sigma^2 (1 - \delta^2)\right)$, assumindo que $T_i \sim N(0, 1) I_{(0, \infty)}$ assim como $\beta \sim N(0, 10^{-6})$, $\lambda \sim N(0, 10^{-3})$ $\frac{1}{\sigma^2} \sim \text{Gamma}(10^{-3}, 10^{-3})$ como distribuições *a priori*.

Como valores iniciais para os parâmetros, foram tomados valores próximos às estimativas frequentistas e consideraram-se significativos *a posteriori*, em nível de 5%. A obtenção das distribuições marginais *a posteriori* para todos os parâmetros foi por meio do pacote *BRugs* do programa R (R DEVELOPMENT CORE TEAM,

distribuição skew-normal. Os autores concluem tanto em um estudo simulado quanto em uma aplicação a dados reais, que ao utilizar uma curva de crescimento do tipo logística apresentada na Equação 7, o modelo skew-*t* para os erros é o mais adequado por apresentar maior probabilidade de cobertura dos intervalos de credibilidade para os verdadeiros valores dos parâmetros e menor DIC.

2012). Foram gerados 1.010.000 de valores em um processo MCMC considerando um período de descarte amostral de 10.000 valores iniciais, assim a amostra final tomada em saltos de tamanho igual a 100, contém 10.000 valores gerados. A convergência das cadeias de todos os parâmetros analisados foi confirmada por meio do pacote CODA do programa R, pelos critérios de Heidelberger e Welch (1983) e de Geweke (1992).

Para comparar as curvas de crescimento entre gêneros, foram tomadas as diferenças entre as distribuições *a posteriori* das médias dos respectivos parâmetros de interesse. Consideraram-se como diferentes, em nível de 5% de significância, os parâmetros cujos intervalos de credibilidade para as diferenças médias não contemplaram o valor zero.

Resultados e Discussão

É sabido na literatura da produção animal, que de modo geral, há diferenças nas observações quando os dados são analisados separadamente por gênero. A Tabela 1 apresenta estimativas Bayesianas (média, desvio-padrão e intervalo com 95% de credibilidade) da variável peso corporal (em gramas) ao nascimento, 7, 14, 21, 28 e 35 dias de idade das aves analisadas. Apesar de a evolução do peso apresentar comportamento parecido, os valores são distintos, principalmente a partir dos 21 dias de idade, já que foram detectadas diferenças significativas nas médias *a posteriori*, o que nos

induz a ajustar modelos para as curva de crescimento separadamente.

No contexto do melhoramento genético, Paiva (2007) comenta que as correlações genéticas obtidas

entre os pesos nas diferentes idades analisadas indicam que a progressos genéticos podem ser obtidos com sucesso, para peso ao abate, usando-se como critério de seleção o peso aos 28 dias de idades.

Tabela 1. Estimativas Bayesianas (média e desvio-padrão (dp) *a posteriori*) para peso corporal (g) de codornas, por sexo, nas idades (dias) consideradas e seus intervalos com 95% de credibilidade.

Idade (dias)	Pesos (g)			
	Machos		Fêmeas	
	Média (dp)	$P_{2.5\%} - P_{97.5\%}$	Média (dp)	$P_{2.5\%} - P_{97.5\%}$
1	9,50 ^a (0,031)	9,45 – 9,57	9,58 ^a (0,033)	9,52 – 9,65
7	29,66 ^a (0,164)	29,34 – 29,98	29,81 ^a (0,169)	29,47 – 30,14
14	76,56 ^a (0,314)	75,95 – 77,18	76,88 ^a (0,365)	76,17 – 77,60
21	139,00 ^b (0,472)	138,10 – 139,90	145,50 ^a (0,592)	140,40 – 142,70
28	192,30 ^b (0,553)	191,20 – 193,40	197,50 ^a (0,656)	196,20 – 198,80
35	225,70 ^b (0,639)	224,40 – 226,90	235,90 ^a (0,724)	234,50 – 237,40

dp = desvio-padrão da média; $P_{2.5\%} - P_{97.5\%}$ = intervalo com 95% de credibilidade;

^{a, b} Letras distintas na linha, indicam diferenças significativas entre as médias dos pesos (g), por meio de comparações Bayesianas em nível de 95% de credibilidade.

Fonte: Elaboração dos autores.

Na primeira etapa de análise, verificou-se por meio do menor DIC que o modelo de Gompertz, assumindo erros normais, é o que melhor se ajusta aos dados, para ambos os gêneros (Tabela 2). Optou-se pela forma parametrizada f_2 do modelo não-linear de Gompertz devido à velocidade de processamento computacional. Mazucheli, Souza e Philippsen (2011) trabalharam com a mesma forma paramétrica assumindo também erros heteroscedásticos na forma $\varepsilon_{ij} \sim N(0, \sigma_e^2 x_j^\rho)$ sugerida por Bellio, Jensen e Seiden (2000), em dados de ganho de peso de codornas. Estimaram os parâmetros por máxima verossimilhança e por MCMC e concluem que a última modelagem é a mais adequada a situação além de flexível e preciso. Ressaltam que apesar de não apresentados no artigo, os modelos Richards e Von Bertalanffy superestimaram, de maneira significativa, o peso médio assintótico das aves.

Tabela 2. DIC para os modelos considerados – erros normais.

Modelo	DIC	
	Machos	Fêmeas
<i>Brody</i>	5.019	4.969
<i>Gompertz - f₁</i>	4.858	4.816
<i>Gompertz - f₂</i>	4.858	4.816
<i>Logístico - f₁</i>	4.860	4.830
<i>Logístico - f₂</i>	4.860	4.830
<i>Logístico - f₃</i>	4.913	4.830
<i>Von Bertalanffy</i>	4.887	4.846

Fonte: Elaboração dos autores.

Na segunda etapa de análise, para os dados de peso corporal de aves macho verificou-se, que o modelo de Gompertz- f_2 , assumindo erros skew-normais via procedimento P_2 , de Cancho, Lachos e Ortega (2010), apresentou o menor DIC (4.555)

(Tabela 3). Entretanto, observou-se que a estimativa para $\sigma = 19,6$ foi a mais alta entre todos os modelos considerados. Este fato nos leva a escolher P_1 como melhor procedimento por ser mais parcimonioso (DIC = 4.563, $\sigma = 8,16$) considerando-se erros

skew-normais, conforme sugerido por De la Cruz e Branco (2009). Desta forma, ao assumir erros skew-normais e skew- t , respectivamente, para machos e fêmeas, obteve-se os melhores ajustes para os dados de pesos corporais de codornas (Figura 1).

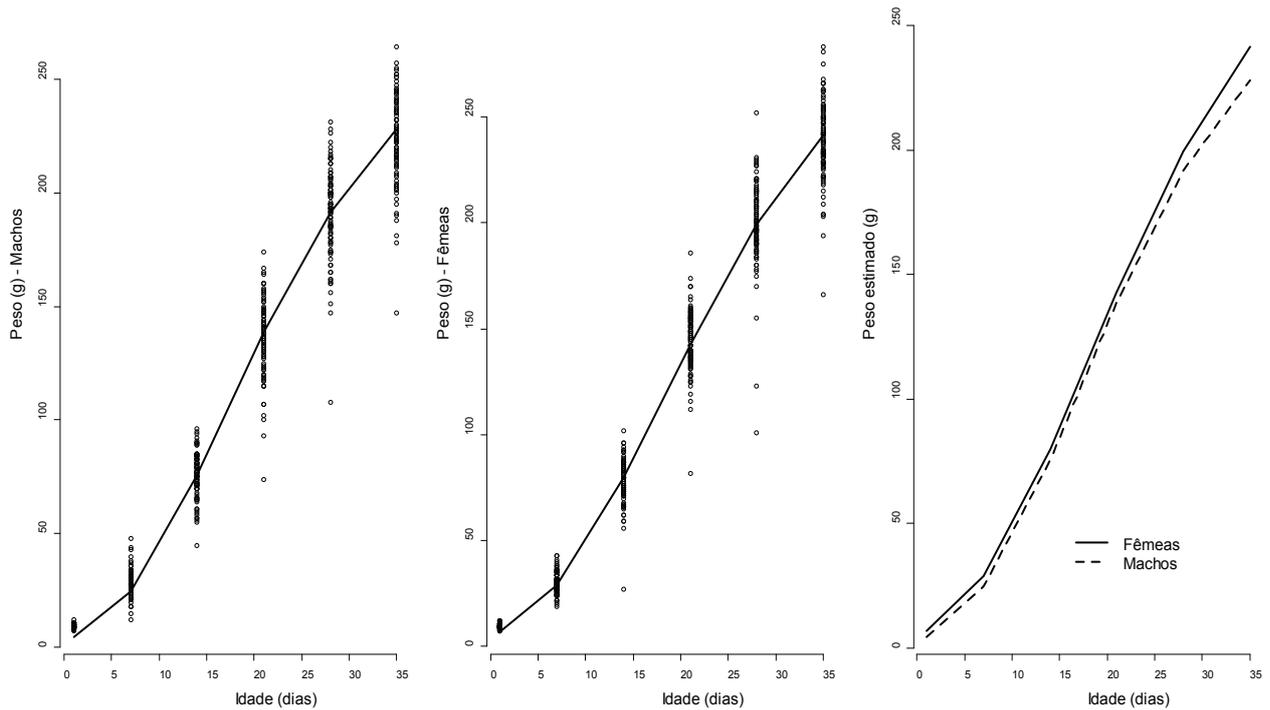
Tabela 3. Estimativas Bayesianas (média e desvio-padrão (dp) *a posteriori*) para os parâmetros da curva Gompertz - f_2 em diferentes distribuições para os erros, seus respectivos DIC's e intervalos com 95% de credibilidade.

Distribuição para o erro		Machos		Fêmeas	
		Média (dp)	$P_{2,5\%} - P_{97,5\%}$	Média (dp)	$P_{2,5\%} - P_{97,5\%}$
Normal	β_1	285,90 (5,69)	(275,00 ; 297,70)	310,20 (5,91)	(299,10 ; 322,00)
	β_2	1,40 (0,03)	(1,35 ; 1,45)	1,41 (0,02)	(1,37 ; 1,46)
	β_3	0,08 (0,003)	(0,076 ; 0,086)	0,079 (0,002)	(0,075 ; 0,084)
	σ	13,83 (0,40)	(13,07 ; 14,65)	13,36 (0,39)	(12,64 ; 14,15)
	DIC	4.858		4.816	
t -Student	β_1	292,90 (6,58)	(280,30 ; 305,20)	319,60 (7,68)	(308,90 ; 331,40)
	β_2	1,38 (0,02)	(1,35 ; 1,42)	1,38 (0,01)	(1,36 ; 1,41)
	β_3	0,08 (0,003)	(0,075 ; 0,084)	0,076 (0,002)	(0,073 ; 0,079)
	ν	2,28 (0,23)	(2,01 ; 2,86)	2,42 (0,29)	(0,02 ; 3,11)
	σ	7,55 (0,47)	(6,74 ; 8,54)	7,19 (0,43)	(6,40 ; 8,09)
DIC	4.720		4.621		
Skew-normal- P_1	β_1	278,80 (5,32)	(267,60 ; 289,00)	308,80 (28,45)	(295,20 ; 324,20)
	β_2	1,51 (0,04)	(1,45 ; 1,59)	1,48 (0,04)	(1,41 ; 1,53)
	β_3	0,09 (0,003)	(0,083 ; 0,096)	0,083 (0,005)	(0,077 ; 0,088)
	λ	-18,18 (1,19)	(-20,49 ; -15,87)	-15,97 (1,74)	(-18,21 ; -13,49)
	σ	8,16 (0,67)	(6,89 ; 9,49)	8,89 (0,95)	(7,68 ; 10,22)
DIC	4.563		4.603		
Skew- t - P_1	β_1	283,40 (5,03)	(273,80 ; 293,60)	296,61 (4,84)	(288,13 ; 307,45)
	β_2	1,46 (0,03)	(1,41 ; 1,52)	1,47 (0,03)	(1,43 ; 1,52)
	β_3	0,09 (0,003)	(0,080 ; 0,090)	0,08 (0,002)	(0,080 ; 0,089)
	λ	-5,53 (1,22)	(-7,86 ; -3,08)	-5,52 (1,22)	(-7,77 ; -3,07)
	ν	2,67 (0,17)	(2,50 ; 3,12)	2,76 (0,23)	(2,51 ; 3,35)
σ	7,12 (0,48)	(6,20 ; 8,09)	7,32 (0,42)	(6,51 ; 8,15)	
DIC	4.568		4.520		
Skew-normal- P_2	β_1	280,10 (5,51)	(270,60 ; 290,70)	298,30 (8,70)	(277,80 ; 313,20)
	β_2	1,50 (0,04)	(1,39 ; 1,57)	1,55 (0,11)	(1,29 ; 1,68)
	β_3	0,09 (0,003)	(0,080 ; 0,094)	0,090 (0,006)	(0,080 ; 0,099)
	λ	-2,12 (0,61)	(-2,83 ; -0,09)	-3,99 (1,74)	(-5,66 ; -1,56)
	σ	19,60 (1,63)	(13,93 ; 21,70)	22,54 (3,50)	(17,24 ; 28,11)
DIC	4.555		4.369		

dp = desvio-padrão da média; $P_{2,5\%} - P_{97,5\%}$ = intervalo com 95% de credibilidade.

Fonte: Elaboração dos autores.

Figura 1. Curvas de *Gompertz* - f_2 ajustadas a dados de peso (g), por meio do procedimento P_1 , com erros skew-normais e skew- t , respectivamente, para machos e fêmeas.



Fonte: Elaboração dos autores.

Analogamente, para os dados de peso corporal de aves fêmeas verificou-se que ao assumir erros skew- t por meio do procedimento P_2 , o ajuste apresentou o menor DIC (4.369) (Tabela 3). Entretanto, também se observou por esse procedimento que a estimativa para $\sigma = 22,54$ foi a mais alta entre todos os modelos considerados. Optou-se então pela modelagem que assume erros skew- t (DIC = 4.520, $\sigma = 7,32$), por meio do procedimento P_1 .

Em ambos os gêneros, foram observadas reduções em torno de 10% nas amplitudes dos intervalos de credibilidade, somente para o parâmetro β_1 , quando são contrastados modelos com erros normais e skew- t - P_1 . Isto implica em baixo impacto no uso de diferentes metodologias nas estimativas. Em contrapartida, se observam reduções nas estimativas do parâmetro σ , respectivamente, 52% e 55% ao se considerarem os mesmos modelos. Isto se deve, em parte, ao fato de o modelo skew- t , sendo ele mais flexível ao incorporar o parâmetro de assimetria λ ,

ser mais sensível aos dados, isto é, captura a realidade biológica por meio da forma da distribuição. Em grandes sistemas de produção animal, e até mesmo no melhoramento genético, este fator possibilita ganhos, em especial financeiros, já que animais são selecionados baseados em características como, por exemplo, o peso corporal ou ganho de peso ao longo do tempo.

Obviamente a escolha do melhor modelo não depende somente da estrutura dos dados, mas também da área em que foram coletados entre outros aspectos, já que a interpretação biológica para os seus parâmetros pode ser distinta.

De la Cruz e Branco (2009) modelaram dados de β -HCG, um tipo de hormônio protéico, produzido pela placenta, que aumenta progressivamente durante a gravidez. O estudo foi realizado em gestantes acompanhadas e mensuradas de 1 até 4 vezes ao longo do tempo em uma clínica de obstetria privada em Santiago, Chile. Utilizara a

curva tipo Logística- f_3 (Equação 6). Relatam que ao assumir erros correlacionados skew- t , obtiveram melhores ajustes para os dados, por meio dos critérios DIC e LPML.

Campos e Andrade Filho (2010), por meio do procedimento frequentista NLGAMLSS (Modelo generalizado não-linear) considerando os modelos não-lineares de Brody, Gompertz, Logístico e Von Bertalanffy em diferentes distribuições para o erro: normal, t , normal generalizada, skew-normal e skew- t , aplicados a dados de peso de bovinos machos da raça Canchim desde o nascimento até os 40 meses de idade, concluíram via AIC e BIC que o modelo de Brody assumindo erros skew- t foi o mais versátil.

Campos (2011) apresentou uma aplicação a dados de crescimento de clones de árvores *Eucalyptus sp.* da Votorantim Celulose e Papel na região de São Paulo, tanto no caso homoscedástico quanto no caso heteroscedástico, com e sem assimetria nos erros. Concluiu que de modo geral, considerando os mesmos modelos de Campos e Andrade Filho (2010), o modelo de Brody juntamente com a suposição da distribuição skew-normal heteroscedástico foi o que apresentou a melhor opção de modelagem utilizando o procedimento Bayesiano P_2 , sugerido por Cancho, Lachos e Ortega (2010). Para o diagnóstico do bom ajuste dos modelos considerados utilizaram os critérios Bayesianos EAIC, EBIC, CPO e DIC.

Conclusões

O modelo não-linear de Gompertz na forma parametrizada f_2 , com erros skew-normais e skew- t , respectivamente, para machos e fêmeas, por meio do procedimento proposto por De la Cruz e Branco (2009), foi o que melhor ajustou os dados de pesos corporal de codornas.

Agradecimentos

O(s) autor(es) agradece(m) a prof^a. Msc. Emília de Paiva Porto (DMV-UENP) e ao prof. Dr. Elias

Nunes Martins (DZO-UEM) pelos dados usados neste trabalho.

Referências

AMARAL, M. T. R. *Abordagem Bayesiana para curva de crescimento com restrições nos parâmetros*. 2008. Dissertação (Mestrado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, Recife.

AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, Stockholm, v. 12, n. 2, p. 171-8, 1985.

BELLIO, R.; JENSEN, J. E.; SEIDEN, P. Applications of likelihood asymptotic for nonlinear regression in herbicide bioassays. *Biometrics*, Arlington, v. 56, n. 4, p. 1204-12, 2000.

BERTALANFFY, L. V. Quantitative laws in metabolism and growth. quarterly review of biology. *Stony Brook*, New York, v. 32, n. 3, p. 217-31, 1957.

BRODY, S. *Bioenergetics and growth*. New York: Rinhold Publishing Corp, 1945. 1023 p.

CAMPOS, A. M.; ANDRADE FILHO, M. G. Ajuste de curvas de crescimento usando NLGAMLSS. In: SINAPE, 19., 2010, São Pedro. *Anais...* São Pedro: Associação Brasileira de Estatística, 2010. p. 1-6.

CAMPOS, A. M. *Uma abordagem Bayesiana para alguns modelos de crescimento na presença de assimetria e heteroscedasticidade*. 2011. Dissertação (Mestrado em Ciências Matemáticas e Computação) - Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo, São Paulo.

CANCHO, V. G.; LACHOS, V. H.; ORTEGA, E. M. M. A nonlinear regression model with skew-normal errors. *Statistical Papers*, New York, v. 51, n. 3, p. 547-558, 2010.

DE LA CRUZ, R.; BRANCO, M. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. *Biometrical Journal*, Weinheim, v. 51, n. 4, p. 588-609, 2009.

FREITAS, A. R. Curvas de crescimento na produção animal. *Revista Brasileira de Zootecnia*, Viçosa, MG, v. 34, n. 3, p. 786-95, 2005.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). *Bayesian statistics 4*. Oxford: Oxford University Press, 1992. p.

169-93.

GOMPERTZ, B. On the nature of the function expressive of the law of human mortality, and a new mode of determining the value of live contingencies. *Philosophical Transactions of the Royal Society*, London, v. 115, p. 513-85, 1825.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. *Operations Research*, Baltimore, v. 31, n. 6, p. 1109-144, 1983.

HENZE, N. A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, Ithaka, v. 13, n. 4, p. 271-275, 1986.

MAGALHÃES, M. N. *Probabilidade e variáveis aleatórias*. São Paulo: EDUSP, 2006. 428 p.

MARSAGLIA, G.; TSANG, W. W.; WANG, J. Evaluating kolmogorov's distribution. *Journal of Statistical Software*, Hong Kong, v. 8, n. 18, p. 1-4, 2003.

MAZUCHELI, J.; SOUZA, R. M.; PHILIPPSEN, A. S. Gompertz growth model in the presence of normal heteroscedastic errors: A case study. *Revista Brasileira de Biometria*, São Paulo, v. 29, n. 1, p. 91-101, 2011.

PAIVA, E. *Estimativas de parâmetros genéticos e fenotípicos, para peso corporal e características de carcaça em codornas de corte*. 2007. Dissertação (Mestrado em Zootecnia) - Programa de pós-graduação em Zootecnia, Centro de Ciências Agrárias. Universidade Estadual de Maringá, Maringá.

R DEVELOPMENT CORE TEAM. 2012. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.r-project.org>>. Acesso em: 01 jan. 2012.

RATKOWSKY, D. A. *Nonlinear regression modeling*. New York: Marcel Dekker, 1983. 276 p.

ROSSI, R. M. *Introdução aos métodos bayesianos na análise de dados zootécnicos com uso do WinBUGS e R*. Maringá: Eduem, 2011. 191 p.

SAHU, S. K.; DEY, D. K.; BRANCO, M. D. A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, Malden, v. 31, n. 2, p. 129-50, 2003.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER INDE, A. Bayesian measures of model complexity and fit (with discussion). *Journal Royal of Statistical Society*, London, v. 64, n. 4, p. 583-639, 2002.

SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; LUNN, D. *Winbugs version 1.4 user manual*. Mrc biostatistics unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine. Cambridge: MRC, 2003. Disponível em: <<http://www.mrc-bsu.cam.ac.uk>>. Acesso em: 01 jan. 2012.

VERHULST, P. F. Notice sur la loi que la population suit dans son accroissement. *Correspondence Mathematic et Physic*, Brussel, v. 10, p. 113-21, 1838.

