

Estimação do fator de condição de peixes da espécie *Tracydoras paraguayensis*: uma perspectiva bayesiana

Robson Marcelo Rossi^{1*}, Terezinha Aparecida Guedes¹, Vanderly Janeiro¹ e Elias Nunes Martins²

¹Departamento de Estatística, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brasil.

²Programa de Pós-graduação em Zootecnia, Universidade Estadual de Maringá, Maringá, Paraná, Brasil. *Autor para correspondência. E-mail: rrossi@uem.br

RESUMO. O fator de condição de peixes, expressa a relação peso-comprimento e é aceita como um indicador da avaliação do “bem-estar” do animal; quanto maior o peso, melhor deve ser sua condição de sobrevivência. Este trabalho objetivou ajustar um modelo Gama, da família exponencial com função de ligação potência, a dados de medidas de comprimento e peso de peixes da espécie *Tracydoras paraguayensis*, observados na bacia do Rio Paraná, Paraná-Brasil. As estimativas dos parâmetros do modelo foram obtidos por dois processos distintos: pelo método clássico da máxima verossimilhança e pelo método Bayesiano (MCMC). Observou-se que todas as estimativas Bayesianas (médias à posteriori) para os parâmetros de locação foram similares as estimativas clássicas, contudo o parâmetro de dispersão é inferior ao obtido pela estimativa clássica, além de fornecer um erro padrão 30% menor.

Palavras-chaves: fator de condição, inferência Bayesiana, modelos lineares generalizados, *Tracydoras paraguayensis*.

ABSTRACT. Estimate of the condition factor of fish of the *Tracydoras paraguayensis* species: a Bayesian perspective. The factor of fish condition expresses the weight-length relationship and it is accepted as an indicator for the evaluation of the animal "well-being"; as larger the weight, best should be your survival condition. This work aims to fit a Gamma model, from the exponential family with power link function, to data of length measurements and weight of fish from the *Tracydoras paraguayensis* species, observed in the basin of the Paraná River, Paraná-Brazil. The estimates of the parameters of the model were obtained by two different processes: by a classic method (maximum likelihood) and by Bayesian method (MCMC). It was observed that the Bayesian estimates (posterior mean) of the parameters of interest were similar to the classic estimates, the dispersion parameter, it underestimated the classic estimate, however the dispersion parameter is inferior to the obtained by the classic estimate, besides supplying a standard mistake 30% smaller.

Key words: condition factor, Bayesian inference, generalized linear models, *Tracydoras paraguayensis*.

Introdução

A crescente preocupação com a natureza ocorre devido a um consenso cada vez mais generalizado, de que a humanidade estará se conduzindo para uma situação irreversível de penúria se continuar indiscriminadamente utilizando matéria-prima, energia e lançando ao meio ambiente toda espécie de efluentes (Clemente, 1994). Embora os peixes tenham a capacidade natural de responder fisiologicamente, adaptando-se às alterações provocadas por um estresse moderado, Barton e Iwana (1991) *apud* Merighe *et al.* (2004) complementam que, no caso de estresse crônico, o animal perde a capacidade de adaptabilidade, torna-se mais susceptível às doenças e, em casos mais extremos, morre. Os efeitos metabólicos do estresse

social provavelmente são maiores em condições de cativeiro, pois a oportunidade para a fuga é limitada, influenciando o desempenho dos peixes na aquicultura (Merighe *et al.*, 2004).

A ictiofauna da bacia do rio Paraná é composta por cerca de 600 espécies (Bonetto, 1986). Dentre estas, a espécie *Tracydoras paraguayensis* (o peixe conhecido como “Armadinho”) da família Doradidae, é considerada uma espécie comum.

A Universidade Estadual de Maringá (UEM), Estado do Paraná, no ano de 1983, iniciou um projeto de estudo sobre a população de peixes do reservatório de Itaipu, na bacia do rio Paraná, com o objetivo de avaliar o impacto da construção da usina sobre a ictiofauna da região. Do projeto originou o Nupélia, Núcleo de Pesquisas em Limnologia,

Ictiologia e Aqüicultura, que é hoje referência nacional e internacional em estudos da área.

A relação peso-comprimento do peixe é aceita como um indicador da avaliação do "bem-estar" do peixe; quanto melhor esta relação, melhor deve ser sua sobrevivência (Agostinho e Gomes, 1997). Tal relação é denominada fator de condição do peixe.

Em geral a relação peso (W) e comprimento (L) padrão é ajustada pelo modelo dado por $W_t = aL_s^b$ (Agostinho e Gomes, 1997), onde a indica o estado nutricional do peixe, e b representa o tipo de incremento que pode assumir valores próprios de acordo com a espécie, onde assumindo valor igual a três, gera o modelo para o fator de condição ($K = a$) de Fulton (Connolly e Bolzer, 1989):

$$K = \frac{W}{L^3}.$$

Outro ajuste é o modelo linear com transformação logarítmica, também descrita por Connolly e Bolzer (1989):

$$\log(W_t) = \log(K) + bL_s.$$

De acordo com a estrutura dos modelos lineares generalizados, Acorsi (2002) sugere que o fator de condição estimado pode ser dado por:

$$K = \left(\hat{\mu} \right)^p - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2.$$

Segundo Gulland, Weatherley e Gill (*apud* Agostinho e Gomes, 1997) se o valor de b for próximo de três o crescimento é definido como isométrico, ou seja, as taxas de incremento das diferentes partes do corpo do peixe são iguais. A utilização do fator de condição justifica-se por possibilitar inferências acerca da condição do indivíduo em relação ao seu ambiente, por refletir a ação conjunta de fatores biótipos e abiótipos.

Weatherley (1972) indica algumas funções de K , como comparar duas ou mais populações vivendo sob estas condições de alimentação, densidade, clima etc.; determinar a duração do período de maturação gonadal; acompanhar o grau de atividade alimentar de uma espécie, verificando se ela está ou não fazendo um bom uso da fonte alimentar; determinar a evolução comportamental na reprodução (Rossi-Wongtschowski, 1977), entre outras.

Acorsi (2002) justifica sua pesquisa pelo fato de que, embora existam na literatura vários fatores de condição disponíveis, a maioria deles refere-se a

populações que não pertencem a espécies cujo habitat natural pertence a áreas com históricos de barramentos, que é a característica fundamental da população envolvida. Complementa ainda que na ausência da suposição de normalidade, modelar por meio de modelos mais parcimoniosos como os modelos lineares generalizados (MLG), capazes de relatar o peso total do animal de maneira mais simplificada e coerente, é mais adequados que através de modelos lineares simples (ML).

Dentro da metodologia por meio dos modelos lineares generalizados, dentre várias propriedades, uma pode fornecer uma medida estatística que representa a discrepância (discordância) entre o modelo ajustado e o modelo saturado, isto é, entre o modelo e os dados, denominada *Deviance* (McCullagh e Nelder, 1989).

Outra forma de inspecionar um modelo ajustado, especialmente quando o modelo ajustado é usado para análise ou predição, é realizando análise dos resíduos que evidenciará os prováveis *outliers* (atípicos: valores extremos ou discrepantes) e fornecerá várias medidas de influência para identificar observações cuja remoção causaria mudanças no modelo ajustado. Um dos procedimentos utilizados para esta verificação é por meio da Distância de Cook (Cook e Weisberg, 1982), que opera com os resíduos padronizados.

Há vários estudos sobre a utilização de distribuições *a priori* nos modelos lineares generalizados (MLG). Dellaportas e Smith (1993) consideram uma distribuição *a priori* Normal Multivariada para os parâmetros de regressão. Esta é, a distribuição mais usada, talvez por ser de mais fácil implementação. Ibrahim e Chem (1998) consideram o uso de dados históricos para construir distribuições *a priori* informativas para os parâmetros de regressão em modelos lineares generalizados mistos. Por outro lado, Ibrahim e Laud (1991) deduzem a distribuição *a priori* de Jeffreys para os Modelos Lineares Generalizados (MLG) e mostram que para a maior parte desses modelos conduz a distribuições *a posteriori* próprias. Gelman *et al.* (1995) consideram várias situações de distribuições *a priori* possíveis.

Não há relato na literatura de estudos por meio da metodologia Bayesiana, que modelam dados que expressam a relação peso-comprimento de peixes. O presente trabalho visa avaliar este procedimento no intuito de fornecer estimativas Bayesianas melhores ou tão precisas quanto às clássicas, já que esta última, utiliza-se da teoria assintótica para construção de intervalos de confiança para os parâmetros populacionais.

Os procedimentos Bayesianos em geral, produzem estimativas mais apuradas, com maior representatividade para os modelos estatísticos, em especial na determinação da relação peso-comprimento de peixes. Contudo, a distribuição *a posteriori* conjunta, na maioria das vezes, é muito complexa, sendo difícil a obtenção analítica das distribuições marginais, as quais são as distribuições de interesse na análise. Segundo Sorensen (1996), a "Marginalização" é a maior dificuldade técnica na implementação Bayesiana, sendo praticamente impossível essa obtenção analítica, mesmo por métodos usuais de integração, em algumas situações (Gianola, 1996). O problema pode ser solucionado com os processos de simulação mediante iterações que utilizam os métodos MCMC (*Markov Chain Monte Carlo*), como o algoritmo de amostrador de Gibbs (Geman e Geman, 1984), que é indicado quando as distribuições condicionais completas são identificáveis. Caso isso não ocorra, pode-se utilizar o algoritmo Metropolis-Hastings (Chib e Greenberg, 1995).

Teoricamente, o processo iterativo teria fim apenas quando o número de iterações tendesse ao infinito. Na prática admite-se convergência quando a série alcança um estado de estacionaridade, significando que as condicionais completas estão suficientemente próximas das distribuições marginais, o que pode ser verificado com a utilização de diferentes critérios, como o "factor R" de Gelman e Rubin (1992), onde R próximo de "1" é indicativo de convergência; "Stationarity" de Heidelberger e Welch (1983), que testa se a cadeia é ou não estacionária, indicando convergência, segundo a Estatística de Cramer-Von-Misses; "z-score" de Geweke (1992), é utilizado para comparação de médias, usando a estatística Z. Onde Valor $|k| < 1,96$ indicativo de convergência, entre outros, inclusive testes gráficos. Gamerman (1996) comenta que, apesar de os resultados teóricos garantirem a convergência do amostrador de Gibbs (Geman e Geman, 1984), sua utilização na prática pode ser dificultada em razão da eventual complexidade dos modelos. Essa complexidade faz com que a convergência do amostrador de Gibbs seja de difícil obtenção. A lentidão na convergência pode estar relacionada com a alta correlação entre os elementos da cadeia de um dado parâmetro.

Com o avanço da tecnologia, *softwares* são desenvolvidos para que a complexidade dos modelos não seja uma barreira na obtenção de boas estimativas para os parâmetros. Um exemplo disto é

o WinBUGS (Spiegelhalter *et al.*, 1996), um *software* desenvolvido especificamente para a teoria Bayesiana. Apesar de ser uma poderosa ferramenta de obtenção das estimativas dos parâmetros, este *software* é uma "caixa-preta" pois utiliza a metodologia que achar conveniente para a obtenção das amostras. A grande vantagem é que não há necessidade de se obter a forma analítica da distribuição *a posteriori* e muito menos a de suas marginais condicionais. As amostras são calculadas rapidamente, dependendo do modelo, e após a verificação de convergência, via CODA (Best *et al.*, 1995), podem-se obter estimativas como média, moda, mediana e desvio-padrão *a posteriori*.

Material e métodos

Os dados analisados neste trabalho referem-se as variáveis: peso total (Y em gramas), comprimento total (X₁ em centímetros) e o sexo (X₂), inicialmente de 3.117 peixes da espécie *Trachydoras paraguayensis*, coletados de forma amostral sistemática e criteriosa pelos pesquisadores do Nupélia no período de 1986 a 2001 na bacia do Rio Paraná, Estado do Paraná, Brazil.

Para assegurar os pressupostos de consistência e representatividade da amostra, os indivíduos são capturados com o uso de redes de Espera Simples com malhagem de 3 a 16 centímetros entre nós opostos, e redes Feiticeira com 7 e 8 centímetros entre nós adjacentes. A escolha do aparelho de pesca em cada reservatório (Ivinheima, Baía e Paraná) é feita com base em amostragem piloto para a estimativa de eficiência de pesca. Após a captura, as amostras são fixadas em formol 4%, acondicionadas, com a identificação do local, data, horário e método de pesca, em tambores de polietileno e conduzidas aos laboratórios de ictiologia do Nupélia onde são medidas e analisadas (Acorsi, 2002).

Os dados referentes à medição do peso e comprimento dos peixes, foram submetidos a uma análise estatística exploratória para a verificação da presença de *outliers*.

Após tal procedimento de retirada dos *outliers*, o banco de dados final reduziu-se para 2.985 peixes, o qual foi utilizado para a obtenção das estimativas dos parâmetros e a avaliação do modelo. Estes procedimentos foram realizados por meio do *software* SAS: *Statistical Analysis System* (SAS, 2003), de acordo com a metodologia dos modelos lineares generalizados, descrito anteriormente.

Modelo linear generalizado - Clássico

Seja y_i o valor da variável dependente, Y e

$X_{i1}, X_{i2}, \dots, X_{ik}$ os valores das k variáveis explicativas, X_1, \dots, X_k , para o i -ésimo elemento de um conjunto de dados.

O modelo linear generalizado tem as seguintes características:

i) Os Y_i ($i = 1, \dots, n$) são variáveis aleatórias independentes compartilhando da mesma distribuição da família exponencial;

ii) As variáveis aleatórias fornecem um conjunto de preditores lineares $\eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$ para $i = 1, \dots, n$ e k variações;

A ligação entre (i) e (ii) é tal que $g(\mu_i) = \eta_i$, onde μ_i é a média de Y_i para $i = 1, \dots, n$; $g(\cdot)$ é chamada de função de ligação do modelo.

As duas extensões para o modelo linear geral que caracterizam o modelo linear generalizado são: sua aplicabilidade para qualquer membro da família de distribuições exponenciais e, a presença de uma função ligação conectando o preditor linear η a média μ de Y .

Em qualquer problema, existem várias possíveis funções de ligação que podem ser usadas. Entretanto, uma simplificação particular é introduzida se a função de ligação é a mesma que define o parâmetro canônico para a distribuição relevante; esta função de ligação é chamada de função de ligação canônica.

Uma variável aleatória Y tem distribuição com família exponencial se a função de densidade de probabilidade $f(y)$ pode ser escrita na forma canônica usada por McCullagh e Nelder (1989):

$$f(y; \theta, \varphi) = \exp \left\{ \frac{[y\theta - b(\theta)]}{a(\varphi)} + c(y; \varphi) \right\}, \quad (1)$$

para algumas funções específicas $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ e parâmetros θ e φ .

Se φ é conhecido então a família é chamada família exponencial linear e θ é o parâmetro natural ou canônico.

Segundo Krzanowski (1998), freqüentemente a função $a(\varphi)$ é dada por $a(\varphi) = k(\varphi)$, onde k é um peso *a priori* que varia de observação para observação e $\varphi = \sigma^2$ ou $\varphi = 1$.

Acorsi (2002) relata que após análise descritiva dos dados, os mesmos sugerem como possíveis distribuições para a variável resposta, as distribuições Gama e Normal Inversa, pois segundo Cordeiro (1997), essas distribuições são associadas a dados contínuos assimétricos e positivos. Acorsi (2002)

conclui que o modelo Gama é mais adequado por apresentar menor dispersão e melhor ajuste, considerando uma função de ligação potência com $p = 0,35$ (fixo), obtido de forma subjetiva, sugerido pelo comportamento do gráfico de dispersão dos valores ajustados versus observados (Demétrio, 2001).

Considerando a variável aleatória "peso" (Y) com distribuição Gama com parâmetros φ e $\frac{\mu}{\varphi}$, respectivamente, forma e escala, dada por:

$$f(y) = \frac{1}{\Gamma(\varphi)} \left(\frac{\varphi}{\mu} \right)^\varphi y^{\varphi-1} e^{-\frac{\varphi y}{\mu}} \quad 0 \leq y < \infty. \quad (2)$$

Pode-se reescrever esta distribuição como membro da família exponencial:

$$f(y) = \exp \left\{ \left(-\frac{y}{\mu} - \log \mu \right) \varphi - \log \Gamma(\varphi) + \varphi \log \varphi + (\varphi - 1) \log y \right\} \quad (3)$$

tal que:

$$\theta = -\frac{1}{\mu},$$

$$b(\theta) = -\log(-\theta),$$

$$a(\varphi) = \frac{1}{\varphi},$$

$$c(y; \varphi) = -\log \Gamma(\varphi) + \varphi \log \varphi + (\varphi - 1) \log y.$$

Isto indica que Y é um membro da família exponencial com parâmetro canônico θ , média $E(Y) = b'(\theta) = \mu$, função de variância $b''(\theta) = \mu^2$ e, portanto, $\text{Var}(Y) = a(\varphi)b''(\theta) = \frac{\mu^2}{\varphi}$.

Para esta distribuição, a função de verossimilhança e seu logaritmo são dados, respectivamente por:

$$L(\mu_i, \varphi) = \prod_{i=1}^n \exp \left\{ \left(-\frac{y_i}{\mu_i} - \log \mu_i \right) \varphi + c(y_i; \varphi) \right\} \quad (4)$$

e

$$l(\mu_i, \varphi) = \sum_{i=1}^n \left\{ \left(-\frac{y_i}{\mu_i} - \log \mu_i \right) \varphi + c(y_i; \varphi) \right\} \quad (5)$$

com $i = 1, \dots, n$. Tomando-se a função de ligação potência, $\eta_i = \mu_i^p$, então para $\mu_i = (X\beta)^{1/p}$ implica em $\eta_i = \mu_i^p = [(X\beta)^{1/p}]^p = X\beta$. Substituindo-a em (5), tem-se:

$$l(\mu_i, \varphi) = \sum_{i=1}^n \left\{ -\frac{y_i}{(X\beta)^{1/p}} - \log(X\beta)^{1/p} \right\} \varphi + c(y_i; \varphi). \quad (6)$$

Assim, para a análise dos dados, tem-se $X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, o qual o logaritmo da função verossimilhança, agora será:

$$l(\beta_i, \varphi) = \sum_{i=1}^n \left\{ [-y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{-1/p} - \log(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{1/p}] \varphi + c(y_i; \varphi) \right\} \quad (7)$$

e considerando que $\mu_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{1/p}$, tem-se:

$$\begin{aligned} g(\mu_i) &= \eta_i = \mu_i^p = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \\ g(\mu_i)' &= p\mu_i^{p-1} = p(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{\frac{(2p-1)}{p}}, \\ a_i(\varphi) &= \frac{1}{\varphi}, \\ \text{Var}(\mu_i) &= b''(\theta_i) = \mu_i^2 = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{\frac{(2)}{p}}. \end{aligned}$$

Modelo Bayesiano

Preservando as características e pressuposições clássicas, considerou-se que a variável aleatória Y tem distribuição Gama com parâmetros positivos: escala e forma dados por, respectivamente, $\frac{\mu}{\varphi}$ e φ , com $y \geq 0$, isto é $Y_i \sim \text{Gama}\left(\frac{\mu_i}{\varphi}, \varphi\right)$, em que $E(Y_i) = \mu_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})^{1/p}$ é a função de ligação do tipo potência com $p = 0,35$ (fixo).

Distribuições a priori para os parâmetros

Para cada parâmetro: β_0, β_1 e β_2 , considerou-se como distribuição *a priori*, uma distribuição Normal não-informativa, isto é: $\beta_i \sim N(0; 10^6)$ e para φ , considerou-se uma distribuição Gama também não-informativa: $\varphi \sim \text{Gama}(1; 1)$.

Distribuição a posteriori

A distribuição *a posteriori* conjunta apresentada em (8) não tem uma forma fechada, sendo assim o *software* WinBUGS (Spiegelhalter *et al.*, 1996), foi utilizado para a obtenção das estimativas amostrais, isto é: Metropolis-Hastings (Hastings, 1970), método da Rejeição Adaptativa (ARMS) ou Envolvimento (Gilks e Wild, 1992) etc.

$$\begin{aligned} P(\varphi, \mu | Y, X_1, X_2) &= L(Y|\varphi, \mu, X_1, X_2, p)P(\varphi)P(\mu) \\ P(\varphi, \mu | Y, X_1, X_2) &= \prod_{i=1}^n \left\{ \frac{1}{\Gamma(\varphi)} \left(\frac{\varphi}{\mu_i} \right)^\varphi y_i^{\varphi-1} e^{-\frac{\varphi}{\mu_i} y_i} \right\} P(\varphi)P(\mu) \\ P(\varphi, \mu | Y, X_1, X_2) &= \prod_{i=1}^n \left\{ \frac{1}{\Gamma(\varphi)} \left(\frac{\varphi}{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})^{1/p}} \right)^\varphi y_i^{\varphi-1} e^{-\frac{\varphi y_i}{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})^{1/p}}} \right\} P(\varphi)P(\beta_0)P(\beta_1)P(\beta_2). \end{aligned} \quad (8)$$

As estimativas Bayesianas, foram obtidas utilizando como valores iniciais para $\beta_0, \beta_1, \beta_2$ e φ , respectivamente:

cadeia 1: 1,9; 0,007; 0,0017; 420;
cadeia 2: 2,0; 0,009; 0,0020; 450.

Foram gerados 100.000 valores para cada uma das cadeias, com uma “queima” ou período de descarte amostral de 4.000 valores iniciais. As amostras finais foram tomadas em saltos iguais a: 5, 20, 23 e 20, produziram amostras de tamanhos: 18.000, 4.500, 3.913 e 4.500, respectivamente, para os parâmetros $\varphi, \beta_0, \beta_1$ e β_2 .

Resultados e discussão

Na Tabela 1 são apresentados os resultados obtidos das estimativas clássicas, considerando a potência $p = 0,35$ (fixa). Pode-se verificar que os parâmetros, obtidos por meio do *software* SAS, são significativos no ajuste (p -valor $< 0,05$), indicando uma robustez do modelo.

Tabela 1. Estimativas clássicas para os parâmetros do MLG, respectivos Erros-padrão, Intervalos de Confiança, e p -valores, obtidos via *software* SAS.

Table 1. Classic estimates for the parameters of GLM, respective standard-errors, Confidence Intervals, and p -values, obtained by SAS software.

Parâmetros Parameters	Estimativas Estimates	Erros-padrão Standard-errors	Intervalo de Confiança (95%) Confidence Interval (95%)		p-valor p-values
			2,5%	97,5%	
φ	587,6350	15,2064	558,5741	618,2078	-
β_0	1,9260	0,0021	1,9219	1,9302	0,0001
β_1	0,0092	0,0001	0,0090	0,0094	0,0001
β_2	0,0017	0,0006	0,0005	0,0029	0,0060

Os valores da *Deviance* e *Deviance Escalada* do modelo oferecidos pelo *software* SAS (SAS, 2003), foram respectivamente, 5,0811 e 2.985,6, indicando um bom ajuste para o modelo, já que estão próximos de 1 e $n = 2.985$, respectivamente (McCullagh e Nelder, 1989).

O modelo possui características lineares na função da média para a variável Y, mesmo sua distribuição apresentando características não normais com variância não constante (Nelder e Wedderburn, 1972). Segundo Acorsi (2002), uma estimativa para o fator de condição do peixe, baseado nos resultados apresentados na Tabela 1, pode ser dada por:

$$K = \mu^{0,35} - 0,0092X_1 - 0,0017X_2$$

Na Tabela 2 são apresentadas as estimativas dos parâmetros obtidos via WinBUGS (Spiegelhalter et al., 1994) e na Tabela 3 os resultados dos testes de convergência das cadeias via CODA (Best et al., 1995).

As estimativas Bayesianas (média da distribuição) dos parâmetros β_0 , β_1 e β_2 , foram praticamente as mesmas das estimativas clássicas (Tabelas 2 e 1, respectivamente). Não havendo nem grandes diferenças entre os intervalos de confiança e de credibilidade para esses parâmetros, assim como em suas amplitudes.

A estimativa Bayesiana de ϕ (média da distribuição) diferiu em 166 unidades da estimativa clássica, com erro-padrão 30% menor, pois no método clássico foi de 15,20 e no Bayesiano de 10,92.

Tabela 2. Estimativas Bayesianas para os parâmetros (Médias e Medianas), respectivos Erros-padrão, Intervalos de Credibilidade (percentis 2,5 e 97,5%), obtidas via software WinBUGS.

Table 2. Bayesian estimates for the parameters (Mean and Median), respective Standard-errors, Credibility Intervals (percentile 2,5 and 97,5%), by WinBUGS.

Parâmetros	Médias	Erros-padrão	Erro de Monte Carlo	Percentil 2,5%	Medianas	Percentil 97,5%
Parameters	Means	Standard-deviation	Error MC	Percentile	Medians	Percentile
ϕ	421,6	10,92	0,0253700	400,6	421,4	443,4
β_0	1,926	0,002468	0,0000593	1,92100	1,92600	1,931
β_1	0,0092	0,000113	0,0000027	0,00898	0,00920	0,00950
β_2	0,0017	0,000715	0,0000077	0,00028	0,00169	0,00308

Além dos testes apresentados na Tabela 3, um teste gráfico para verificação de convergência das cadeias, denominado “fator Shrink” (Gelman e

Tabela 3. Critério de convergência das cadeias via CODA.

Table 3. Criterion of convergence of the chains by CODA.

Parâmetros	Cadeias	Teste*			
		Score Z	Fator R	Estacionaridade	Autocorrelação (Saltos)
Parameters	Chains	z-score	R Factor	Stationarity	Autocorrelations (Lags)
ϕ	1	0,115	1,00	Sim	-0,0028 (lag5)
	2	-1,910	1,00	Sim	0,0005 (lag5)
β_0	1	1,330	1,00	Sim	< 0,01 (lag20)
	2	-0,961	1,02	Sim	< 0,01 (lag20)
β_1	1	-1,930	1,00	Sim	< 0,01 (lag23)
	2	1,000	1,02	Sim	< 0,01 (lag23)
β_2	1	-1,450	1,00	Sim	< 0,01 (lag5)
	2	-0,833	1,02	Sim	< 0,01 (lag5)

*Os critérios adotados para os testes de convergência das cadeias geradas foram, respectivamente: Score Z (Geweke, 1992), Fator R (Gelman e Rubin, 1992), de Estacionaridade (Heidelberger e Welch, 1983) e Autocorreção em Saltos.

Rubin, 1992), foi utilizado. A Figura 1 ilustra que houve a convergência, já que a linha contínua está muito próxima a linha tracejada.

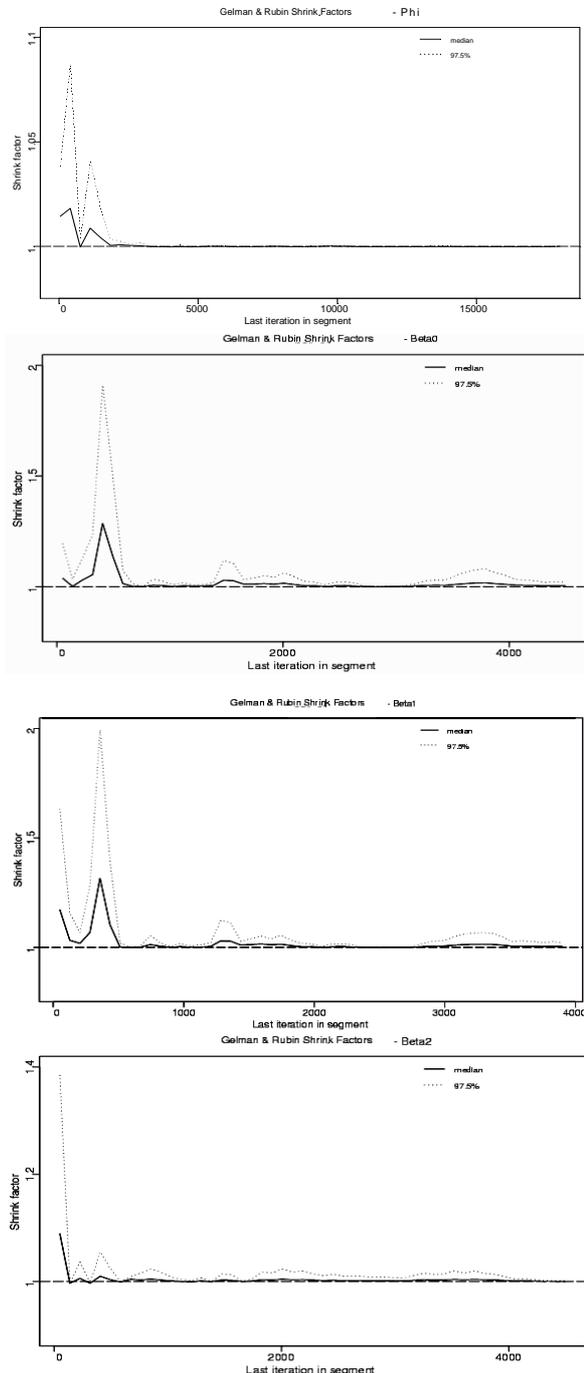


Figura 1. Gráfico do fator Shrink (Gelman e Rubin, 1992), para as cadeias dos parâmetros ϕ , β_0 , β_1 e β_2 , respectivamente.

Figure 1. Factor Shrink graphics (Gelman and Rubin, 1992), for the chains of the parameters, ϕ , β_0 , β_1 e β_2 , respectively.

Segundo Vicentin et al. (2004) o fator de condição indica o estado de higidez ou bem estar do peixe e está associada às variações dessa condição

frente ao meio em que vive o organismo. Pela Tabela 4 e Figura 2, observa-se que a grande massa (75%) de peixes tem seu fator de condição estimado entre os valores 1,667 e 1,792, sendo que a amplitude total é de 0,341, apresentando uma média de 1,692.

Tabela 4. Resumo à posteriori das estimativas do fator de condição.

Table 4. Summary the posteriori of the estimates of the condition factor.

Mínimo <i>Min.</i>	1º Quartil (P _{25%}) <i>1th quartile</i>	Mediana <i>Median</i>	Média <i>Mean</i>	3º Quartil (P _{75%}) <i>3th quartile</i>	Máximo <i>Max.</i>
1,451	1,667	1,692	1,682	1,708	1,792

Na Figura 2 é apresentada a distribuição do fator de condição e seu respectivo gráfico Box-Plot, para os peixes da espécie em estudo.

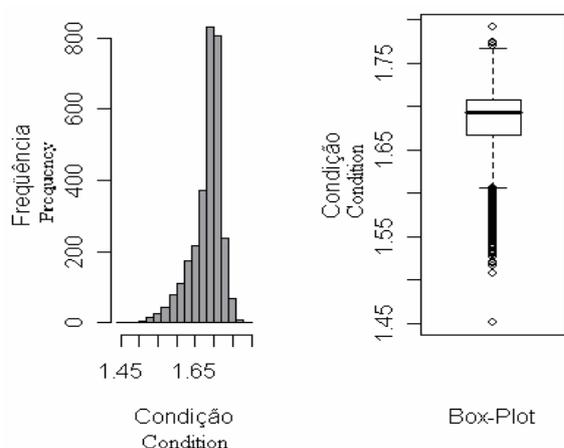


Figura 2. Distribuição de frequência das estimativas para o fator de condição.

Figure 2. Distribution of frequency of the estimates for the condition factor.

A metodologia Bayesiana mostra-se de fácil aplicação e entendimento. A estatística descritiva das distribuições marginais é facilmente obtida para cada parâmetro como, por exemplo, no cálculo de erros-padrão. É uma ferramenta de grande potencial, pois leva em conta a incerteza existente sobre todos os parâmetros do modelo, além de possibilitar a inclusão de informações passadas, pelo uso de distribuições *a priori* informativas, o que podem melhorar a acurácia dos resultados ou predições.

Conclusão

Uma vez que o objetivo deste estudo foi o de modelar dados de medidas de comprimento e peso de peixes por meio da inferência Bayesiana, conclui-se que a mesma pode ser aplicada sem restrição. Para os dados do presente estudo, as estimativas obtidas por este método são consistentes e mais precisas que as obtidas por meio da metodologia clássica.

Referências

- ACORSI, C.R.L. *Estimação do fator de condição para peixes utilizando modelos lineares generalizados*. 2002. Dissertação (Mestrado em Engenharia de Produção)–Universidade Federal de Santa Catarina, Florianópolis, 2002.
- AGOSTINHO, A.; GOMES, L.C. *Reservatório de Segredo*. Bases ecológicas para o manejo. Maringá: Eduem, 1997.
- BEST, N.G. *et al. Convergence diagnostics and output analysis software for Gibbs sampler output*. Cambridge: MRC Biostatistics Unit, 1995.
- BONETTO, A.A. The Paraná river system. In: DAVIES, B.R.; WALKER, E.D. (Ed.). *The ecology of river systems*. Dordrecht: Dr. W. Junk Publishers, 1986. p. 541-555.
- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. *Am. Stat.* Washington, D.C., v. 49, n. 4, p. 327-335, 1995.
- CLEMENTE, A. *Economia regional e urbana*. São Paulo: Editora Atlas, 1994.
- CONNOLLY, P.L.; BOLGER, T. The selection of suitable indices for the measurement and analysis of fish condition. *J. Fish Biol.*, Ireland, v. 34, n. 2, p. 171-182, 1989.
- CORDEIRO, G.M. *Introdução à teoria de verossimilhança*. 10. ed. Rio de Janeiro: Sinape, 1992.
- COOK, R.D.; WEISBERG, S. *Residuals and influence in regression*. London: Chapman and Hall, 1982.
- DELLAPORTAS, P.; SMITH, A.F.M. Bayesian inference for generalized linear and proportional hazard model via Gibbs. *Appl. Stat.*, New York, v. 42, p. 443-459, 1993.
- DEMÉTRIO, C.G.B. *Modelos lineares generalizados na experimentação agrônômica*. Piracicaba: Esalq, 2001.
- GELMAN, A. *et al. Bayesian data analysis*. 1. ed. New York: Chapman and Hall, 1995.
- GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hayward, v. 7, n. 4, p. 457-511, 1992.
- GEMAN, A.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, Canada, v. 6, p. 721-741, 1984.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In: BERNARDO, J.M. *et al.* (Ed.). *Bayesian statistics 4*. Oxford: Oxford University Press, 1992. p. 169-193.
- GIANOLA, D. *Bayesian analyses with applications to genetics e biology*. Class notes of a course tough at the department of genetics. São Paulo: Medical School, 1996.
- GILKS, W.R.; WILD, P. Adaptive rejection sampling for Gibbs sampling. *Appl. Stat.*, New York, v. 41, p. 337-348, 1992.
- HASTINGS, W.K. Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, Oxford, v. 57, p. 97-109, 1970.
- HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. *Op. Res.*, Maryland, v. 31, p. 1109-1144, 1983.

- IBRAHIM, J.G.; LAUD, P.W. On Bayesian analysis of generalized linear models using Jeffrey's prior. *J. Am. Stat. Assoc.*, Alexandria, v. 86, p. 981-986, 1991.
- IBRAHIM, J.G.; CHEN M.H. Prior elicitation and variable selection for generalized linear mixed models. In: DEY D.K. et al. (Ed.). *Generalized linear models: A Bayesian perspective*. New York: Marcel Dekker, 1998. p. 41-53.
- KRZANOWSKI, W. *An introduction to statistical modelling*. London: Arnold, 1998.
- McCULLAGH, P.; NELDER, J.A. *Generalized linear models*. 2. ed. New York: Chapman and Hall, 1989.
- McCULLOCH, C.E.; SEARLE, S.R. *Generalized, linear, and mixed models*. New York: John Wiley and Sons, 2001.
- MERIGHE, G.K. et al. Efeito da cor do ambiente sobre o estresse social em tilápias do Nilo (*Oreochromis niloticus*). *Rev. Bras. Zootec.*, Viçosa, v. 33, n. 4, p. 828-837, 2004.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized Linear Models. *J. Royal Stat. Soc. A*, London, v. 135, n. 3, p. 370-384, 1972.
- ROSSI-WONGTSCHOWSKI, C.L.D.B. Estudo das variações da relação peso total/comprimento total em função do ciclo reprodutivo e comportamento de *Sardinella brasiliensis* da costa do Brasil entre 23°S e 28°S. *Bol. Inst. Oceanog.*, São Paulo, v. 26, p. 131-180, 1977.
- SAS Institute. *Statistics analysis system*. Version 8.02, Cary, 2003.
- SORENSEN, D. *Gibbs sampling in quantitative genetics*. Foulun: [s.n.], p. 186, 1996.
- SPIEGELHALTER, D.J. et al. *BUGS - Bayesian inference using gibbs sampling*. Cambridge: MRC Biostatistics Unit., 1994.
- SPIEGELHALTER, D.J. et al. *BUGS - Bayesian Analysis Using Gibbs Sampling*. Version 0.5, Cambridge: MRC Biostatistics Unit, 1996.
- VICENTIN, W. et al. Fator de condição e relação peso X comprimento de *Prochilodus lineatus*, capturados na cabeceira do Rio Miranda, MS. In: SIMPÓSIO SOBRE RECURSOS NATURAIS E SÓCIO-ECONÔMICOS DO PANTANAL CORUMBÁ, 4., 2004, Corumbá. *Anais...* Corumbá: Embrapa Pantanal, 2004.
- WEATHERLEY, A.H. *Growth and ecology of fish populations*. London: Academic Press Inc., 1972. p. 283.

Received on August 11, 2006.

Accepted on March 16, 2007.