

UM CRITÉRIO PARA A SELEÇÃO DE TOUROS NELORE USANDO MODELOS DE SOBREVIVÊNCIA

Suely Ruiz GIOLO¹
Robin HENDERSON²
Clarice Garcia Borges DEMÉTRIO³

- **RESUMO:** O objetivo deste estudo é propor um critério para avaliar e identificar touros Nelore que são geneticamente superiores quanto a taxa de crescimento dos animais gerados por eles. Foram estudados registros de 6.602 animais da raça Nelore nascidos na primavera, no período de 1990 a 1998, gerados por 64 touros de um mesmo rebanho. O tempo, em dias, que um animal levou para ganhar um desejado peso, do período do nascimento até a desmama, foi usado como variável resposta no modelo estatístico considerado. Diversas covariáveis, bem como um efeito aleatório denominado fragilidade, representando em um sentido estatístico os fatores genéticos não-observáveis, foram consideradas no modelo. Como a variável tempo usualmente não segue a distribuição normal, e diversos animais não ganharam o peso desejado até a desmama, foram utilizados, para a análise desses dados, métodos que vêm sendo estudados em análise de sobrevivência para dados com censura intervalar. Uma abordagem Bayesiana foi considerada e o método de Markov Chain Monte Carlo (MCMC) foi usado para estimação. As estimativas dos parâmetros foram obtidas por meio de suas respectivas distribuições *a posteriori*.
- **PALAVRAS-CHAVE:** Fragilidade; critério de seleção; censura-intervalar; análise de sobrevivência; variância genética.

1 Introdução

O gado da raça Nelore é comumente usado no Brasil para a produção comercial de carne. Tempos não muito longos para um específico ganho de peso no período

¹Departamento de Estatística, Universidade Federal do Paraná - UFPR, CEP: 81531-990, Curitiba, PR, Brasil

²Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

³Departamento de Ciências Exatas, Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - ESALQ/USP, CEP: 13418-900, Piracicaba, SP, Brasil

do nascimento até a desmama, bem como da desmama ao abate, são, portanto, economicamente desejáveis. O principal objetivo deste estudo foi desenvolver um critério para identificar touros que geram animais com ganho rápido de peso, considerando-se como variável resposta de interesse, o tempo em dias que um animal leva para ganhar um específico peso no período do nascimento até a desmama. De acordo com Albuquerque e Fries (1998), em estudo com gado de corte da raça Nelore, um ganho de peso realístico para esse período está em torno de 160kg.

A variável tempo usualmente não segue a distribuição normal e o tempo exato que um animal leva para ganhar o peso desejado também é, em geral, desconhecido, uma vez que a pesagem diária dos animais é inviável na prática. Embora não se saiba o tempo exato em que o ganho de 160kg é atingido, sabe-se, a partir das informações do tempo observado desde o nascimento até a desmama e do respectivo peso na desmama, que o mesmo ocorre em um determinado intervalo de tempo. Por essa razão, modelos paramétricos que vêm sendo estudados em análise de sobrevivência para dados com censura intervalar foram usados para avaliar os touros. No contexto de análise de sobrevivência, classificaram-se os tempos dos animais que ganharam mais de 160kg no período do nascimento até a desmama como tempos de falha e, os tempos dos que não alcançaram esse peso, como *censurados* (Klein e Moeschberger, 1997). Tais tempos foram devidamente considerados nos modelos e os touros, com melhor desempenho, identificados, usando-se o critério proposto. Considerando-se que o peso ao nascer, é de aproximadamente 30kg, um dos objetivos nessa seleção seria o de permitir que os criadores de gado de corte comercializassem animais desmamados com peso de aproximadamente 190kg em um período de tempo mais curto.

2 Material e métodos

Os dados utilizados para ilustrar o critério proposto foram cedidos por GenSys Consultores Associados S.C. Ltda. Foram analisados registros de 6.602 animais nascidos na primavera, que foram gerados por 64 touros e 5.262 vacas de um mesmo rebanho. Desses animais, aproximadamente 65% eram fêmeas, em torno de 74% foram gerados por monta natural e 0,17% deles nasceram em 1990; 10,25% em 1993; 3,53% em 1994; 12,77% em 1995; 27,81% em 1996; 33,64% em 1997 e 11,83% em 1998. O número de animais por touro variou de 3 a 460, o tempo do nascimento até a desmama de 134 a 276 dias e o peso na desmama de 100 a 289kg. O tempo para ganhar 160kg do nascimento até a desmama, foi considerado como a variável resposta de interesse e os fatores fixos considerados foram: sexo, tipo de monta (natural ou inseminação artificial), ano de nascimento do animal e tipo de manejo alimentar (102 tipos no rebanho sob estudo).

Na desmama, foi observado que cerca de 30% dos animais tiveram ganho de peso superior a 160kg, indicando que este peso foi atingido antes daquela data. Além disso, nenhum animal atingiu o peso desejado antes de 120 dias. Por outro lado, muitos outros animais não ganharam 160kg até a desmama, e portanto, tudo o que se sabe a respeito desses animais é que eles alcançaram tal ganho de peso em algum tempo

após aquela data e, supôs-se, no máximo, 500 dias após o nascimento. Isso significa, portanto, que os tempos exatos, T_i ($i = 1, \dots, n$), dos n animais são desconhecidos. Sabe-se, contudo, que tais tempos T_i ocorreram em intervalos de tempos denotados por $[L_i, U_i]$ em que $L_i \leq T_i \leq U_i$. Para o animal que ganhou pelo menos 160kg até a desmama, assumiu-se que este atingiu os exatos 160kg no intervalo $[120 \text{ dias}, U_i]$, sendo neste caso, $U_i =$ tempo em dias do animal i observado do nascimento até a desmama. Do contrário, o intervalo assumido foi $[U_i, 500 \text{ dias}]$. Foi utilizada uma variável indicadora, denominada *status*, indicando se o animal ganhou, ou não, 160kg. Animais com tempos T_i ocorridos no intervalo $[U_i, 500 \text{ dias}]$ foram classificados como censurados, e portanto, *status* = 0 e caso contrário, *status* = 1.

Considerando, portanto, que os tempos T_i ocorreram dentro dos intervalos $[L_i, U_i]$, que alguns desses tempos são censurados e que a variável tempo não é, em geral, normalmente distribuída, foram considerados métodos em análise de sobrevivência, em um contexto Bayesiano. Um efeito aleatório foi incorporado ao modelo a fim de considerar que os tempos dos animais gerados pelo mesmo touro seriam provavelmente associados devido a fatores genéticos compartilhados. Nos modelos de componentes de variância para dados normalmente distribuídos, esse efeito é, em geral, aditivamente incorporado. O modelo proposto nesse artigo incorpora tal efeito multiplicativamente na taxa de risco individual (Hougaard, 2000), e como esse efeito aleatório não é observável, uma distribuição de probabilidade deve ser considerada. A distribuição gama, por sua simplicidade, flexibilidade e popularidade nessa área, foi assumida. O modelo proposto permite que covariáveis, isto é, efeitos fixos sejam considerados e, condicional ao efeito aleatório e às covariáveis, assumiu-se que os tempos T_i são independentes e seguem a distribuição Weibull. Portanto, a probabilidade de um animal i ($i = 1, \dots, n$) gerado pelo touro j ($j = 1, \dots, m$) com vetor p -dimensional de covariáveis \mathbf{x}_i ganhar o desejado peso após o tempo t , é representada por:

$$S(t | \mathbf{x}_i, z_j) = \exp\left\{-z_j \exp\{\alpha + \boldsymbol{\beta}'\mathbf{x}_i\} t^\gamma\right\} \quad (1)$$

em que α é uma constante, $\boldsymbol{\beta}$ denota o vetor p -dimensional de coeficientes desconhecidos associados aos efeitos fixos \mathbf{x}_i , z_j representa os fatores genéticos compartilhados por animais gerados pelo touro j e que não são mensuráveis, e γ é o parâmetro de forma da distribuição Weibull. Um animal com função de sobrevivência $S(t | \mathbf{x}_i, z_j)$ decrescendo rapidamente ao longo do tempo ganha peso rapidamente. Em termos de riscos, o modelo (1) é representado por:

$$\lambda(t | \mathbf{x}_i, z_j) = z_j \exp\{\alpha + \boldsymbol{\beta}'\mathbf{x}_i\} \gamma t^{\gamma-1}. \quad (2)$$

Pode-se ver mais claramente em (2) do que em (1) que o efeito aleatório atua multiplicativamente nos riscos individuais. Assim, se a função de risco para um determinado animal cresce rapidamente isso indica que o peso do animal está aumentando rapidamente. O interesse estará, portanto, naqueles animais cuja função de sobrevivência decresce rapidamente ou, equivalentemente, cuja função de risco cresce rapidamente.

Os componentes aleatórios z_j , $j = 1, \dots, 64$, foram assumidos como independentes e com distribuição gama com parâmetros $(1/\xi, 1/\xi)$ de modo que $E(z_j) = 1$ e $\text{Var}(z_j) = \xi$, chamada variância genética. Cada touro terá, portanto, um valor z_j associado, o que permitirá uma comparação entre eles. Valores grandes de z_j indicarão touros cujos descendentes ganham peso rapidamente. O modelo proposto em (1) permite diferentes números de descendentes por touro. Conclusões sobre aqueles touros com poucos descendentes deveriam, portanto, serem vistas com cuidado.

Para a implementação computacional do modelo proposto, utilizou-se a abordagem Bayesiana em que o método de Markov Chain Monte Carlo (MCMC) foi usado para a estimação dos parâmetros. MCMC é um método computacionalmente intensivo no qual uma cadeia estocástica dos valores dos parâmetros é gerada de modo que, após um conveniente número de iterações (*burn in*), os valores da cadeia são assumidos como amostras de uma específica distribuição de probabilidades. Propriedades como média, variância e outros momentos podem ser estimados a partir dos valores gerados. Formalmente, MCMC é uma técnica Bayesiana que requer informação *a priori* dos parâmetros de interesse. Se existir pouca ou nenhuma informação a respeito dos parâmetros, o uso de *prioris* com variâncias grandes permitirá que os dados dominem a análise, como ocorre na abordagem clássica. As distribuições *a priori* assumidas para os parâmetros e efeitos aleatórios de interesse, isto é, para $\theta = (\alpha, \beta_1, \dots, \beta_p, \gamma, \xi, z_1, \dots, z_{64})$ foram: $\alpha \sim N(0, 10^5)$, $\beta_i \sim N(0, 10^5)$, $1/\xi \sim \Gamma(0,001; 0,001)$, $\gamma \sim \Gamma(1, 10^{-5})$ e $z_j \sim \Gamma(1/\xi, 1/\xi)$ para $j = 1, \dots, 64$. Os valores iniciais considerados foram: $\alpha = 0$, $\beta_i = 1$ para $i = 1, \dots, p$, $\gamma = 1$, $1/\xi = 0,1$, $z_j = 1$ para $j = 1, \dots, 64$ e t_{ij} = tempo até a desmama. Dado que nenhum animal ganhou 160kg antes de 120 dias, por simplicidade, esse valor foi subtraído do tempo observado desde o nascimento até a desmama. Um programa, usando o pacote WinBUGS (Spiegelhalter et al., 2000), foi escrito para realizar a análise dos dados.

3 Resultados e discussão

Considerando-se, para os tempos até ganhar 160kg, o modelo de sobrevivência Weibull com fragilidade gama e censura intervalar definido em (1), com as covariáveis sexo, tipo de monta, ano de nascimento do animal e tipo de manejo, verificou-se que tipo de monta não foi estatisticamente significativo. Então, o modelo para o animal i gerado pelo touro j , ficou representado por:

$$S(t \mid \mathbf{x}_i, z_j) = \exp \left\{ -z_j \exp \{ \alpha + \beta_1 \text{sex}_i + \beta_k \text{an}_i + \beta_l \text{tma}_i \} t^\gamma \right\} \quad (3)$$

sendo, β_1, β_k e β_l os coeficientes de regressão associados, respectivamente, às covariáveis sexo (*sex*), ano de nascimento (*an*) e tipo de manejo alimentar (*tma*) do animal i .

Um resumo das distribuições *a posteriori* de alguns desses parâmetros e efeitos aleatórios z_j 's, baseado no método MCMC em que foram utilizadas 20.000 iterações, após um período *burn in* de 5.000, é apresentado na Tabela 1. Na Figura 1 podem-se observar as médias *a posteriori* de todos os z_j , $j = 1, \dots, 64$, bem como seus intervalos de credibilidade.

Tabela - 1: Resumo das distribuições *a posteriori* dos parâmetros e z_j 's

Parâmetros e z_j 's	Média	Desvio	Mediana	Intervalos de Credibilidade	
	<i>a posteriori</i>	padrão (SD)	<i>a posteriori</i>	2.5%	97.5%
α : constante	-25,21	1,535	-25,97	-27,53	-22,32
β_1 : machos	2,19	0,086	2,189	2,024	2,360
β_2 : an = 1990	2,431	1,034	2,418	0,462	4,557
β_3 : an = 1993	1,454	0,281	1,445	0,925	2,022
β_4 : an = 1994	0,890	0,417	0,880	0,117	1,735
β_5 : an = 1995	0,718	0,204	0,713	0,337	1,132
β_6 : an = 1996	0,012	0,173	0,011	-0,322	0,353
β_7 : an = 1997	0,428	0,164	0,428	0,102	0,747
β_8 : tma = 1	3,289	0,960	3,364	1,649	5,291
β_9 : tma = 2	2,386	1,013	2,335	0,651	4,466
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
γ	4,282	0,172	4,327	3,894	4,529
$1/\xi$	4,317	1,108	4,175	2,549	6,845
ξ	0,247	0,063	0,239	0,146	0,392
z_1 : touro 1	1,845	0,417	1,801	1,144	2,790
z_{25} : touro 25	0,106	0,045	0,098	0,043	0,216
z_{29} : touro 29	1,379	0,418	1,333	0,695	2,350
z_{31} : touro 31	1,746	0,392	1,708	1,103	2,615
z_{34} : touro 34	0,354	0,092	0,345	0,200	0,560
z_{37} : touro 37	1,391	0,384	1,341	0,782	2,277
z_{42} : touro 42	1,422	0,332	1,385	0,883	2,157
z_{43} : touro 43	0,387	0,175	0,359	0,127	0,804
z_{44} : touro 44	1,354	0,352	1,314	0,773	2,160
z_{47} : touro 47	1,432	0,308	1,400	0,917	2,114
z_{49} : touro 49	1,370	0,461	1,306	0,650	2,453
z_{52} : touro 52	1,825	0,319	1,795	1,285	2,532

an = ano de nascimento e tma = tipo de manejo alimentar

A partir da distribuição *a posteriori* de β_1 , foram observadas evidências de efeito do sexo na variável resposta, indicando que os machos ganharam peso mais rapidamente que as fêmeas. De modo análogo, observaram-se evidências de efeito do ano de nascimento e tipo de manejo alimentar. A distribuição *a posteriori* de ξ também mostrou que a variância genética é significativa. A média *a posteriori* da variância genética foi $\xi = 0,247$ (sd = 0,0632), fornecendo o intervalo de credibilidade a 90% para a distribuição $\Gamma(1/0,247, 1/0,247)$ de 0,43 a 1,66. Isso significa que alguns touros são capazes de gerar animais com taxa de crescimento quase 4 vezes maior do que a taxa de outros touros.

A título de ilustração, dois touros foram comparados, dentre os 64 analisados, de modo que um deles tivesse um grande valor de z e o outro um pequeno valor de z . Escolheu-se para essa comparação os touros 1 ($z = 1,845$) e 34 ($z = 0,354$). As curvas de sobrevivência *a posteriori*, obtidas de (1), encontram-se apresentadas na Figura 2, e mostram que animais gerados no ano de 1998 pelo touro 1 e com manejo alimentar do tipo 1 ganharam peso mais rapidamente do que aqueles gerados no mesmo ano e com o mesmo tipo de manejo alimentar, pelo touro 34. Note que no tempo $t = 240$ dias (linha vertical em ambos os gráficos da Figura 2) todos os animais machos gerados pelo touro 1 já haviam atingido o desejado ganho de peso (quando a curva atinge valor próximo de zero), enquanto muitos dos animais machos gerados pelo touro 34 (em torno de 50% uma vez que $\hat{S}(240) \approx 0,50$) permaneciam com ganho de peso

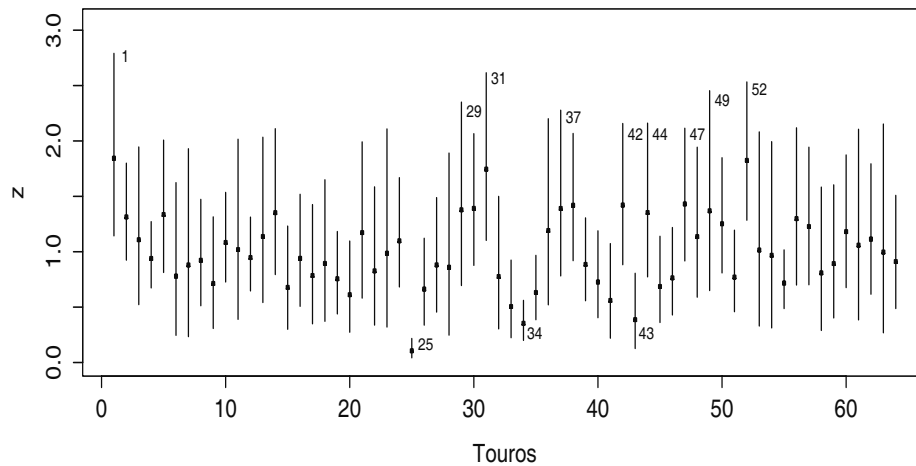


Figura - 1: Médias e intervalos de credibilidade a posteriori dos z_j 's.

inferior ao desejado. Assim, se os machos gerados pelo touro 1 ganharam 160kg no máximo em 240 dias e no mínimo em 150 dias (quando a curva começa a decrescer), tem-se que esses animais obtiveram um ganho de peso diário médio de no mínimo 0,67kg/dia e, no máximo, 1,07kg/dia. A Figura 2 mostra também que os machos ganharam peso mais rapidamente do que as fêmeas. As fêmeas do touro 1 ganharam 160kg em torno de no máximo 325 dias (quando $\hat{S}(t) \approx 0$) e, no mínimo, 175 dias (quando $\hat{S}(t)$ começa a decrescer) o que implica um ganho de peso diário médio de no mínimo 0,5kg/dia e, no máximo, 0,9kg/dia. Analogamente observa-se, a partir da Figura 2, que machos e fêmeas do touro 34 ganharam 160kg em torno de no máximo 300 e 420 dias, respectivamente, e, no mínimo, 175 e 200 dias, respectivamente. Isso implica um ganho de peso diário médio, para os machos, de no mínimo 0,53kg/dia e, no máximo, 0,90kg/dia e, para as fêmeas de, no mínimo 0,38kg/dia e, no máximo 0,80kg/dia. Portanto, touros com valores de z grandes são desejáveis, uma vez que quando z cresce as curvas de sobrevivência *a posteriori* dos animais gerados por eles decrescem mais rapidamente significando que o ganho desejado de peso dos animais são atingidos mais rapidamente. Animais gerados pelo touro 1 apresentaram, portanto, melhor desempenho do que os gerados pelo touro 34.

Pela análise das médias *a posteriori* dos z_j , $j = 1, \dots, 64$, mostradas na Figura 1, pode-se selecionar os touros geneticamente superiores em relação aos animais gerados por eles. Como uma ferramenta auxiliar nesse processo de seleção, pode-se, também, utilizar as distribuições *a posteriori* das classificações atribuídas aos touros. Essas distribuições estão representadas, para alguns touros, na Figura 3 e mostram a incerteza

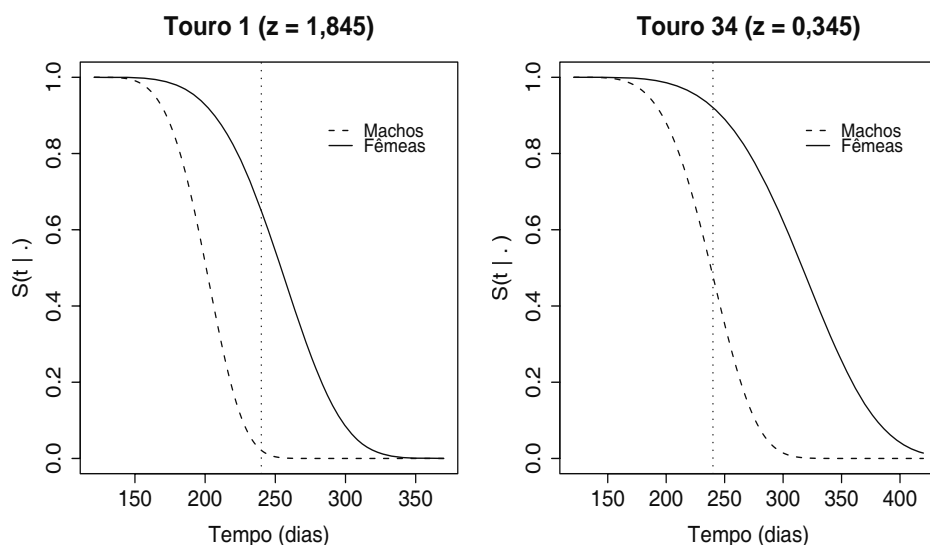


Figura - 2: Curvas de sobrevivência a posteriori para os animais machos e fêmeas dos touros 1 e 34 nascidos em 1998 e com manejo alimentar do tipo 1.

associada à classificação de cada um deles. Como valores grandes de z estão associados a classificações de ordem alta, pode-se notar pela Figura 3 alguns touros superiores tais como os touros 1, 31 e 52. Das distribuições *a posteriori* das classificações nota-se uma probabilidade grande de esses touros terem classificações de ordem alta. Por outro lado, os touros 25, 34 e 43 apresentaram probabilidade grande de terem classificações de ordem baixa o que, conseqüentemente, significa pequeno valor de z . Para esses touros espera-se desempenho pobre dos animais gerados por eles, uma vez que irão levar mais tempo para ganhar 160kg. Alguns touros, tais como 29, 37, 42, 44, 47 e 49 apresentaram desempenho mediano.

Os dados analisados neste trabalho tiveram como propósito ilustrar a metodologia proposta. Outros fatores fixos e aleatórios julgados relevantes são, portanto, possíveis de serem estudados incorporando-os ao modelo proposto.

4 Conclusão

Para a análise da variável de interesse, tempo até um ganho de peso de 160kg, foi proposto um modelo para dados de análise de sobrevivência com censura intervalar e com efeito aleatório. O efeito aleatório foi incorporado multiplicativamente ao modelo para considerar os efeitos genéticos compartilhados pelos animais gerados pelo mesmo touro. O modelo proposto atendeu aos objetivos de avaliação dos touros, fornecendo um critério de seleção dos animais. Dos resultados obtidos, verificou-se a existência de diferenças estatisticamente significativas entre os touros e, dentre os 64 touros

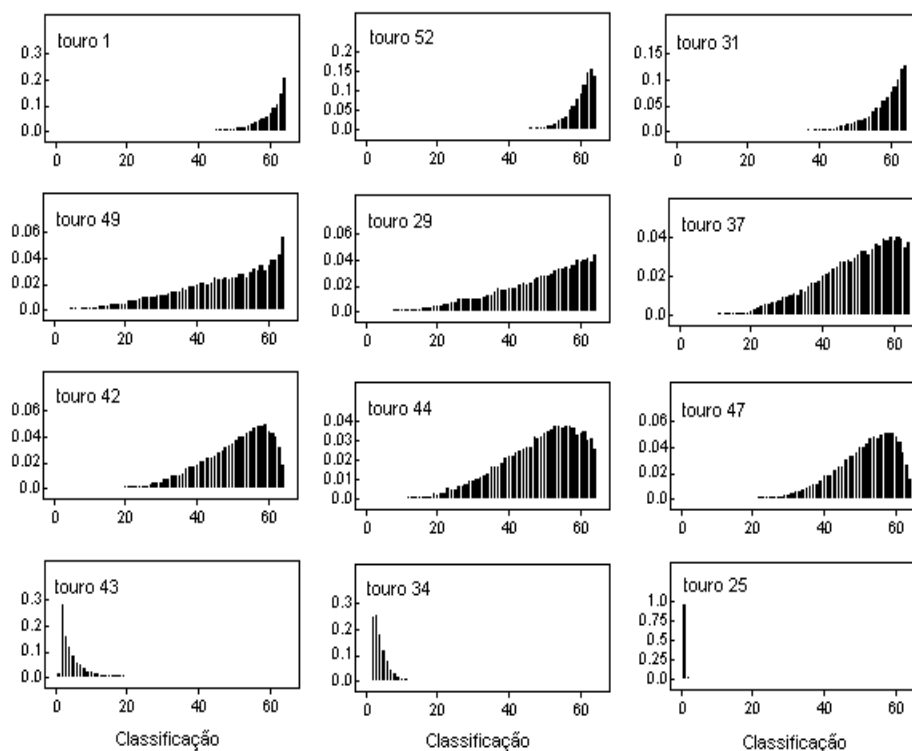


Figura - 3: Distribuições a posteriori das classificações de 12 touros.

considerados no estudo, 3 deles puderam ser identificados, em função da distribuição *a posteriori* da ordem de classificação, como tendo melhor desempenho do que os demais no sentido de que os animais gerados por eles ganharam peso mais rapidamente. Alguns touros foram capazes de produzir animais com taxa de crescimento quase 4 vezes maior do que a taxa dos animais de outros touros. Evidências de que os fatores sexo, ano de nascimento e tipo de manejo alimentar produziram efeito no tempo para ganhar 160kg foram obtidas.

Agradecimentos: À GenSys Consultores Associados S/C Ltda pelos dados utilizados nesse artigo e à CAPES pelo suporte financeiro.

GIOLO, S. R.; ENDERSON, R.; DEMÉTRIO, C.G.B. A criterion for selecting Nelore sires using survival models. *Rev. Mat. Estat.*, São Paulo, v.21, n.3, p.115-123, 2003.

- **ABSTRACT:** *The aim of this study is to propose a criterion to evaluate and identify Nelore sires who are genetically superior in terms of offspring growth rates. Data on 6602 Nelore cattle born in the spring from 1990 to 1998 and generated by 64 sires at a same farm were studied. The time in days that an animal needs to gain a desirable weight from birth to weaning was used as a response variable in the statistical model considered. Several observed covariates as well as genetic factors, which are unmeasurable, were also considered. As the response variable time is not usually distributed as Normal and there were animals which had not gained the desirable weight at weaning, methods developed in survival analysis for interval-censored data were used. A random effect, called frailty, representing, in a statistical sense unmeasurable genetic factors, was considered in the model. A Bayesian approach was considered and the Markov chain Monte Carlo (MCMC) method was used in the estimation. Estimated parameters were obtained from the posterior distributions.*

- **KEYWORDS:** *Frailty; selection criterion; interval-censored; survival analysis; genetic variance.*

Referências

- ALBUQUERQUE, L. G.; FRIES, L. A. Selection for reducing ages of marketing units in beef cattle. In: WORLD CONGRESS GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armide. *Proceedings...* Armide, 1998. v.27, p.235-238.
- HOUGAARD, P. *Analysis of multivariate survival data*. New York: Springer-Verlag, 2000. 543p.
- KLEIN, J. P.; MOESCHBERGER M. L. *Survival analysis*. New York: Springer-Verlag, 1997. 502p.
- SPIEGELHALTER, D.; THOMAS A.; BEST, N. *User manual WinBUGS 1.3*, 2000. Disponível em: <http://www.mrc-bsu.cam.ac.uk/bugs/>. Acesso em: 19 de maio de 2003.

Recebido em 10.10.2002.

Aprovado após revisão em 15.10.2003.