

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

**MODELOS BAYESIANOS PARA ANÁLISE DE
CARACTERÍSTICAS REPRODUTIVAS DE FÊMEAS
NELORE**

Diogo Anastácio Garcia

Zootecnista

2014

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**MODELOS BAYESIANOS PARA ANÁLISE DE
CARACTERÍSTICAS REPRODUTIVAS DE FÊMEAS
NELORE**

Diogo Anastácio Garcia

Orientador: Profa. Dr. Lucia Galvão de Albuquerque

**Tese apresentada à Faculdade de
Ciências Agrárias e Veterinárias
- Unesp, Câmpus de Jaboticabal,
como parte das exigências para a
obtenção do título de Doutor em
Genética e Melhoramento Animal**

2014

Garcia, Diogo Anastácio
G216m Modelos Bayesianos para análise de características reprodutivas de
fêmeas Nelore / Diogo Anastácio Garcia. – – Jaboticabal, 2014
xix, 84 p.; 28cm

Tese (Doutorado) – Universidade Estadual Paulista, Faculdade de
Ciências Agrárias e Veterinárias, 2011.

Orientador: Lucia Galvão de Albuquerque
Banca examinadora: Arione Augusti Boligon, Fabyano Fonseca e
Silva, Fernando Sebastián Baldi Rey, Henrique Nunes de Oliveira.
Bibliografia

1. Dados censurados. 2. Idade ao primeiro parto. 3. Dias para
o parto. 4. Longevidade. 5. Gado de corte. I. Título. II. Jaboticabal-
Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.2:636.082

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da
Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus
de Jaboticabal.

CERTIFICADO DE APROVAÇÃO

TÍTULO: MODELOS BAYESIANOS PARA ANÁLISE DE CARACTERÍSTICAS REPRODUTIVAS DE FÊMEAS NELORE

AUTOR: DIOGO ANASTÁCIO GARCIA

ORIENTADORA: Profa. Dra. LUCIA GALVAO DE ALBUQUERQUE

Aprovado como parte das exigências para obtenção do Título de DOUTOR EM GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:


Profa. Dra. LUCIA GALVAO DE ALBUQUERQUE

Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Prof. Dr. HENRIQUE NUNES DE OLIVEIRA

Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Prof. Dr. FERNANDO SEBASTIÁN BALDI REY

Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Prof. Dr. FABYANO FONSECA E SILVA

Universidade Federal de Viçosa / Viçosa/MG


Profa. Dra. ARIONE AUGUSTI BOLIGON

Universidade Federal de Pelotas / Pelotas/RS

Data da realização: 20 de fevereiro de 2014.

DADOS CURRICULARES DO AUTOR

Diogo Anastácio Garcia, nascido em Dionísio - MG em 05 de Maio de 1985, iniciou o curso de Zootecnia na Universidade Federal do Vales do Jequitinhonha e Mucuri (UFVJM), Diamantina - MG, em Agosto de 2003, tornando-se Bacharel em Zootecnia em Janeiro de 2008. Em Março de 2008, ingressou no curso de Pós-Graduação em Zootecnia da UFVJM, sob orientação do professor Idalmo Garcia Pereira, obtendo o grau de Mestre em Zootecnia no dia 22 de Fevereiro de 2010. Em Março de 2010, iniciou o curso de Doutorado no programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias (UNESP), Jaboticabal - SP, sob orientação da professora Lucia Galvão de Albuquerque. De Agosto de 2012 à Julho de 2013, realizou estágio de Doutorado no exterior na University of Wisconsin, sob orientação do professor Guilherme Jordão de Magalhães Rosa.

“Por vezes, do fundo obscuro da alma humana, soterrada de paixões e terrores, nasce um impulso de libertar-se da densa confusão dos tempos e erguer-se até um ponto onde seja possível enxergar, por cima do caos e das tormentas, dos prazeres e das dores, um pouco da harmonia cósmica ou mesmo, para além dela, um fragmento de luz da secreta ordem transcendente que - talvez - governa todas as coisa.

É o impulso mais alto e mais nobre da alma humana ...

Acontece que a esse impulso fundamental corresponde um outro, derivado mas não menos forte: aquele que leva o homem que entreviu a ordem e o sentido a desejar repartir com os outros homens um pouco daquilo que viu. Não há certamente maior benefício que se possa fazer a um semelhante: mostrar-lhe o caminho do espírito e da liberdade, pelo qual ele pode se elevar a uma condição que, dizia o salmista, é apenas um pouco inferior à dos anjos. Tal é, substancialmente, a forma concreta do amor ao próximo: dar ao outro o melhor e mais alto do que um homem obteve para si mesmo ...”

Olavo de Carvalho

(O espírito e cultura: o Brasil ante o sentido da vida)

Aos meus queridos pais, José e Rosângela.

As minha irmãs Gra e Tina.

A Matilde, grande companheira.

Todos aqueles que me iluminaram com suas orações, ações e palavras de sabedoria.

Dedico e Ofereço.

AGRADECIMENTOS

À Deus, por iluminar a minha caminhada.

À Profa. Dra. Lucia Galvão de Albuquerque, pela orientação, amizade e confiança.

Aos membros da banca examinadora, Profa. Dra. Arione Augusti Boligon, Prof. Dr. Fabyano Fonseca e Silva, Prof. Dr. Fernando Sebastián Baldi Rey e Prof. Dr. Henrique Nunes de Oliveira pelos importantes questionamentos e sugestões.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de Doutorado (Processo nº 2009/15010-6).

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de Doutorado Sanduíche (BEX 1817-12-4).

Ao Prof. Dr. Guilherme Jordão de Magalhães Rosa, pela orientação e valiosas discussões e sugestões durante o estágio na University of Wisconsin.

Aos criadores que integram o programa Nelore Aliança pela concessão do banco de dados.

Aos companheiros de república em Jaboticabal, Luis, Diana, Rafael Medeiros, Raúl, Jorge, Rafael Keith (Bingola) e Orlando pelos momentos descontração e amizade.

Aos amigos da salinha, pelos momentos de aprendizado, trabalho, descontração e comemorações.

SUMÁRIO

SUMÁRIO	i
RESUMO	iii
ABSTRACT	v
CAPÍTULO 1 CONSIDERAÇÕES GERAIS	7
1.1 Introdução	7
1.2 Revisão de Literatura	9
1.2.1 Idade ao Primeiro Parto	9
1.2.2 Dias para o Parto	10
1.2.3 Longevidade	11
1.3 Modelos Estatísticos	11
1.3.1 Modelo Hierárquico Bayesiano Unicaracterístico	11
1.3.2 Modelo Hierárquico Bayesiano Unicaracterístico para Dados Cen- surados	13
1.3.3 Modelo Linear-Limiar	14
1.3.4 Análise de Sobrevida	17
1.4 Objetivos	19
REFERÊNCIAS	20
CAPÍTULO 2 ANÁLISE DE IDADE AO PRIMEIRO PARTO CONSIDERANDO A INCLUSÃO DE DADOS CENSURADOS	26
2.1 Introdução	27
2.2 Material e Métodos	28
2.3 Resultados e Discussão	34
2.4 Conclusões	39
REFERÊNCIAS	40
CAPÍTULO 3 MODELAGEM DE INFORMAÇÕES CENSURADAS NA AVALIA- ÇÃO DE DIAS PARA O PARTO	46
3.1 Introdução	47
3.2 Material e Métodos	47
3.3 Resultados e Discussão	55
3.4 Conclusões	59
REFERÊNCIAS	61

CAPÍTULO 4	AVALIAÇÃO DA LONGEVIDADE PRODUTIVA MODELANDO RE-	
	GISTROS CENSURADOS	67
4.1	Introdução	67
4.2	Material e Métodos	68
4.3	Resultados e Discussão	75
4.4	Conclusões	78
REFERÊNCIAS		79

MODELOS BAYESIANOS PARA ANÁLISE DE CARACTERÍSTICAS REPRODUTIVAS DE FÊMEAS NELORE

RESUMO - As características reprodutivas são importantes para o sistema produtivo de carne bovina e, algumas delas, como a idade ao primeiro parto (IPP), dias para o segundo parto (DPP) e longevidade produtiva (LONG), são de fácil mensuração e não oneram o sistema produtivo para a obtenção fenotípica de sua expressão. Porém, estas características reprodutivas apresentam herdabilidade de baixa magnitude e, uma das razões, pode ser o fato de estas serem mensuradas apenas nas fêmeas consideradas férteis, pois no momento da avaliação genética, uma porção das fêmeas, normalmente, não apresentam fenótipos devido a falhas reprodutivas, descarte voluntário e involuntário anterior à estação de monta e, em alguns casos, por registros inconsistentes. Estes registros parciais ou desconhecidos são denominados informações censuradas. De acordo com alguns estudos, a inclusão adequada das informações censuradas poderiam evitar distorções nas inferências e melhoria na predição dos valores genéticos. Algumas metodologias têm sido utilizadas para lidar com a censura nas avaliações genéticas como modelos com penalização, modelos hierárquicos Bayesianos para dados censurados, modelo linear-limiar e o de análise de sobrevivência. Porém no Brasil, essas abordagens não vêm sendo muito utilizadas. Este estudo tem como objetivo avaliar as características reprodutivas de idade ao primeiro parto, dias para o segundo parto e longevidade produtiva, de fêmeas da raça Nelore, utilizando modelos que tratam as expressões fenotípicas destas características como não censuradas e censuradas, com a finalidade de encontrar modelos adequados para avaliação genética dos animais para estas características. Foram utilizadas fêmeas nascidas entre 1993 a 2007, da região sudeste e centro-oeste do Brasil, as quais fazem parte da base de dados Aliança Nelore. Para IPP, os grupos de contemporâneos (GC) foram definidos pela combinação de safra, mês e fazenda de nascimento, fazenda e grupo de manejo à desmama e fazenda e grupo de manejo ao sobreano. Os GC das características DPP e LONG, foram definidos como anteriormente, porém com a adição da classe de precocidade, matrizes com IPP inferior a 30 meses foram classificadas como precoces. Apenas GC com mais de 10 fêmeas com informação fenotípica não censuradas foram utilizadas nas análises de IPP, por outro lado, a restrição utilizada para DPP e LONG foi de pelo menos 5 animais com registros fenotípicos conhecidos. Os modelos lineares Gaussiano (ML), Gaussiano com penalização (MLP), Gaussiano com censura (MLC), linear-limiar (ML), linear-limiar com censura (MLC) e o modelos de riscos proporcionais Weibull (MW) foram utilizados na avaliação de IPP e DPP, porém para LONG apenas os modelos ML, MLC e MW foram empregados. A comparação dos modelos foi realizada utilizando correlações de Pearson (CP), dos valores genéticos preditos, e a porcentagem de coincidência na classificação dos 10% melhores touros (TOP10), utilizando todos os reprodutores ou touros com mais de 10 filhas com registros fenotípicos. Além disso, um esquema de validação cruzada com 10 repetições foi realizado, em que o banco de dados foi particionado em treinamento e validação, sendo estes compostos, respectivamente, por 80 e 20% dos registros. A habilidade de predição dos modelos foi testada utilizando informações de eventos reprodutivos esperados e observados de progênies dos touros, com mais de 10 filhas não censuradas, por meio da estatística χ^2 . A inclusão de informações censuradas na análise de

IPP acarretou diminuição da estimativa de herdabilidade para os modelos lineares (0,18 para 0,12). As CP entre os modelos lineares foram de alta magnitude (0,92 a 0,99; 0,93 a 0,99). Por outro lado, menores associações foram obtidas entre MW-ML, MW-MLL e MW-MLLP (-0,74 a -0,88; -0,76 a -0,89) quando comparadas aos modelos MW-MLP e MW-MLC (-0,92 e -0,90; -0,92 e -0,90). O TOP10 entre ML, MLP, MLC, MLL e MLLP variaram em torno de 89,66 a 97,41% e 86,11 a 100,00% e, destes com MW foram de 71,55 a 81,03% e 69,44 a 80,56%. Em termos de habilidade de predição, o MW apresentou a menor média de χ^2 e o maior número de vezes em que possuiu o menor valor desta estatística. Em contrapartida ao observado para IPP, as estimativas de herdabilidade de DPP aumentaram com a inclusão de registros censurados (0,03 para 0,08). As CP entre os modelos MLP, MLC, MLLP e MW foram de alta magnitude (-0,94 a 0,99; -0,93 a 0,99), entretanto uma menor associação foi obtida entre as predições destes modelos e os modelos ML (-0,60 a 0,80; -0,55 a 0,78) e MLL (0,84 a 0,92; -0,82 a 0,93). O TOP10 variou de 81,52 a 91,30% e 88,89 a 94,44% entre os modelos MLP, MLC, MLLP e MW. Por outro lado, a correspondência de classificação destes com os modelos ML e MLL variaram, respectivamente, de 58,70 a 90,22% e 50,00 a 88,89% e, entre ML e MLL foi de 63,04 e 55,56%. O modelo MLC apresentou a maior habilidade de predição, pois apresentou menor média da estatística χ^2 e o maior número de repetições em que obteve o menor valor desta estatística. Para LONG, a inclusão de informações censuradas não alterou a estimativa de herdabilidade. Maiores CP foram obtidas entre MLC e MW (-0,89 e -0,91) e, destes modelos com ML foram, respectivamente, 0,82 e -0,68, 0,87 e -0,74. O TOP10, quando todos os touros foram considerados, o maior valor de coincidência observado foi entre ML e MLC (74,60%), por outro lado, quando reprodutores com mais de 10 filhas foram avaliados, o maior TOP10 foi entre MLC e MW (77,14%). Quando à habilidade de predição, o modelo MW apresentou melhor performance. A inclusão de informações censuradas na análise acarretou mudança nas estimativas de herdabilidade das características idade ao primeiro parto e dias para o segundo parto. Para avaliação genética de IPP e LONG o modelo de riscos proporcionais Weibull seria recomendado e, para DPP, o modelo linear Gaussiano com censura seria o mais indicado.

Palavras-chave: dados censurados, idade ao primeiro parto, dias para o parto, longevidade, gado de corte.

BAYESIAN MODELS FOR REPRODUCTIVE TRAITS ANALYSIS OF THE NELLORE FEMALES

ABSTRACT - Reproductive traits are very important to beef cattle production system and, some of them, such as age at first calving (AFC), days to calving (DC), and longevity (LONG) are easy to measure and do not bring further measurement costs to the production system. However, these reproductive traits show low heritabilities and, one of the reasons might be the fact that only fertile females have phenotypic records, because when the genetic evaluation is performed some cows have missing records due to reproductive failure, as well as due to voluntary and involuntary culling, which are made before the breeding season, and, in some cases, due to inconsistent phenotypic informations. These partially or unobserved phenotypes are termed as censored records. According to some studies, properly accounting for censored records in genetic analysis could avoid distorted inferences and better genetic breeding value predictions. Several methodologies have been proposed to cope with censored records in genetic evaluations, such as models in which censored records are penalized, Bayesian hierarchical models to deal with censored data, threshold-linear models, and frailty survival models. However, these approaches have not been widely applied in Brazil. The objectives of this study were to evaluate age at first calving, days to calving, and longevity, of the Nellore breed females, applying models which handle phenotypic informations as uncensored and censored, aiming to find appropriate models to genetic evaluation for these traits. Data were provided by Nellore Alianca and reproductive records of animals born from 1993 to 2007, in farms located in south-east and middle-west of Brazil, were considered in the study. Contemporary groups (CG) for AFC were defined by the combination of farm, month, and year of birth, farm and management group at weaning, and farm and management group at long-yearling. For DC and LONG, the CG were defined as for AFC, however the precocious class was included, in which heifers with AFC less the 30 months were classified as precocious. Only CG with more than 10 uncensored animals were included in the analysis for AFC, on the other hand, the GC restriction for DC and LONG were at least 5 animals with observed records by CG. Gaussian linear model (LM), Gaussian linear model with penalty (PLM), right-censored Gaussian linear model (CLM), threshold-linear model (TLM), threshold-linear model with penalty (PTLM), and Weibull proportional hazard model (WM) were applied in the AFC and DC analysis, however only LM, CLM, and WM were used in the LONG evaluation. The alternative methodologies were compared, in terms of breeding value prediction, using Pearson correlations (r) and the percentage of bulls in common considering only 10% of sires with the highest breeding values (TOP10%), in which all sires and bulls with at least 10 uncensored daughters were considered. Moreover, predictive ability of models were assessed by 10 repetitions of a cross validation scheme. The data set was divided into training and validation in which 80 and 20% of animals were randomly allocated, respectively. Chi-squared statistic (χ^2) was applied to evaluate model prediction ability in terms of expected and observed fertility events of daughters of sires with at least 10 uncensored progenies in the whole data. Censored data inclusion in the IPP analysis result in lower heritability estimates for linear models (0.18 to 0.12). The r among linear models showed high estimates (0.92 to 0.99; 0.93 to 0.99). On the other hand, lower associations were

observed between WM-LM, WM-TLM, WM-PTLM (-0.74 to -0.88; -0.76 to -0.89) when compared to WM-PLM and WM-CLM (-0.92 and -0.90; -0.92 and -0.90). TOP10 among LM, PLM, CLM, TLM, and PTLM varied around 89.66 to 97.41% and 86.11 to 100.00% and, between these models and WM the TOP10 observed were of 71.55 to 81.03% and 69.44 to 80.56%. In terms of prediction ability, WM showed the lowest χ^2 mean and the highest number of repetitions in which had the smallest value of this statistics. For DPP, heritability estimates increased when censored information was handled using linear models (0.03 to 0.08). Higher associations were observed among PLM, CLM, PTLM, and WM (-0.94 to 0.99; -0.93 to 0.99), however lower r estimates were found between breeding values predictions of these approaches and the models LM (-0.60 to 0.80; -0.55 to 0.78) and TLM (0.84 to 0.92; -0.82 to 0.93). The TOP10 varied from 81.52 to 91.30% and 88.89 to 94.44% among PLM, CLM, PTLM, and WM. On the other hand, the coincidence between these approaches and the models LM and TLM were, respectively, from 58.70 to 90.22% and 50.00 to 88.89% and, between LM and TLM the observed values were of 63.04 and 55.56%. CLM showed the better predictive ability, because the lowest χ^2 mean and the highest number of repetitions in which had the smallest value of this statistics were observed. In the LONG analysis, heritability estimates did not change with the inclusion of censored records. Higher r estimates were observed between CLM and WM (-0.89 and -0.91) and, between these models and LM were, respectively, 0.82 and -0.68, 0.87 and -0.74. The highest TOP10 observed was between LM and CLM (74.60%), when all bull were considered. On the other hand, when sires with more than 10 uncensored daughters were taking into account, the most favourable TOP10 was between CLM and WM (77.14%). In terms of prediction ability, the WM showed the better performance. Dealing with censored data in the analysis result in heritability estimates changes for AFC and DC. For age at first calving and longevity genetic evaluation, the Weibull proportional hazard model would be recommended and, for days to calving, the right-censored Gaussian linear model would be the most adequate model to perform genetic analysis.

Key words: censored data, age at first calving, days to calving, longevity, beef cattle.

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1.1 INTRODUÇÃO

Na condução de programas de melhoramento genético as características alvo, para a seleção, devem apresentar as seguintes atribuições desejáveis: ter importância econômica, serem herdáveis e de fácil mensuração e, se possível, não gerarem custos adicionais para o sistema produtivo quando de sua obtenção.

As características reprodutivas são importantes para o sistema produtivo de carne bovina e, algumas delas, como a idade ao primeiro parto, dias para o parto e longevidade, são de fácil mensuração e não oneram o sistema produtivo para a obtenção fenotípica de sua expressão. Entre as principais vantagens de se emprenhar as novilhas mais jovens estão o menor tempo para obter retorno do investimento, aumento da vida produtiva da vaca e do número de bezerros produzidos (SHORT et al., 1994). A característica dias para o parto permite a identificação de animais com maior fertilidade no rebanho, das fêmeas que emprenham mais cedo dentro da estação de monta e dos touros que produzem filhas que emprenham mais cedo na estação de monta (PEREIRA; ELER; FERRAZ, 2000; FORNI; ALBUQUERQUE, 2006). A maximização da longevidade gera otimização dos lucros com a diminuição das taxas de descarte involuntário, permitindo ao produtor realizar maior taxa de descarte voluntário, aumentando o ganho genético (VAN ARENDONK, 1986). Adicionalmente, o decréscimo dos custos de reposição e o aumento do número de vacas no rebanho, no ápice de produção, também geram maior rentabilidade (BOLDMAN; FREEMAN; HARRIS, 1992; ALLAIRE; GIBSON, 1992).

Estimativas de herdabilidade dessas características reprodutivas descritas na literatura apresentam-se na faixa de classificação denominada de baixa magnitude (MERCADANTE; LÔBO; OLIVEIRA, 2000; PEREIRA; ELER; FERRAZ, 2000; PEREIRA; ELER; FERRAZ, 2002; DIAS; EL FARO; ALBUQUERQUE, 2004; FORNI; ALBUQUERQUE, 2005; FORNI; ALBUQUERQUE, 2006; BOLIGON; RORATO; ALBUQUERQUE, 2007; BALDI et al., 2008; BOLIGON; ALBUQUERQUE; RORATO, 2008). Entretanto, estimativas de herdabilidade moderadas têm sido encontradas para idade ao primeiro parto (MERCADANTE; LÔBO; OLIVEIRA, 2000; AZEVÊDO et al., 2006; FARIA et al., 2007) e longevidade no rebanho (GIANLORENÇO et al., 2003; BERTAZZO et al., 2004; MELLO et al., 2006).

Uma das razões destas baixas herdabilidades pode ser o fato das características de reprodução, normalmente registradas a campo, serem mensuradas apenas nas fêmeas consideradas férteis, que parem pelo menos uma vez. Com isso, as amostras são, usualmente, viesadas, contribuindo para mascarar a variabilidade das

características e as diferenças genéticas entre os animais (NOTTER, 1988; MEYER; HAMMOND; PARNELL, 1990; MERCADANTE, 1995).

Características como idade ao primeiro parto, dias para o parto e longevidade são exemplos de dados censurados, pois muitos animais ainda estão em reprodução no momento da avaliação e, portanto, somente o limite inferior do seu valor fenotípico é conhecido (VUKASINOVIC; MOLL; KUNZI, 1997a; VUKASINOVIC; MOLL; KUNZI, 1997b; CARDOSO et al., 2009). Tratando os dados censurados como não censurados ou excluindo estes das avaliações genéticas leva-se a um favorecimento de touros que possuem grande proporção de filhas com informações censuradas (HOU et al., 2009). De acordo com Guo, Gianola e Short (2001), a avaliação genética sem incorporar as informações dos dados censurados pode conduzir a distorções na inferência e na classificação dos animais e, a estimação de valores genéticos dos touros pode ser melhorada incluindo de forma adequada os dados censurados.

Alguns métodos têm sido usados para lidar com a censura nas avaliações genéticas das características reprodutivas. Uma abordagem é tratar o registro censurado como uma informação verdadeira após adicionar uma penalidade, como por exemplo, quando IPP é avaliada podemos aplicar uma penalização correspondente à 21 dias, a qual corresponderia a um ciclo estral (JOHNSTON; BUNTER, 1996; DONOGHUE; REKAYA; BERTRAND, 2004a; DONOGHUE; REKAYA; BERTRAND, 2004b; URIOSTE; MISZTAL; BERTRAND, 2007a; HOU et al., 2009). Outra abordagem é utilizar uma densidade condicional completa *a posteriori* na amostragem de valores para as informações censuradas utilizando uma distribuição normal truncada (DONOGHUE; REKAYA; BERTRAND, 2004a; DONOGHUE; REKAYA; BERTRAND, 2004b; KORSGAARD et al., 2003; ARANGO et al., 2005; CHANG et al., 2006; GONZÁLEZ-RECIO et al., 2006; HOU et al., 2009). Além desses, modelos de análise de sobrevivência têm sido aplicados para avaliar características relacionadas à reprodução (DUCROCQ, 1993; DUCROCQ; CASELLA, 1996; SCHNEIDER et al., 2005; SCHNEIDER et al., 2006; PEREIRA et al., 2005; PEREIRA et al., 2007; GONZÁLEZ-RECIO et al., 2006; HOU et al., 2009). Outra alternativa é o uso do modelo linear-limiar, que inclui uma característica binária indicadora do estado de censura a uma característica de fertilidade contínua (DONOGHUE et al., 2004c; URIOSTE; MISZTAL; BERTRAND, 2007b; URIOSTE; MISZTAL; BERTRAND, 2007a; HOU et al., 2009). Entretanto no Brasil, estas abordagens ainda têm sido pouco empregadas (PEREIRA et al., 2005; PEREIRA et al., 2007; PEREIRA, 2008; CARDOSO et al., 2009; SILVA et al., 2010; CAETANO, 2011).

1.2 REVISÃO DE LITERATURA

1.2.1 Idade ao Primeiro Parto

A característica idade ao primeiro parto (IPP), segundo Azevêdo et al. (2006), é de fácil mensuração e está relacionada à longevidade potencial da fêmea. De acordo com Short et al. (1994), o procedimento de emprenhar novilhas mais jovens promove o retorno mais rápido do capital investido, além de aumentar a vida reprodutiva e o número de progênies por vaca.

As estimativas de herdabilidade para IPP, na raça Nelore, apresentam valores variando de 0,08 a 0,16 (PEREIRA; ELER; FERRAZ, 2000; DIAS; EL FARO; ALBUQUERQUE, 2004; BOLIGON; RORATO; ALBUQUERQUE, 2007; BOLIGON et al., 2008). Segundo Mercadante (1995), as características de reprodução normalmente registradas a campo são mensuradas apenas nas fêmeas consideradas férteis. Assim, as amostras são, usualmente, viesadas, o que pode contribuir para mascarar a variabilidade das características estudadas e as diferenças genéticas existentes entre os animais. Entretanto, Mercadante, Lôbo e Oliveira (2000) e Azevêdo et al. (2006), estimaram coeficientes de herdabilidade entre 0,21 e 0,27. Faria et al. (2007), utilizando a inferência bayesiana, encontraram herdabilidade média de 0,40 e 0,41, em análises bi-características com perímetro escrotal medido aos 365 e 450 dias.

Boligon et al. (2008) salientaram que diferentes populações e modelos causam disparidades nas estimativas de parâmetros genéticos dessa característica. Isso ocorre porque a idade ao primeiro parto é muito dependente do manejo reprodutivo e da idade em que as fêmeas são desafiadas à concepção. Assim, dependendo da forma que tais efeitos são considerados ou não nos modelos, as estimativas obtidas podem sofrer grandes variações.

Silva et al. (2010), trabalhando com animais da raça Nelore, utilizaram a metodologia de penalização de 21 dias (P21) (JOHNSTON; BUNTER, 1996), modelo de censura utilizando uma distribuição normal truncada (NT) para a geração dos dados censurados (SORENSEN; GIANOLA; KORSGAARD, 1998) e, uma terceira alternativa metodológica considerando um modelo categórico (MC) (NOTTER, 1988), em que a expressão fenotípica é categorizada com 1 e 2, sendo respectivamente, para as fêmeas que não possuem e que apresentam informação de IPP. Os autores obtiveram resultados semelhantes de estimativas de herdabilidade para IPP nas três metodologias, sendo estas de 0,071, 0,071 e 0,077 para os modelos P21, NT e MC, respectivamente. Entretanto, os mesmos ressaltaram que a baixa variabilidade genética e a coincidência de resultados para as três metodologias podem estar associados ao fato

de que a grande maioria das fazendas que disponibilizaram os dados não expõem as novilhas à reprodução precocemente.

1.2.2 Dias para o Parto

A característica dias para o parto (DPP), calculada como o intervalo, em dias, do primeiro dia da estação de monta à data do parto subsequente, tem sido considerada como alternativa para se avaliar a fertilidade das fêmeas, especialmente em gado de corte (PEREIRA; ELER; FERRAZ, 2000; FORNI; ALBUQUERQUE, 2005). Esta característica reflete a variabilidade das fêmeas em apresentar atividade de estro durante a estação de monta, o número de serviços requeridos para a prenhez, a duração da gestação e, conseqüentemente, a data do parto (NEWMAN; MORRIS; BAKER, 1992; FORNI; ALBUQUERQUE, 2006).

Estudos com animais da raça Nelore reportam herdabilidade para DPP variando de 0,04 a 0,08 (PEREIRA; ELER; FERRAZ, 2000; PEREIRA; ELER; FERRAZ, 2002; FORNI; ALBUQUERQUE, 2005; FORNI; ALBUQUERQUE, 2006; BOLIGON; ALBUQUERQUE; RORATO, 2008). As baixas estimativas de herdabilidade podem ser reflexo da utilização nas análises apenas das fêmeas que emprenharam (FORNI; ALBUQUERQUE, 2005). Segundo Boligon, Albuquerque e Rorato (2008), mesmo utilizando um arquivo de genealogia completo, muitas vezes não é possível obter estimativa acurada da variabilidade genética na população.

Mercadante et al. (2002) obtiveram estimativas de herdabilidade variando de 0,12 a 0,16 quando registros de fêmeas não paridas foram considerados. A observação de caracteres de reprodução somente das fêmeas férteis, que parem pelo menos uma vez, contribui para mascarar a variabilidade das características e as diferenças genéticas entre os animais (NOTTER, 1988; MEYER; HAMMOND; PARNELL, 1990).

Donoghue, Rekaya e Bertrand (2004a) e Donoghue, Rekaya e Bertrand (2004b), compararam por meio de dados simulados e reais de dias para o primeiro parto de novilhas em rebanhos de animais da raça Angus australianos, a adequação de duas estratégias para lidar com dados censurados de DPP. A penalização dentro de grupo de contemporâneo, com a utilização do modelo animal tradicional, e o modelo hierárquico bayesiano com ampliação para dados censurados. Os autores não observaram grande distinção entre os modelos em termos de correlações de Pearson entre os valores genéticos preditos e verdadeiros. Entretanto, o modelo hierárquico de ampliação de dados proporcionou melhor ajuste, estimativas mais precisas da variância residual e da herdabilidade da característica e também melhor identificação de quais touros estavam entre os 10% piores na classificação, quando o percentual de regis-

tros censurados era de 20%. Além disso, no estudo com dados reais, concluíram que utilizando o modelo em que os dados censurados são ignorados há subestimação da variância genética aditiva e classificação inadequada dos touros.

Estudos associando dias para o parto com idade ao primeiro parto apresentam correlações genéticas positivas. Forni e Albuquerque (2005) e Boligon, Albuquerque e Rorato (2008), estimaram correlação genética de DPP com IPP de 0,76 e 0,73, respectivamente. Portanto, indicando a possibilidade de obtenção de resposta correlacionada favorável entre as características.

1.2.3 Longevidade

Uma importante característica de fêmeas é a longevidade (LONG), que indica sua habilidade reprodutiva, uma vez que matrizes que apresentam falhas na reprodução tendem a ser descartadas mais cedo (MERCADANTE; LÔBO; OLIVEIRA, 2000).

Mercadante, Lôbo e Oliveira (2000) e Baldi et al. (2008) estimaram coeficientes de herdabilidade de 0,08 e 0,06 para tempo de permanência no rebanho, trabalhando com animais da raça Nelore e Canchim, respectivamente. Entretanto, Gianlorenço et al. (2003) e Mello et al. (2006), ambos utilizando inferência bayesiana na estimação de parâmetros genéticos de animais da raça Canchim, obtiveram estimativas de herdabilidade médias de 0,24 e 0,22. Bertazzo et al. (2004), trabalhando com informações de fêmeas Nelore, estimaram herdabilidade de 0,46 para LONG. Por outro lado, Caetano (2011), trabalhando com idade de fêmeas Nelore ao último parto, reportou herdabilidade de 0,25 utilizando modelo de análise de sobrevivência.

Trabalhos associando longevidade com idade ao primeiro parto são escassos e apresentam resultados discrepantes. Bertazzo et al. (2004), estimaram correlação genética de LONG com IPP de 0,46. Porém, Baldi et al. (2008), obtiveram estimativas igual a -0,20. Portanto, há carência de estudos que relacionem longevidade com características reprodutivas economicamente importantes.

1.3 MODELOS ESTATÍSTICOS

1.3.1 Modelo Hierárquico Bayesiano Unicaracterístico

Considerando a teoria dos modelos lineares generalizados (MCCULLAGH; NELDER, 1989), a especificação do modelo é definida pelos seguintes componentes: i) Componente aleatório determinando a distribuição da variável resposta, a qual deve possuir densidade pertencente à família exponencial, como por exemplo, as distribuições multinomial ou normal; ii) Preditor linear (η), no qual as variáveis sistemáticas

são combinadas a partir da soma de seus efeitos; iii) Função de ligação ($g(\cdot)$), a qual relaciona os componentes sistemáticos e aleatórios do modelo, sendo o preditor linear associado à esperança da característica em análise, $E[Y | \eta] = g(\eta)$. Logo, quando a característica em análise é assumida possuir distribuição normal, a função de ligação a ser adotada é a identidade, sendo portanto a associação entre preditor linear e a esperança representada por $E[Y | \eta] = \eta$.

Segundo Cardoso et al. (2009), na formação em estágios hierárquicos do modelo animal o primeiro estágio corresponde à distribuição condicional de amostragem do vetor de dados y . No segundo estágio, especificam-se as pressuposições *a priori* das incógnitas usadas para determinar a média e variância da distribuição normal, determinada no primeiro estágio. Para os parâmetros de locação adotam-se densidades *a priori* uniformes para os efeitos fixos (β) e normal multivariada para os efeitos aleatórios (u), e uma distribuição qui-quadrado invertida escalonada para o componente de variância residual (σ_e^2). No terceiro estágio, é feita a pressuposição de uma *priori* conjugada para a variância genética aditiva (σ_u^2) por meio de uma distribuição qui-quadrado invertida escalonada.

A distribuição *a posteriori* conjunta de todos os parâmetros do modelo é obtida pelo do produto das densidades especificadas nos três estágios:

$$p(\beta, u, \sigma_e^2, \sigma_u^2 | y, V_e, S_e^2, V_u, S_u^2) \propto p(y | \beta, u, \sigma_e^2) p(\beta) p(u | \sigma_u^2) p(\sigma_e^2 | V_e, S_e^2) p(\sigma_u^2 | V_u, S_u^2) \quad (1.1)$$

onde V_e e V_u são os graus de liberdade das distribuições *a priori* de σ_e^2 e σ_u^2 , respectivamente.

As distribuições condicionais completas de cada parâmetro ou conjunto de parâmetros necessários para implementar a inferência por meio de métodos Monte Carlo via Cadeias de Markov (MCMC), são derivadas a partir de (1.1) (WANG; RUTLEDGE; GIANOLA, 1994). Para os parâmetros de locação $\theta' = [\beta' \ u']$, pode ser demonstrado que a distribuição condicional completa é normal multivariada (GIANOLA; FERNANDO, 1986):

$$\theta | y, \sigma_e^2, \sigma_u^2, V_e, S_e^2, V_u, S_u^2 \sim N(\hat{\theta}, C^{-1}) \quad (1.2)$$

em que: $\hat{\theta}$ é a solução para o sistema de equações $C\theta = b$; $C = W'R^{-1}W + \Sigma^{-1}$;

$$W' = \begin{bmatrix} X' & Z' \end{bmatrix}; R^{-1} = I_{\frac{1}{\sigma_e^2}}; \Sigma^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & A^{-1} \frac{1}{\sigma_u^2} \end{bmatrix}; b = W'R^{-1}y.$$

Para os parâmetros de dispersão temos as seguintes distribuições condicionais completas qui-quadrado invertida escalonadas:

$$\sigma_e^2 | \mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, V_e, S_e^2, V_u, S_u^2 \sim \left[\sum_{j=1}^m (y_j - \mu_j) + V_e S_e^2 \right] \chi_{(V_e+q)}^{-2} \quad (1.3)$$

$$\sigma_u^2 | \mathbf{y}, \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, V_e, S_e^2, V_u, S_u^2 \sim \left(\mathbf{u}' A^{-1} \mathbf{u} + V_u S_u^2 \right) \chi_{(V_u+q)}^{-2} \quad (1.4)$$

onde A^{-1} representa a inversa da matriz de parentesco.

A estratégia para implemetar a amostragem de Gibbs é baseada na geração de uma cadeia de amostras sucessivas de (1.2), (1.3) e (1.4).

1.3.2 Modelo Hierárquico Bayesiano Unicaracterístico para Dados Censurados

Para contemplar a análise de dados censurados, utilizando modelos hierárquicos bayesianos é preciso reformular o primeiro estágio. Neste caso, o vetor de dados $\mathbf{y}$ pode ser particionado em dois sub-vetores $\mathbf{y}_{nc}$, contendo os registros observados, e $\mathbf{y}_c$, contendo os $n - m$ registros não observados dos animais com dados censurados, portanto $\mathbf{y}' = [\mathbf{y}_{nc}' \quad \mathbf{y}_c']$, apresentada a seguinte densidade condicional nos parâmetros (SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001; CARDOSO et al., 2009):

$$p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{m}{2}} \exp \left\{ \left[-\frac{1}{2\sigma_e^2} \sum_{j=1}^m (y_{ncj} - \mu_j)^2 + \sum_{j=m+1}^n (y_{cj} - \mu_j)^2 \right] \right\} \quad (1.5)$$

em que: $y_{cj} > c_j$, representando que o registro não observado (y_{cj}) deve ser maior ao ponto de truncamento (c_j) definido como registro censurado.

A distribuição *a posteriori* conjunta de todos os parâmetros do modelo é obtida pelo do produto de todas as densidades especificadas nos três estágios (CARDOSO et al., 2009):

$$p(\boldsymbol{\beta}, \mathbf{u}, \mathbf{y}_c, \sigma_e^2, \sigma_u^2 | \mathbf{y}_{nc}, V_e, S_e^2, V_u, S_u^2) = p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) p(\boldsymbol{\beta}) p(\mathbf{u} | \sigma_u^2) p(\sigma_e^2 | V_e, S_e^2) p(\sigma_u^2 | V_u, S_u^2) \prod_{j=m+1}^n I(y_{cj} > c_j) \quad (1.6)$$

onde $I(.)$ é uma variável indicadora de censura dos registros.

Os fenótipos não observados (y_{c_j}) possuem densidade condicional completa *a posteriori* dada pela distribuição normal truncada, a qual pode ser representada com a seguir (SORENSEN; GIANOLA; KORSGAARD, 1998):

$$p(y_{c_j} | \beta, \mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y}_{nc}; y_{c_j} > c_j) = \frac{\phi_{y_{c_j}}(\mu_j, \sigma_e^2)}{1 - \Phi\left(\frac{c_j - \mu_j}{\sigma_e}\right)} \quad (1.7)$$

em que: $\phi(.)$ e $\Phi(.)$ são, respectivamente, as funções densidade e acumulada da distribuição normal.

As distribuições condicionais completas dos parâmetros de locação, $\theta' = [\beta' \quad \mathbf{u}']$, e para os parâmetros de dispersão, σ_e^2 e σ_u^2 , permanecem as mesmas do modelo hierárquico Bayesiano unicaracterístico, apenas refletindo a mudança no modelo amostral ocasionada pela utilização da técnica de aumento de dados:

$$\sigma_e^2 | \mathbf{y}, \beta, \mathbf{u}, \sigma_u^2, V_e, S_e^2, V_u, S_u^2 \sim \left[\sum_{j=1}^m (y_{ncj} - \mu_j) + \sum_{j=m+1}^n (y_{c_j} - \mu_j)^2 + V_e S_e^2 \right] \chi_{(V_e+q)}^{-2} \quad (1.8)$$

$$\sigma_u^2 | \mathbf{y}, \beta, \mathbf{u}, \sigma_e^2, V_e, S_e^2, V_u, S_u^2 \sim \left(\mathbf{u}' A^{-1} \mathbf{u} + V_u S_u^2 \right) \chi_{(V_u+q)}^{-2} \quad (1.9)$$

A estratégia para implementar a amostragem de Gibbs é baseada na geração de uma cadeia de amostras sucessivas de (1.2), (1.7), (1.8) e (1.9).

1.3.3 Modelo Linear-Limiar

Nesta abordagem usa-se o estado de censura das observações como uma característica correlacionada para melhorar a acurácia das estimativas dos parâmetros genéticos e da predição dos valores genéticos dos animais. Então, o modelo linear-limiar inclui um modelo limiar para indicar o estado de censura e um modelo linear Gaussiano, onde os dados censurados são considerados como informações verdadeiras sem nenhuma penalidade. Supõe-se que a correlação entre a característica de interesse e a predisposição de censura pode diminuir o problema de censura dos dados (ARNASON, 1999; URIOSTE; MISZTAL; BERTRAND, 2007a; HOU et al., 2009).

A característica reprodutiva contínua (y_1) é assumida possuir modelo amostral Gaussiano e a relação entre o preditor linear (η_1) e a esperança de y_1 é determinada

pela função de ligação identidade. Por outro lado, para característica binária indicadora de censura (y_2) a distribuição binomial é utilizada e a função de ligação comumente empregada em melhoramento animal é a probit, outra alternativa possível seria a adoção da função de ligação logit. A associação entre preditores lineares e esperança de y_1 e y_2 pode ser descrita como a seguir:

$$E[Y_1 | \eta_1] = \eta_1 = X_1\beta_1 + Z_1u_1 \quad (1.10)$$

$$E[Y_2 | \eta_2] = \Phi[-(t - \eta_2)] = \Phi[-(t - X_2\beta_2 + Z_2u_2)] \quad (1.11)$$

em que: $\Phi(\cdot)$ é a distribuição normal padrão acumulada; e t é o limiar, o qual determina a expressão da característica categórica, pois $y_{2i} = 1$ se $\eta_{2i} > t$ e $y_{2i} = 0$ se $\eta_{2i} \leq t$; β_1 e β_2 são os vetores de efeitos fixos; u_1 e u_2 são vetores dos efeitos genéticos; X_1 , X_2 , Z_1 e Z_2 são matrizes de incidência dos efeitos fixos e aleatórios genéticos aditivos.

Quando o limiar é assumido ser igual a zero, a expressão (1.11) é representada por $E[Y_2 | \eta_2] = \Phi(\eta_2) = \Phi(X_2\beta_2 + Z_2u_2)$. A associação entre preditor linear e esperança realizada pela função de ligação pode ser ilustrada utilizando a variável indicadora binária. A esperança de uma variável aleatória que segue uma distribuição binomial é expressa por p , a probabilidade da matriz ser não censurada, e este parâmetro pode assumir valores entre 0 e 1. Por outro lado, o espaço paramétrico do preditor linear pode variar de $-\infty$ a ∞ , pois nenhuma restrição quanto aos valores de β_2 e u_2 são postulados. A predisposição de censura, a qual é representada por η_2 , é assumida possuir densidade normal e a probabilidade de sucesso, $P(Y = 1 | t, \eta_2)$, é obtida por meio da função de ligação probit, $\Phi[-(t - \eta_2)]$ ou $\Phi(\eta_2)$ quando $t = 0$. Logo, a discrepância entre os espaços paramétricos da esperança e do preditor linear, da variável binária, é solucionado a partir da utilização da função de ligação.

Os estágios hierárquicos do modelo animal linear-limiar apresentam distintos componentes, devido a uma maior complexidade, em relação ao modelo Bayesiano unicaracterístico, embora o número de estágios sejam semelhantes. Além das informações fenotípicas observadas, $y' = (y'_1, y'_2)$, o modelo amostral compreenderá a inclusão da predisposição à censura (η_2), a qual pode ser amostrada *a posteriori* pela aplicação da metodologia de aumento de dados (TANER; WONG, 1987; ALBERT; CHIB, 1993). Portanto, o modelo amostral pode ser representado como a seguir (SORENSEN; GIANOLA, 2002):

$$\begin{aligned} p(y_1, y_2 | \Omega, \eta_2) &= p(y_2 | y_1, \Omega, \eta_2) p(y_1 | \Omega, \eta_2) p(\eta_2 | \Omega) \\ p(y_1, y_2 | \Omega, \eta_2) &\propto p(y_1, \eta_2 | \Omega) p(y_2 | \Omega, \eta_2) \end{aligned} \quad (1.12)$$

$$\mathbf{y}_1, \boldsymbol{\eta}_2 | \boldsymbol{\Omega} \sim N \left(\begin{bmatrix} X_1 \boldsymbol{\beta}_1 + Z_1 \mathbf{u}_1 \\ X_2 \boldsymbol{\beta}_2 + Z_2 \mathbf{u}_2 \end{bmatrix}, R_0 \otimes I \right) \quad (1.13)$$

em que: $\boldsymbol{\Omega} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, R_0)$; R_0 representa a matriz de variância e covariância residual.

No segundo estágio, as densidades *a priori* para os parâmetros de locação e escala do modelo amostral assumidas são: i) $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ distribuições uniforme; ii) $\mathbf{u}_1$ e $\mathbf{u}_2$, densidade normal multivariada centrada em 0 e dispersão dada por $G_0 \otimes A$, em que G_0 é a matriz de variância e covariância genética entre as características reprodutiva contínua e indicadora de censura; iii) Para as variâncias e covariâncias residuais (R_0), a distribuição adotada é a Wishart invertida. No terceiro estágio, é feita a suposição *a priori* da densidade Wishart invertida para as variâncias e covariâncias genéticas (G_0).

A distribuição completa *a posteriori* é obtida pelo produtos das densidades descritas nos três estágios. Para inferência, os métodos de MCMC são empregados e as distribuições condicionais completas *a posteriori* são necessárias para implementá-la. Os parâmetros de locação, $\boldsymbol{\theta}' = [\boldsymbol{\beta}'_1 \ \boldsymbol{\beta}'_2 \ \mathbf{u}'_1 \ \mathbf{u}'_2]$, a densidade condicional completa é normal multivariada (SORENSEN; GIANOLA, 2002):

$$\boldsymbol{\theta} | \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\eta}_2, G_0, R_0 \sim N(\hat{\boldsymbol{\theta}}, C^{-1}) \quad (1.14)$$

em que: $\hat{\boldsymbol{\theta}}$ é a solução para o sistema de equações $C\boldsymbol{\theta} = \mathbf{b}$; $C = W'R^{-1}W + \Sigma^{-1}$; $W = \begin{bmatrix} X_1 & 0 & Z_1 & 0 \\ 0 & X_2 & 0 & Z_2 \end{bmatrix}$; $R = R_0 \otimes I$; $\Sigma^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & G_0^{-1} \otimes A^{-1} \end{bmatrix}$; $\mathbf{b} = W'R^{-1}\mathbf{z}$; $\mathbf{z}' = [\mathbf{y}'_1 \ \boldsymbol{\eta}'_2]$.

A densidade condicional completa *a posteriori* para a predisposição de censura é dada por uma distribuição normal, a qual possui média e variância para a amostragem de η_{2i} representadas como a seguir (SORENSEN; GIANOLA, 2002):

$$E[\eta_{2i} | \mathbf{y}_1, \boldsymbol{\Omega}, G_0] = \mathbf{x}'_{2i} \boldsymbol{\beta}_2 + \mathbf{z}'_{2i} \mathbf{u}_2 + \frac{\sigma_{e1,2}}{\sigma_{e1}^2} \left(y_{1i} - \mathbf{x}'_{1i} \boldsymbol{\beta}_1 - \mathbf{z}'_{1i} \mathbf{u}_1 \right) \quad (1.15)$$

$$Var[\eta_{2i} | \mathbf{y}_1, \boldsymbol{\Omega}, G_0] = \sigma_{e2}^2 \left[1 - \frac{(\sigma_{21,2})^2}{\sigma_{e1}^2 \sigma_{e2}^2} \right] \quad (1.16)$$

As matrizes de covariância genética e residual possuem densidades condicionais *a posteriori* Wishart invertida:

$$G_0|y_1, y_2, \eta_2, \Omega, R_0 \sim WI \left([S_u + V_u^{-1}]^{-1}, v_u + q \right) \quad (1.17)$$

$$R_0|y_1, y_2, \eta_2, \Omega, G_0 \sim WI \left([S_e + V_e^{-1}]^{-1}, v_e + n \right) \quad (1.18)$$

em que: $S_u = \begin{bmatrix} u_1' A^{-1} u_1 & u_1' A^{-1} u_2 \\ u_2' A^{-1} u_1 & u_2' A^{-1} u_2 \end{bmatrix}$; $S_e = \begin{bmatrix} e_1' e_1 & e_1' e_2 \\ e_2' e_1 & e_2' e_2 \end{bmatrix}$; $e_1 = y_1 - X_1 \beta_1 - Z_1 u_1$ e $e_2 = \eta_2 - X_2 \beta_2 - Z_2 u_2$; S_u , S_e , v_u e v_e são hiperparâmetros; q e n representam, respectivamente, o número de animais na matriz A e número de observações.

Logo, a amostragem de Gibbs é baseada na geração de amostras a partir de (1.14), densidades normal com média e variância denotadas por (1.15) e (1.16), (1.17) e (1.18).

1.3.4 Análise de Sobrevivência

Em análise de sobrevivência uma série de procedimentos estatísticos estão disponíveis para o estudo de variáveis que representam o tempo de ocorrência de um evento de interesse, como por exemplo, tempo até o primeiro parto (IPP), tempo até a prenhez em uma estação de monta (DPP) ou tempo em que uma matriz permanece produtiva em um rebanho (LONG). Logo, a variável em estudo deve ser estritamente contínua e um par de variáveis aleatórias é observado (Y, δ) , os quais representam, respectivamente, o tempo até o evento de interesse e uma variável indicadora de censura. Os registros observados podem ser representados como $y_i = \min(t_i, c_i)$, em que t_i e c_i são o tempo até a expressão fenotípica e o tempo de censura, respectivamente, e uma observação é não censurada ($\delta_i = 1$) se $t_i \leq c_i$ e censurada ($\delta_i = 0$) quando $t_i > c_i$.

Na análise de sobrevivência, duas funções derivadas da distribuição acumulada da característica em estudo ($F(t)$) são importantes: as funções de sobrevivência ($S(t)$) e taxa de falha ($h(t)$). A função de sobrevivência representa a probabilidade de um animal possuir o fenótipo acima de um tempo t , ou seja, $P(T > t)$. Por outro lado, a função taxa de falha expressa o potencial instantâneo da expressão fenotípica ocorrer dado que o fenótipo é superior a um tempo t específico, $P(t \leq T < t + \Delta t | T \geq t)$. Logo, estas funções podem ser expressas como a seguir:

$$S(t) = 1 - F(t) \quad (1.19)$$

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$h(t) = \frac{f(t)}{S(t)} \quad (1.20)$$

em que, $f(t)$ representa a função densidade de probabilidade da variável analisada.

As funções taxa de falha e de sobrevivência podem possuir formas desconhecidas, como na abordagem utilizando o modelo de Cox (COX, 1972; SMITH; QUASS, 1984), ou modeladas por meio de funções paramétricas conhecidas, como por exemplo, exponencial, Weibull e lognormal (DUCROCQ et al., 1988; DUCROCQ; CASELLA, 1996).

No modelo amostral, a contribuição de observações não censuradas e censuradas é realizada de forma distinta, pois indivíduos cujos fenótipos são observados proveem $f(t)$, por outro lado, animais censurados contribuem com $S(t)$. Portanto, como descrito por Ducrocq e Casella (1996), a densidade amostral quando a distribuição Weibull é assumida pode ser representada como:

$$\begin{aligned}
 p(\mathbf{y}|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) &= \prod_{i=1}^n [f(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2)]^{\delta_i} [S(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2)]^{1-\delta_i} \\
 p(\mathbf{y}|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) &= \prod_{i=1}^n [h(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) S(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2)]^{\delta_i} [S(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2)]^{1-\delta_i} \\
 p(\mathbf{y}|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) &= \prod_{i=1}^n [h(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2)]^{\delta_i} S(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) \quad (1.21) \\
 h(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) &= \rho y_i^{\rho-1} \exp \left[\rho \log(\lambda) + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u} \right] \\
 S(y_i|\lambda, \rho, \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2) &= \exp \left[-y_i^\rho \exp \left(\rho \log(\lambda) + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u} \right) \right]
 \end{aligned}$$

em que, λ e ρ são os parâmetros de escala e forma da distribuição Weibull.

Ducrocq e Casella (1996) adotaram como distribuições *a priori* dos efeitos genéticos aditivos uma distribuição normal multivariada, centrada em zero e matriz de variância e covariâncias denotada por $A\sigma_u^2$, e para λ , ρ , $\boldsymbol{\beta}$ e σ_u^2 densidades uniforme.

A distribuição completa *a posteriori* é obtida pelo produto das densidades condicional de $\mathbf{y}$ e *a priori* dos parâmetros do modelo. Ducrocq e Casella (1996), descreveram a utilização dos métodos quasi-Newton e integração de Laplace na inferência *a posteriori* dos parâmetros $\boldsymbol{\Theta}' = \left[\tilde{\lambda} \quad \tilde{\rho} \quad \tilde{\boldsymbol{\beta}} \quad \tilde{\mathbf{u}} \right]'$ e variância genética aditiva, respectivamente. Logo, estes métodos foram empregados nas seguintes distribuições condicionais *a posteriori*:

$$\tilde{\boldsymbol{\Theta}} = \arg_{\boldsymbol{\Theta}} \max p(\boldsymbol{\Theta}|\mathbf{y}, \sigma_u^2) \quad (1.22)$$

$$p(\sigma_u^2|\mathbf{y}) = \int \int p(\boldsymbol{\Theta}, \sigma_u^2|\mathbf{y}) d\boldsymbol{\Theta} \quad (1.23)$$

1.4 OBJETIVOS

Objetivo Geral

- Estudar as características reprodutivas de idade ao primeiro parto, dias para o parto e longevidade, de fêmeas da raça Nelore, utilizando modelos que tratam as expressões fenotípicas destas características como não censuradas e censuradas, com o objetivo de encontrar modelos adequados para avaliação genética dos animais para estas características.

Objetivos Específicos

- Estimar parâmetros genéticos para as características reprodutivas utilizando diferentes modelos de análise;
- Predizer os valores genéticos e realizar comparações na classificação dos animais nos distintos modelos estatísticos;
- Determinar o modelo mais adequado para a avaliação genética destas características reprodutivas.

CAPÍTULO 2 - ANÁLISE DE IDADE AO PRIMEIRO PARTO CONSIDERANDO A INCLUSÃO DE DADOS CENSURADOS

2.1 INTRODUÇÃO

A eficiência reprodutiva apresenta-se como um dos principais objetivos de seleção na bovinocultura de corte devido ao seu grande impacto econômico no sistema produtivo. Entre as principais vantagens do aumento da precocidade reprodutiva das novilhas estão a redução do tempo para o retorno do investimento, aumento da vida produtiva da matriz e o maior número de bezerros produzidos (SHORT et al., 1994). Um dos critérios de seleção mais utilizados visando o progresso genético da eficiência reprodutiva em gado de corte é a idade ao primeiro parto (IPP).

Na avaliação genética da IPP, uma parte das fêmeas, normalmente, não apresentam fenótipos devido a falhas reprodutivas, descarte voluntário e involuntário anterior à estação reprodutiva e, em alguns casos, por registros inconsistentes. Estas observações parciais ou desconhecidas do fenótipo dos animais são denominadas censuras. De acordo com Guo, Gianola e Short (2001), a inclusão adequada destas observações nas avaliações genéticas poderiam evitar distorções nas inferências e melhoria na predição dos valores genéticos.

Alguns métodos tem sido utilizados no tratamento das censuras nas avaliações genéticas de características reprodutivas. Uma alternativa é considerar a informação censurada como verdadeira após a adição de uma penalidade (JOHNSTON; BUNTER, 1996; DONOGHUE; REKAYA; BERTRAND, 2004a; DONOGHUE; REKAYA; BERTRAND, 2004b; URIOSTE; MISZTAL; BERTRAND, 2007b; HOU et al., 2009). Outra abordagem é amostrar valores para os registros censurados a partir de uma distribuição normal truncada (SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001; KORSGAARD et al., 2003; DONOGHUE; REKAYA; BERTRAND, 2004a; DONOGHUE; REKAYA; BERTRAND, 2004b; ARANGO et al., 2005; CHANG et al., 2006; GONZÁLEZ-RECIO et al., 2006; HOU et al., 2009). Outra alternativa utilizando modelos lineares é o uso do modelo linear-limiar, que inclui uma característica binária indicadora do estado de censura a uma característica reprodutiva contínua (ARNASON, 1999; DONOGHUE et al., 2004c; URIOSTE; MISZTAL; BERTRAND, 2007b; URIOSTE; MISZTAL; BERTRAND, 2007a; HOU et al., 2009). Além destes, modelos de análise de sobrevivência têm sido aplicados para avaliar características relacionadas à reprodução (DUCROCQ, 1993; DUCROCQ; CASELLA, 1996; SCHNEIDER et al., 2005; PEREIRA et al., 2005; GONZÁLEZ-RECIO et al., 2006; SCHNEIDER et al., 2006; PEREIRA et al., 2007; HOU et al., 2009). Entretanto no Brasil, estas abordagens ainda têm sido pouco empregadas (PEREIRA et al., 2005;

PEREIRA et al., 2007; PEREIRA, 2008; CARDOSO et al., 2009; SILVA et al., 2010; CAETANO, 2011).

Logo, o objetivo deste trabalho foi comparar estimativas de parâmetros genéticos e habilidade de predição de modelos nos quais a expressão fenotípica da característica idade ao primeiro parto foi considerada como não censurada ou censurada.

2.2 MATERIAL E MÉTODOS

Dados

O banco de dados utilizado é oriundo do arquivo zootécnico de 171 fazendas, localizadas no sudeste e centro-oeste do Brasil, que se dedicam à bovinocultura de corte e constituem a base de dados Aliança Nelore, a qual agrega informações de diferentes programas de melhoramento genético da raça Nelore.

O manejo reprodutivo adotado pelas fazendas consiste na realização de duas estações reprodutivas: a) Estação de outono, na qual novilhas com idade em torno de 15 a 17 meses são desafiadas por 60 dias à reprodução, a fim de identificar animais sexualmente precoces; b) Estação de verão, onde todas as fêmeas do rebanho participam, sendo as novilhas com idade em torno de 24 meses e multíparas expostas, respectivamente, por 60 e 90 dias. As novilhas que não concebem na primeira estação têm uma segunda oportunidade na estação de verão, e as fêmeas, novilhas e matrizes, que não concebem nesta estação são descartadas.

Todas as fêmeas, nascidas entre 1993 e 2007, que possuíam registros ao sobreano, pais conhecidos e que tiveram a oportunidade de estarem na estação de monta de outono foram consideradas na consistência dos dados (71.398 animais). Matrizes que apresentaram idade ao primeiro parto superior a 40 meses foram excluídas (7.297), visto que, nas fazendas, após a estação de verão as novilhas que não emprenham são descartadas. Logo, essa prática foi adotada para se evitar possíveis erros de registros.

Os grupos de contemporâneos (GC) foram definidos a partir da combinação de safra, mês e fazenda de nascimento, fazenda e grupo de manejo à desmama e fazenda e grupo de manejo ao sobreano. Os GC com mais de dez animais com informação fenotípica permaneceram no banco de dados para análise.

Os valores dos registros censurados foram definidos utilizando duas estratégias distintas: a) Considerando como informação fenotípica censurada o maior valor de idade ao primeiro parto observado em cada grupo de contemporâneos; b) Adicionando

ao maior valor de IPP do grupo de contemporâneos uma penalização de 21 dias, a qual corresponderia a um ciclo estral (JOHNSTON; BUNTER, 1996). Uma descrição completa dos bancos de dados é apresentada na Tabela 2.1.

Tabela 2.1 – Descrição da estrutura do banco de dados da característica idade ao primeiro parto quanto ao número de registros, estatística descritiva, informações de grupos de contemporâneos e pedigree

Registros	
Censurados	13.034
Não censurados	22.045
	35.079
Estatística descritiva dos registros não censurados (meses)	
Mínimo	22,68
Máximo	39,98
Média	34,23
Desvio-padrão	3,56
Estatística descritiva dos registros censurados (meses) ^a	
Mínimo	25,81
Máximo	39,98
Média	37,25
Desvio-padrão	2,20
Grupos de Contemporâneos	
Número de grupos	1.107
Mínimo de animais não censurados	10
Máximo de animais não censurados	98
Mínimo de animais censurados	0
Máximo de animais censurados	74
Pedigree	
Número de gerações	5
Número de touros	1.160
Número total de animais	73.874

^a A estatística descritiva, quando a penalização é adotada, é obtida somando aos valores da tabela 21 dias, exceto para o desvio-padrão que permanece o mesmo

A idade ao primeiro parto foi avaliada utilizando os seguintes modelos:

Modelo linear Gaussiano (ML)

Neste modelo, todas as observações censuradas foram excluídas do banco de dados. Sendo assim, o modelo pode ser descrito com se segue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2.1)$$

em que: $\mathbf{y}$ é o vetor de observações de idade ao primeiro parto; $\boldsymbol{\beta}$ é o vetor de efeitos fixos de grupos de contemporâneos; $\mathbf{u}$ é o vetor dos valores genéticos dos animais, com a pressuposição de que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, sendo $\mathbf{A}$ e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente; $\mathbf{e}$ o vetor de resíduos, os quais são assumidos normalidade e homocedasticidade de variância, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, em que, $\mathbf{I}$ e σ_e^2 são a matriz identidade e variância residual, respectivamente.

Modelo linear Gaussiano com penalização (MLP)

Novilhas que não possuíam informação fenotípica receberam como informação censurada o maior valor do grupo de contemporâneos adicionado a uma penalização de 21 dias. Portanto, o modelo MLP é similar ao descrito em (2.1) com a distinção em relação ao vetor de observações $\mathbf{y}$, que possui informações de fêmeas não censuradas e censuradas.

Modelo Linear Gaussiano com censura (MLC)

O modelo linear Gaussiano com censura é similar ao modelo anteriormente descrito, entretanto a metodologia de aumento de dados foi utilizada no tratamento dos registros censurados (TANER; WONG, 1987; SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001; HOU et al., 2009). O modelo pode ser representado como se segue:

$$\begin{bmatrix} \mathbf{y}_{nc} \\ \mathbf{y}_c \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2.2)$$

em que: $\mathbf{y}_{nc}$ e $\mathbf{y}_c$ representam, respectivamente, as observações não censuradas e os registros aumentados para as fêmeas censuradas.

As observações aumentadas do vetor y_c são obtidas em cada passo da amostragem de gibbs de uma distribuição normal truncada com densidade dada por (SORENSEN; GIANOLA; KORSGAARD, 1998):

$$p(y_{c_i} | \beta, u, \sigma_u^2, \sigma_e^2, y_{nc}; y_{c_i} > c_i) = \frac{\phi_{y_{c_i}}(x'_i \beta + z'_i u_i, \sigma_e^2)}{1 - \Phi\left(\frac{c_i - x'_i \beta + z'_i u_i}{\sigma_e}\right)} \quad (2.3)$$

em que: $\phi(\cdot)$ e $\Phi(\cdot)$ são, respectivamente, as funções densidade e acumulada da distribuição normal; c_i é o ponto de truncamento de cada animal, ou seja, o maior valor de idade ao primeiro parto do grupo de contemporâneos ao qual a novilha censurada pertence.

Modelo Linear-Limiar (MLL)

Nesta abordagem uma variável binária indicadora de censura é definida, sendo 1 para matrizes com informação fenotípica e 0 caso contrário, e modelada como característica correlacionada à idade ao primeiro parto. Entretanto, as novilhas sem informações fenotípicas são incluídas na análise considerando seus registros de IPP como informações perdidas (ARNASON, 1999; URIOSTE; MISZTAL; BERTRAND, 2007b). Logo, o modelo pode ser representado como a seguir (HOU et al., 2009):

$$\begin{bmatrix} y \\ l \end{bmatrix} = \begin{bmatrix} X_y & 0 \\ 0 & X_l \end{bmatrix} \begin{bmatrix} \beta_y \\ \beta_l \end{bmatrix} + \begin{bmatrix} Z_y & 0 \\ 0 & Z_l \end{bmatrix} \begin{bmatrix} u_y \\ u_l \end{bmatrix} + \begin{bmatrix} e_y \\ e_l \end{bmatrix} \quad (2.4)$$

em que: $y' = \begin{bmatrix} y_{nc} & 0 \end{bmatrix}$, representando os registros das novilhas que possuem informação fenotípica e zeros para as fêmeas sem observações; l é o vetor de predisposição do estado de censura dos registros.

As seguintes suposições do modelo linear-limiar são adotadas:

$$\begin{bmatrix} u_y \\ u_l \end{bmatrix} \sim N(0, A \otimes G)$$

$$\begin{bmatrix} e_y \\ e_l \end{bmatrix} \sim N(0, I \otimes R) \quad (2.5)$$

em que: G e R são as matrizes de variâncias e covariâncias genéticas e residuais, respectivamente.

Modelo Linear-Limiar com penalização (MLLP)

Este modelo apresenta as características descritas em (2.4) e (2.5), entretanto, em lugar de zero, os maiores valores de idade ao primeiro parto dentro de cada grupo de contemporâneos foram atribuídos aos animais sem informação fenotípica.

Modelo de Riscos Proporcionais Weibull (MW)

O seguinte modelo misto de análise de sobrevivência foi utilizado (DUCROCQ; CASELLA, 1996):

$$h(t) = \lambda_0(t) \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \mathbf{Z}s) \quad (2.6)$$

em que: $h(t)$ é a função taxa de falha; $\lambda_0(t)$ a função taxa de falha base Weibull, $\lambda_0 = \lambda\rho(\lambda t)^{\rho-1}$ sendo λ e ρ os parâmetros de escala e forma, respectivamente; $\boldsymbol{\beta}$ vetor dos efeitos tempo-independente de grupos de contemporâneos; $\boldsymbol{\theta}$ efeito tempo-dependente de período, com três períodos os quais serão definidos posteriormente; s efeito aleatório tempo-independente de touro, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $\mathbf{A}\sigma_s^2$, sendo $\mathbf{A}$ e σ_s^2 a matriz de numerador de parentesco entre os reprodutores e a variância de touro, respectivamente.

O modelo de touro, descrito em (2.6), foi utilizado para obter a estimativa da variância de touro e, este componente foi multiplicado por quatro para se obter a estimativa da variância genética aditiva. Posteriormente, este componente de variância foi considerado como conhecido para a predição dos valores genéticos utilizando o modelo animal, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $\mathbf{A}\sigma_u^2$, sendo $\mathbf{A}$ e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente

O estudo da adequabilidade do modelo Weibull como distribuição da taxa de falha base foi realizado graficamente plotando o logaritmo natural ($\ln$) da idade ao primeiro parto e o $\ln \left[-\ln \left(\hat{S}(IPP) \right) \right]$, em que $\hat{S}(IPP)$ é a estimativa da função de sobrevivência obtida pelo método não-paramétrico Kaplan-Meier (Figura 2.1). Para a escolha do modelo Weibull, a relação entre $\ln(IPP)$ vs $\ln \left[-\ln \left(\hat{S}(IPP) \right) \right]$ deveria ser linear, comportamento não observado na Figura 2.1. Entretanto, é possível distinguir três intervalos onde a linearidade pode ser considerada e, em cada período a modelagem utilizando a distribuição Weibull é aceitável. Portanto, a variável tempo-dependente período foi definida e os seus limites foram inicialmente determinados

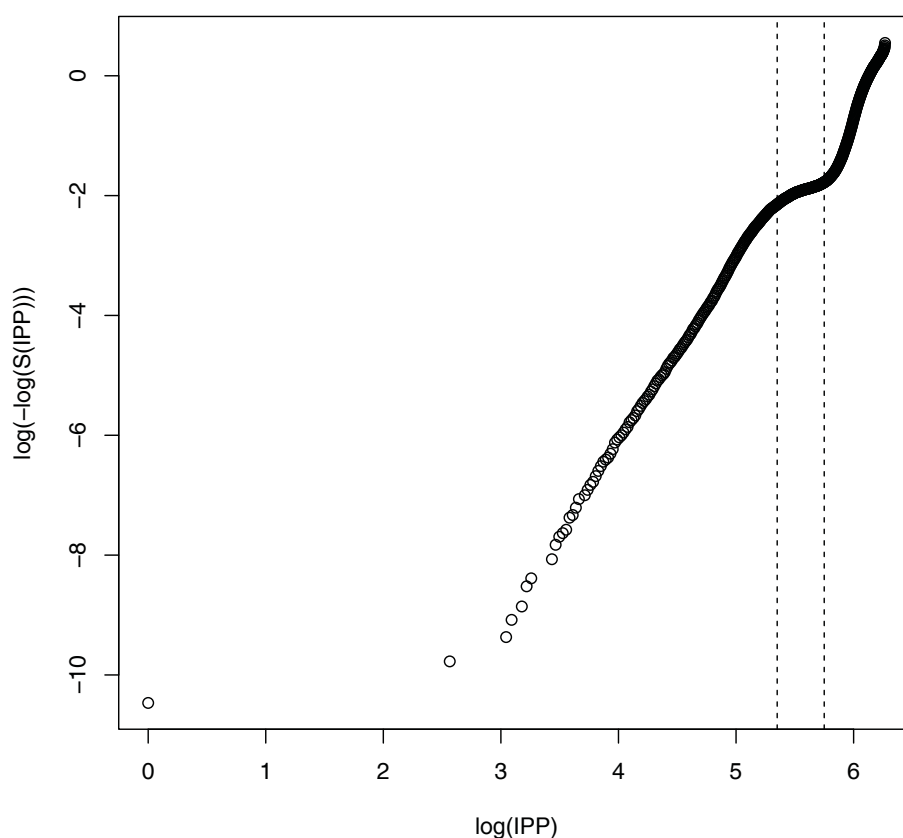


Figura 2.1 – Verificação gráfica do ajuste da distribuição Weibull como taxa de falha base na avaliação da idade ao primeiro parto (IPP). A estimativa da função de sobrevivência, $S(IPP)$, foi obtida utilizando o estimador de Kaplan-Meier e as linhas verticais indicam o limite dos períodos utilizados (0 a 29,61; 29,62 a 33,01; acima de 33,01 meses)

visualmente e, posteriormente, definidos comparando-se valores de verossimilhança quando pequenas mudanças foram realizadas (PEREIRA et al., 2005; PEREIRA et al., 2007). Logo, os períodos utilizados foram, respectivamente, de 0 a 29,61, 29,62 a 33,01 e acima de 33,01 meses.

Os componentes de variância e os valores genéticos foram estimados e preditos utilizando os programas GIBBS1F90, GIBBS2CEN e THRGIBBS1F90, da família BGF90 (MISZTAL et al., 2002), respectivamente, para os modelos ML e MLP, MLC, MLL e MLLP. Já para o modelo MW, o programa Survival Kit v6.0 (DUCROCQ; SOLKNER; MÉSZÁROS, 2010) foi empregado. Nas análises em que a amostragem de Gibbs foi utilizada, três cadeias de 150.000 iterações com distintos valores iniciais de componentes de variância, os quais representam herdabilidade iguais a 0,10, 0,40 e 0,80. Inspeção gráfica e os testes de Brooks, Gelman e Rubin (GELMAN; RUBIN,

1992) e Heldelberger e Welch (HELDELBERGER; WELCH, 1983) utilizando o pacote *coda* (PLUMMER et al., 2006) do programa R (R Core Team, 2013) foram realizados, sendo obtidos resultados positivos de convergência das análises. Para inferência, apenas uma das cadeias foi utilizada com descarte amostral de 15.000 amostras.

Comparação dos modelos

Correlações de Pearson e a porcentagem de coincidência na classificação dos 10% melhores touros (TOP10) e reprodutores com mais de 10 progênes foram calculadas para avaliar o efeito de cada metodologia na predição dos valores genéticos utilizando o banco de dados completo.

Além disso, a habilidade de predição dos modelos foi observada utilizando um esquema de validação com 10 repetições, em que, 80% e 20% das fêmeas com informações fenotípicas pertenciam aos arquivos de treinamento e validação, respectivamente. A alocação dos animais foi realizada aleatoriamente respeitando a restrição de pelo menos 3 animais por GC no treinamento e que cada touro possuísse filhas não censuradas e censuradas em cada partição do banco de dados. Os componentes de variância, estimados na base de dados completa para cada modelo, foram utilizados como verdadeiros na predição dos valores genéticos dos animais do treinamento.

A estatística χ^2 foi utilizada para a comparação dos modelos em termos de eventos reprodutivos esperados e observados das filhas de touros que possuíam mais de 10 progênes não censuradas (356 reprodutores). Para tanto, a idade ao primeiro parto foi discretizada em duas categorias: 1) fêmeas com informações fenotípicas de IPP; 2) matrizes censuradas. Após obter as diferenças esperadas nas progênes (DEP) no treinamento, probabilidades do touro possuir filhas em cada uma das duas classes foi obtida utilizando uma regressão logística dos eventos reprodutivos observados no treinamento sobre as DEP dos reprodutores. As probabilidades estimadas foram utilizadas para prever a frequência esperada das progênes de cada touro em cada classe de IPP no banco de dados de validação. Logo, a soma da estatística χ^2 sobre os touros foi calculada como se segue:

$$\chi^2 = \sum_{i=1}^{nt} \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.7)$$

em que: O_{ij} e E_{ij} são o número de fenótipos observados e esperados, respectivamente, das filhas do touro i serem da classe de IPP j (1 ou 2); nt é o número total de touros, variando de 1 até 356 animais.

2.3 RESULTADOS E DISCUSSÃO

As estimativas de herdabilidade da idade ao primeiro parto de 0,18 e 0,17 (Tabela 2.2), utilizando os modelos ML e MLL, estão de acordo com trabalhos na raça Nelore, em que, este parâmetro genético apresentou estimativas variando entre 0,07 e 0,27 (MERCADANTE; LÔBO; OLIVEIRA, 2000; PEREIRA; ELER; FERRAZ, 2000; DIAS; EL FARO; ALBUQUERQUE, 2004; AZEVÊDO et al., 2006; BOLIGON; RORATO; ALBUQUERQUE, 2007; BOLIGON et al., 2008; SILVA et al., 2010). Entretanto, Faria et al. (2007) obtiveram estimativas de 0,40 e 0,41 em análises bi-características com perímetro escrotal medido aos 365 e 450 dias. Boligon et al. (2008), salientaram que diferentes populações e modelos causaram disparidades nas estimativas de parâmetros genéticos da IPP. Isto ocorre porque a idade ao primeiro parto é dependente do manejo reprodutivo e da idade em que as fêmeas são desafiadas à concepção.

Tabela 2.2 – Média e desvio padrão, entre parênteses, a posteriori para a variância genética ($\hat{\sigma}_u^2$), variância residual ($\hat{\sigma}_e^2$) e herdabilidade ($\hat{h}^2$) da idade ao primeiro parto, obtidas utilizando diferentes modelos de avaliação genética

Parâmetro	Modelo					
	ML	MLP	MLC	MLL	MLLP	MW
$\hat{\sigma}_u^2$	1,361 (0,220)	1,027 (0,181)	0,897 (0,157)	1,290 (0,126)	1,036 (0,155)	0,123 (0,012)
$\hat{\sigma}_e^2$	6,072 (0,173)	7,590 (0,148)	6,664 (0,127)	6,425 (0,926)	6,556 (0,130)	- -
$\hat{h}^2$	0,18 (0,025)	0,12 (0,014)	0,12 (0,018)	0,17 (0,019)	0,14 (0,011)	0,07 (0,020)

^a ML: modelo linear Gaussiano;
MLP: modelo linear Gaussiano com penalização;
MLC: modelo linear Gaussiano com censura;
MLL: modelo linear-limiar;
MLLP: modelo linear-limiar com penalização;
MW: modelo de riscos proporcionais Weibull

Os modelos lineares MLP, MLC e MLLP apresentaram estimativas de herdabilidade semelhantes (Tabela 2.2). Resultados similares foram reportados por Silva et al. (2010), os quais obtiveram herdabilidades de IPP iguais a 0,07 para MLP e MLC. Da mesma forma, Donoghue, Rekaya e Bertrand (2004a) e Donoghue, Rekaya e Bertrand (2004b), trabalhando com simulação e dados reais para a característica dias para o parto, relataram resultados semelhantes entre as estimativas de herdabilidade obtidas utilizando os modelos MLP e MLC. Em gado Holandês, Hou et al. (2009), avaliando

período de serviço e dias entre o parto e a primeira inseminação, obtiveram herdabilidade de 0,07 e 0,10, 0,07 e 0,11, 0,07 e 0,11, respectivamente, para os modelos MLP, MLC e MLLP.

A estimativa de herdabilidade para IPP obtida na escala logarítmica utilizando o modelo de riscos proporcionais Weibull de 0,07 (Tabela 2.2) foi inferior a reportada por Pereira et al. (2005) e Pereira et al. (2007), sendo esta igual a 0,51. Entretanto, estes autores avaliaram dados de idade à concepção de novilhas desafiadas à reprodução em torno dos 14 meses de idade, portanto maior variabilidade genética quando a precocidade é avaliada nesta fase é esperada (ELER et al., 2002; ELER et al., 2004; PEREIRA et al., 2007).

A magnitude da estimativa de ρ é um indicativo da forma que a distribuição taxa de falha Weibull possui. Portanto, quando temos $\rho < 1$, $\rho = 1$ ou $\rho > 1$ é esperado que a taxa de expressão da característica em análise diminua, mantenha-se constante ou aumente com o passar do tempo, respectivamente. Logo, $\hat{\rho}$ com o valor estimado de 4,86 ($\pm 0,051$) representa o que esperamos para as fêmeas, pois com o aumento da idade as novilhas apresentam um maior desenvolvimento fisiológico e, consequentemente, uma maior chance de engravidar na estação reprodutiva.

As correlações de Pearson entre os valores genéticos preditos utilizando os modelos Gaussianos (ML, MLP e MLC) e linear-limiar (MLL e MLLP) foram de alta magnitude nas duas situações delineadas, para todos os touros e touros com mais de 10 progênie (Tabela 2.3). Logo, as predições obtidas com estes modelos apresentam uma forte associação. Hou et al. (2009), estudando dias entre o parto e a primeira inseminação e a característica período de serviço em gado Holandês, observaram correlações próximo da unidade entre os modelos MLP, MLC e MLLP e, valores em torno de -0,90 entre estes modelos e MW. Donoghue, Rekaya e Bertrand (2004b), trabalhando com a característica dias para o parto em gado de corte, obtiveram correlações entre valores genéticos preditos pelos modelos MLP e MLC semelhantes as observadas neste estudo, entretanto, reportaram associações inferiores para os modelos ML-MLP (0,77 e 0,80) e ML-MLC (0,81 e 0,84).

Uma menor associação entre as predições dos modelos ML, MLL e MLLP com o modelo de riscos proporcionais Weibull foi observada, correlação variando de -0,74 a -0,88. Por outro lado, os valores obtidos entre MW-MLP e MW-MLC foram superiores (Tabela 2.3). Investigando idade a primeira concepção em fêmeas da raça Nelore, Pereira et al. (2007) relataram associação superior à obtida neste trabalho para os modelos MLC e MW, sendo estas de -0,99 e -0,92 utilizando diferença esperada na progênie de todos os touros e apenas touros classificados entre os 10% superio-

Tabela 2.3 – Correlações de Pearson entre os valores genéticos preditos de todos os touros (acima da diagonal) e de touros com mais de 10 progênie não censuradas (abaixo da diagonal) utilizando distintos modelos de avaliação genética para idade ao primeiro parto^a

Modelos ^b	ML	MLP	MLC	MLL	MLLP	MW
ML		0,92	0,94	0,98	0,94	-0,74
MLP	0,93		0,99	0,97	0,99	-0,92
MLC	0,95	0,99		0,97	0,99	-0,90
MLL	0,99	0,97	0,97		0,97	-0,84
MLLP	0,95	0,99	0,99	0,97		-0,89
MW	-0,76	-0,92	-0,90	-0,82	-0,89	

^a 1.160 e 356 touros;

^b ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull.

res, respectivamente. É importante ressaltar que as DEP nos modelos lineares e de análise de sobrevivência devem ser interpretadas de maneira distintas, sendo estas, respectivamente, a diminuição que se espera na idade ao primeiro parto das filhas de um reprodutor e a chance das progênie de um touro apresentarem idade ao primeiro parto inferiores. Por isso, a correlação entre as predições dos modelos lineares e o modelo de riscos proporcionais Weibull são negativas.

A porcentagem de coincidência na classificação dos 10% melhores touros foram semelhantes e de alta magnitude entre os modelos MLP, MLC, MLL e MLLP (Tabela 2.4). Estes resultados corroboram com os obtidos por Donoghue, Rekaya e Bertrand (2004b), os quais observaram grande correspondência na classificação de animais para MLP e MLC, assim como, Hou et al. (2009) para os modelos MLP, MLC e MLLP, estudando dias entre o parto e primeira inseminação e dias de serviço.

O TOP10 entre ML, MLP, MLC, MLL e MLLP variaram em torno de 90% e, destes quatro modelos com MW foram ao redor de 76%, considerando todos os touros ou apenas reprodutores com mais de 10 produtos. Logo, existem evidências de reclassificação de animais entre os modelos ML, MLP, MLC, MLL e MLLP e o MW. Entre os modelos lineares, pode-se destacar uma maior coincidência de classificação dentro dos grupos MLP-MLC-MLLP e ML-MLL. Mudança na classificação de animais entre os modelos ML-MLP e ML-MLC foram relatadas por Donoghue, Rekaya e Ber-

Tabela 2.4 – Porcentagem de coincidência na classificação dos 10% melhores touros (acima da diagonal) e touros com mais de 10 progênes não censuradas (abaixo da diagonal) utilizando distintos modelos de avaliação genética para idade ao primeiro parto^a

Modelos ^b	ML	MLP	MLC	MLL	MLLP	MW
ML		89,66	91,38	96,55	89,66	71,55
MLP	86,11		97,41	92,24	96,55	81,03
MLC	86,11	100,00		93,97	96,55	79,31
MLL	97,22	88,89	88,89		92,24	74,14
MLLP	86,11	94,44	94,44	88,89		78,45
MW	69,44	80,56	80,56	72,22	77,78	

^a 116 e 36 touros;

^b ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull.

trand (2004b) e Silva et al. (2010), trabalhando com dias para o parto na raça Angus e idade ao primeiro parto em Nelore, respectivamente. Entretanto, González-Recio et al. (2006) e Pereira et al. (2007) investigando, respectivamente, dias de serviço em gado Holandês e idade a primeira concepção em novilhas Nelore, reportaram elevada correspondência na classificação de touros para os modelos MLC e MW.

O modelo de riscos proporcionais Weibull apresentou uma maior habilidade de predição, com menor média de χ^2 e maior número de vezes em que possuiu o menor valor desta estatística (Tabela 2.5). Os modelos linear Gaussiano com penalização e linear-limiar obtiveram os menores valores de χ^2 em uma e duas repetições, respectivamente, e apresentaram média nas 10 repetições da validação cruzada que os classificaram como segundo e terceiro modelos. Já os modelos MLLP, ML e MLC não obtiveram o menor valor de χ^2 em nenhuma das repetições e apresentaram a quarta, quinta e sexta melhores médias de habilidade de predição. Resultados similares foram relatados por González-Recio et al. (2006), investigando dias de serviço em animais da raça Holandesa, os quais obtiveram melhor habilidade de predição para o modelo MW em relação a MLC e ML. Em gado de corte, Pereira et al. (2007), trabalhando com idade à primeira concepção em novilhas Nelore, não obtiveram diferenças em termos de resposta a seleção entre os modelos MW e MLC.

Os modelos lineares Gaussiano, Gaussiano com penalização, Gaussiano com censura, linear-limiar e linear-limiar com penalização apresentaram associação das

Tabela 2.5 – Média da estatística χ^2 , eficiência e o número de vezes em que os distintos modelos de avaliação genética da idade ao primeiro parto apresentaram o menor valor de χ^2 (NM χ^2), num esquema de validação cruzada com 10 repetições

Modelo ^a	Média χ^2	Eficiência ^b	NM χ^2
ML	715,429	1,034	0
MLP	698,433	1,009	1
MLC	759,869	1,098	0
MLL	705,183	1,019	2
MLLP	705,672	1,020	0
MW	691,954	1,000	7

^a ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull

^b Eficiência calculada como a razão da média de cada modelo em relação a menor média de χ^2 observada

predições e classificação de touros semelhantes e distintas do modelo de riscos proporcionais Weibull. Em termos de habilidade de predição de eventos reprodutivos das fêmeas o modelo MW apresentou a melhor performance no esquema de validação proposto, obtendo a menor média de χ^2 e o menor valor desta estatística em 7 das 10 repetições avaliadas. Portanto, o modelo de riscos proporcionais Weibull seria o mais indicado para a avaliação genética de idade ao primeiro parto nesta população.

2.4 CONCLUSÕES

A utilização de observações censuradas na análise de idade ao primeiro parto acarretou diminuição nas estimativas de herdabilidade. Além disso, o modelo em que as informações fenotípicas são analisadas considerando a metodologia de análise de sobrevivência apresentou melhor habilidade de predição de eventos reprodutivos das progênes dos reprodutores. Portanto, para a avaliação genética de idade ao primeiro parto, nesta população, seria recomendado a utilização do modelo de riscos proporcionais Weibull.

CAPÍTULO 3 - MODELAGEM DE INFORMAÇÕES CENSURADAS NA AVALIAÇÃO DE DIAS PARA O PARTO

3.1 INTRODUÇÃO

Na bovinocultura de corte, o desempenho reprodutivo do rebanho determina diretamente o sucesso do empreendimento agropecuário. Logo, realizar mensurações que permitam controlar e avaliar o desempenho reprodutivo é ponto essencial e estratégico. Neste contexto, pode-se destacar a característica dias para o parto (DPP), pois esta permite a identificação de animais que apresentam maior fertilidade no rebanho e de touros que produzem filhas com capacidade de engravidar precocemente no período de estação de monta (FORNI; ALBUQUERQUE, 2005).

Estimativas de herdabilidade reportadas para DPP na raça Nelore são de baixa magnitude, variando de 0,04 a 0,16 (PEREIRA; ELER; FERRAZ, 2000; MERCADANTE et al., 2002; FORNI; ALBUQUERQUE, 2005; BOLIGON; ALBUQUERQUE; RORATO, 2008). Logo, grande parte da expressão de DPP é atribuída à combinação de fatores ambientais e interações gênicas, entretanto outro fator que possivelmente influencia a variabilidade genética de DPP está relacionado a estrutura dos bancos de dados normalmente analisados em caracteres reprodutivos (MEYER; HAMMOND; PARNELL, 1990; MERCADANTE et al., 2002; GUO; GIANOLA; SHORT, 2001; HOU et al., 2009). Pois, as fêmeas que não possuem expressão fenotípica, devido a falhas reprodutivas, ou que apenas os limites inferiores de seus valores fenotípicos são conhecidos, por ainda estarem em reprodução ou serem descartadas voluntariamente, normalmente são excluídas das avaliações genéticas.

Portanto, a utilização de amostras usualmente viesadas podem contribuir para mascarar a variabilidade genética e as diferenças genéticas entre os animais para características reprodutivas (NOTTER, 1988; MEYER; HAMMOND; PARNELL, 1990; MERCADANTE, 1995; HOU et al., 2009). De acordo com Guo, Gianola e Short (2001), a inclusão de observações censuradas nas avaliações genéticas de forma adequada poderia evitar distorções nas inferências e melhoria na predição dos valores genéticos.

Logo, o objetivo deste trabalho foi comparar modelos que tratam a expressão fenotípica da característica dias para o parto como não censurada e censurada quanto a estimativa de parâmetros genéticos e habilidade de predição.

3.2 MATERIAL E MÉTODOS

Dados

Um banco de dados de 82.846 fêmeas Nelore, nascidas entre 1993 e 2007, que possuíam registros de idade ao primeiro parto foi utilizado para a consistência. Estes animais pertencem ao arquivo zootécnico de 171 fazendas de bovinocultura de corte, localizadas na região sudeste e centro-oeste do Brasil, participantes da base de dados Aliança Nelore, a qual agrega informações de diferentes programas de melhoramento genético da raça Nelore.

Os animais são criados em sistema de pastejo e o manejo reprodutivo de múltiparas é baseado em uma estação reprodutiva realizada no verão, em que, as matrizes são desafiadas à reprodução durante 90 dias, de Novembro a Janeiro. Após o diagnóstico de gestação, fêmeas que não emprenharam são descartadas.

Como a informação da data em que a estação de monta iniciou não estava disponível no banco de dados, a característica dias para o segundo parto (DPP) foi calculada da seguinte forma: a) Identificação da menor data de segundo parto (MDSP) em cada grupo de contemporâneos, considerou-se que estas fêmeas emprenharam no primeiro dia de estação de monta; b) Dentro de cada GC, a data do segundo parto de cada matriz foi subtraída do MDSP e adicionado o período de gestação, sendo este considerado 280 dias. Animais com DPP maior que 370 dias foram excluídos do banco de dados (11.976 matrizes).

Além do efeito genético, grupos de contemporâneos (GC) foram considerados como efeitos sistemáticos no modelo. Os GC foram definidos a partir da combinação de mês e safra de nascimento, grupos de manejo e fazenda da desmama e sobreano e precocidade, matrizes com idade ao primeiro parto inferior a 30 meses foram classificadas como precoces. Apenas fêmeas pertencentes a GC com mais de 5 animais com informações fenotípicas não censuradas foram utilizadas nas análises.

A definição dos registros censurados foi realizada considerando-se como informação fenotípica o maior valor de dias para o segundo parto em cada GC ou adicionando 21 dias ao maior valor de DPP de cada grupo de contemporâneos (JOHNSON; BUNTER, 1996). Uma descrição completa dos bancos de dados é apresentada na Tabela 3.1.

Modelos Estatísticos

Modelos lineares Gaussianos

Nesta classe três modelos foram utilizados: a) Modelo linear Gaussiano (ML): em que todas as observações censuradas foram excluídas do banco de dados de análise; b) Modelo linear Gaussiano com penalização (MLP): em que as informações

Tabela 3.1 – Descrição da estrutura do banco de dados da característica dias para o segundo parto quanto ao número de registros, estatística descritiva, informações de grupos de contemporâneos e pedigree

Registros	
Censurados	10.432
Não censurados	22.131
	32.563
Estatística descritiva dos registros não censurados (dias)	
Mínimo	280,00
Máximo	369,00
Média	312,20
Desvio-padrão	31,58
Estatística descritiva dos registros censurados (dias) ^a	
Mínimo	286,00
Máximo	369,00
Média	347,20
Desvio-padrão	18,29
Grupos de Contemporâneos	
Número de grupos	2.101
Mínimo de animais não censurados	5
Máximo de animais não censurados	57
Mínimo de animais censurados	0
Máximo de animais censurados	93
Pedigree	
Número de gerações	5
Número de touros	921
Número total de animais	64.244

^a A estatística descritiva, quando a penalização é adotada, é obtida somando aos valores da tabela 21 dias, exceto para o desvio-padrão que permanece o mesmo

censuradas foram obtidas a partir da combinação do maior valor de DPP do grupo de contemporâneos acrescido de penalização de 21 dias; c) Modelo linear Gaussiano com censura (MLC): os registros censurados foram simulados a partir de uma distribuição normal truncada, sendo o maior valor de dias para o parto observado no GC adotado como ponto de truncamento (SORENSEN; GIANOLA; KORSGAARD, 1998;

GUO; GIANOLA; SHORT, 2001).

Os modelos lineares Gaussianos podem ser representados matricialmente da seguinte forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (3.1)$$

em que: $\mathbf{y}$ é o vetor de observações de dias para o parto; $\boldsymbol{\beta}$ é o vetor de efeitos sistemático de grupos de contemporâneos; $\mathbf{u}$ é o vetor dos valores genéticos dos animais, com a pressuposição de que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, sendo $\mathbf{A}$ e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente; $\mathbf{e}$ o vetor de resíduos, os quais são assumidos normalidade e homocedasticidade de variância, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, em que, $\mathbf{I}$ e σ_e^2 são a matriz identidade e variância residual, respectivamente.

A densidade condicional do vetor de observações ($\mathbf{y}$) quando as informações censuradas são consideradas, no modelo linear gaussiano com censura, pode ser expressa da seguinte forma (SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001):

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = (2\pi\sigma_e^2)^{-m/2} \exp \left[-\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) \right] \prod_{i=m+1}^n \left[1 - \Phi \left(\frac{c_i - x_i'\boldsymbol{\beta} - z_i'\mathbf{u}}{\sigma_e} \right) \right] \quad (3.2)$$

em que: m e n são, respectivamente, o número de observações não censuradas e o total de observações; $\Phi(\cdot)$ é a distribuição normal acumulada; c_i o valor de DPP censurado adotado para a fêmea i , neste caso o maior valor de dias para o parto do GC da matriz i .

Quando a expressão (3.2) é reescrita utilizando a técnica de aumento de dados é observada uma simplificação no modelo (TANER; WONG, 1987). Logo, a representação da distribuição conjunta do vetor de informações com base nos registros observados ($\mathbf{y}_{nc}$) e não observados ($\mathbf{y}_c$), pode ser representada como a seguir (SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001):

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_e^2} \left[\sum_{i=1}^m (y_{nci} - x_i'\boldsymbol{\beta} - z_i'\mathbf{u}) + \sum_{i=m+1}^n (y_{ci} - x_i'\boldsymbol{\beta} - z_i'\mathbf{u}) \right] \right\} \quad (3.3)$$

em que: $y_{c_i} \geq c_i$, representando que o valor fenotípico não observado para a matriz i deve ser maior ou igual a ponto de truncamento definido como registro censurado.

As observações aumentadas (y_c) são obtidas em cada iteração do amostrador de Gibbs a partir de uma distribuição condicional completa *a posteriori* normal truncada (SORENSEN; GIANOLA; KORSGAARD, 1998; CARDOSO et al., 2009).

Modelos Linear-Limiar (MLL)

Nestes modelos, uma variável indicadora de censura categórica foi determinada, sendo esta definida como 0, para matrizes que não possuíam registros de dias para o segundo parto, e 1, para fêmeas não censuradas. Logo, esta característica foi modelada como correlacionada à DPP. Os seguintes modelos foram adotados: a) Modelo linear-limiar (MLL): em que matrizes sem fenótipos foram incluídas na análise considerando seus registros de DPP como informações perdidas (ARNASON, 1999; URIOSTE; MISZTAL; BERTRAND, 2007b); b) Modelo linear-limiar com penalização (MLLP): aos registros censurados foram atribuídos os maiores valores de dias para o segundo parto em cada grupo de contemporâneos.

Os modelos linear-limiar e linear-limiar com penalização podem ser representados matricialmente como a seguir (HOU et al., 2009):

$$\begin{bmatrix} y \\ l \end{bmatrix} = \begin{bmatrix} X_y & 0 \\ 0 & X_l \end{bmatrix} \begin{bmatrix} \beta_y \\ \beta_l \end{bmatrix} + \begin{bmatrix} Z_y & 0 \\ 0 & Z_l \end{bmatrix} \begin{bmatrix} u_y \\ u_l \end{bmatrix} + \begin{bmatrix} e_y \\ e_l \end{bmatrix} \quad (3.4)$$

em que: $y' = \begin{bmatrix} y_{nc} & 0 \end{bmatrix}$ e $y' = \begin{bmatrix} y_{nc} & y_c \end{bmatrix}$ são os vetores de observações para os modelos MLL e MLLP, respectivamente; l é o vetor de predisposição do estado de censura dos registros.

Os programas utilizados nas análises foram: THRGIBBS1F90 para os modelos linear-limiar e linear-limiar com penalização; GIBBS1F90 nos modelos linear Gaussiano e linear Gaussiano com penalização; GIBBS2CEN empregado no modelo linear Gaussiano com censura (MISZTAL et al., 2002). Na amostragem de Gibbs foram utilizadas três cadeias de 500.000 iterações com distintos valores iniciais para os componentes de variância, os quais expressavam herdabilidades de 0,10, 0,40 e 0,80. Inspeção gráfica e os testes de Brooks, Gelman e Rubin (GELMAN; RUBIN, 1992) e Heldelberger e Welch (HELDELBERGER; WELCH, 1983) utilizando o pacote *coda* (PLUMMER et al., 2006) do programa R (R Core Team, 2013) foram realizados, sendo obtidos resultados positivos de convergência das análises. Para inferência apenas

uma das cadeias foi utilizada com descarte amostral de 50.000, modelos ML, MLC e MLLP, e 100.000, modelo MLL.

Modelo de Riscos Proporcionais Weibull (MW)

Nos modelos de análise de sobrevivência Weibull a função taxa de falha base é denotada por $\lambda_0(t) = \lambda\rho(\lambda t)^{\rho-1}$, em que, ρ e λ são os parâmetros de forma e escala desta distribuição, respectivamente. A inclusão de covariáveis, é realizada de forma multiplicativa à função de taxa de falha base utilizando a função de ligação exponencial. Quando um efeito aleatório é incluído como variável explicativa, como o efeito genético aditivo de animal, o modelo passa a ser denominado modelo de fragilidade e, pode ser representado como a seguir (DUCROCQ; CASELLA, 1996):

$$h(t) = \lambda_0(t) \exp(\mathbf{X}\beta + \mathbf{Z}s) \quad (3.5)$$

em que: $h(t)$ é a função taxa de falha; $\lambda_0(t)$ a função taxa de falha base Weibull, $\lambda_0 = \lambda\rho(\lambda t)^{\rho-1}$ sendo λ e ρ os parâmetros de escala e forma, respectivamente; β vetor dos efeitos fixos tempo-independente de grupos de contemporâneos; s efeito aleatório tempo-independente de touro, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $\mathbf{A}_s\sigma_s^2$, sendo $\mathbf{A}_s$ e σ_s^2 a matriz de numerador de parentesco entre os touros e a variância de touro, respectivamente.

O modelo de touro, descrito em (3.5), foi utilizado para a estimativa da variância de touro e, este componente foi multiplicado por quatro para se obter a estimativa da variância genética aditiva. Posteriormente, este componente de variância foi considerado como fixo para a predição dos valores genéticos utilizando o modelo animal, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $\mathbf{A}\sigma_u^2$, sendo $\mathbf{A}$ e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente. A edição do dados foi realizada de forma semelhante à utilizada para o banco de dados completo e a descrição da estrutura dos dados é apresentada na Tabela 3.2.

A adequabilidade do modelo Weibull como distribuição da taxa de falha base foi avaliada plotando o logaritmo natural ($\ln$) de DPP e o $\ln[-\ln(\hat{S}(DPP))]$, em que $\hat{S}(DPP)$ é a estimativa da função de sobrevivência obtida pelo método não-paramétrico Kaplan-Meier (Figura 3.1). Para a escolha do modelo Weibull, a relação entre $\ln(DPP)$ vs $\ln[-\ln(\hat{S}(DPP))]$ deve ser linear, comportamento observado na Figura (3.1). Logo, o ajuste utilizando a distribuição Weibull é aceitável e o programa

Tabela 3.2 – Descrição da estrutura dos dados da característica dias para o segundo parto, utilizada no modelo de touro de riscos proporcionais Weibull, quanto ao número de registros, estatística descritiva, informações de grupos de contemporâneos e pedigree

Registros	
Censurados	4.920
Não censurados	15.532
	20.452
Estatística descritiva dos registros não censurados (dias)	
Mínimo	280,00
Máximo	369,00
Média	312,90
Desvio-padrão	24,25
Estatística descritiva dos registros censurados (dias)	
Mínimo	286,00
Máximo	369,00
Média	348,40
Desvio-padrão	17,42
Grupos de Contemporâneos	
Número de grupos	999
Mínimo de animais não censurados	5
Máximo de animais não censurados	57
Mínimo de animais censurados	0
Máximo de animais censurados	37
Pedigree	
Número de gerações	5
Número de touros	921
Número total de animais	1.354

Survival Kit v6.0 foi empregado nas análises (DUCROCQ; SOLKNER; MÉSZÁROS, 2010).

Comparação dos modelos

A concordância entre as predições dos valores genéticos obtidos pelos dife-

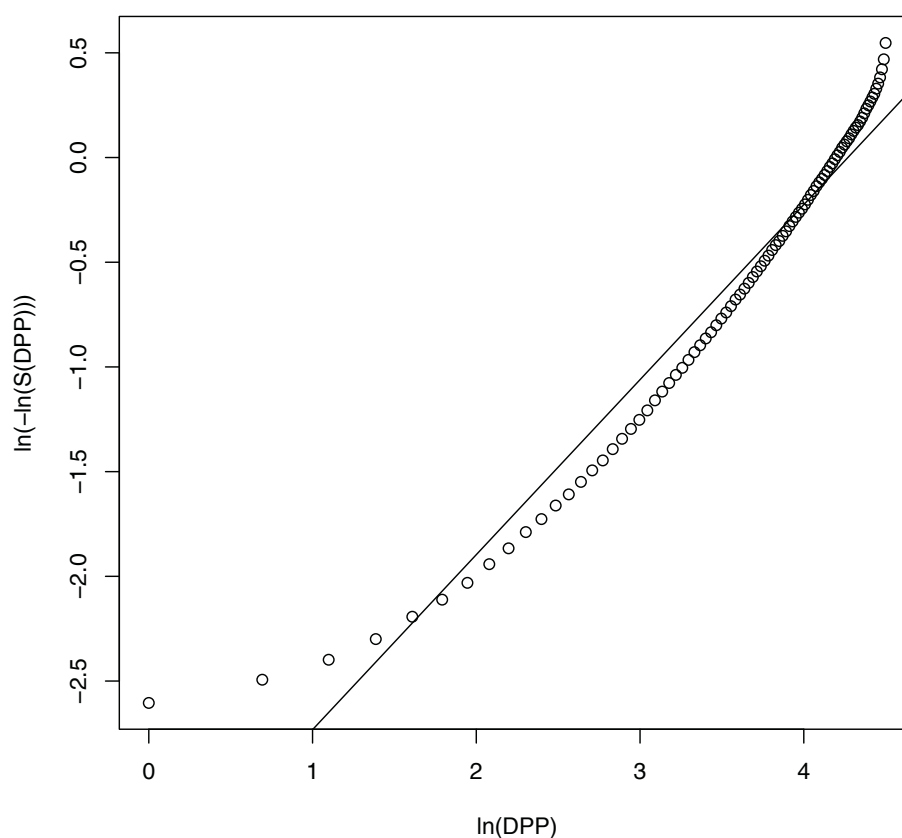


Figura 3.1 – Verificação gráfica do ajuste da distribuição Weibull como taxa de falha base na avaliação de dias para o parto (DPP). A estimativa da função de sobrevivência, $S(t)$, foi obtida utilizando o estimador de Kaplan-Meier.

rentes modelos foi avaliada a partir de correlações de Pearson e da porcentagem de coincidência na classificação dos 10% melhores touros (TOP10), utilizando todos os reprodutores ou touros com mais de 10 filhas.

Um esquema de validação cruzada, com 10 repetições, foi delineado para avaliar a habilidade de predição dos modelos utilizados. O banco de dados foi dividido em partições de treinamento e validação, sendo que em cada uma destas 80 e 20% dos animais estavam presentes, respectivamente. As matrizes foram aleatoriamente alocadas entre as partições do banco de dados seguindo a restrição de pelo menos 3 animais por GC no treinamento e que cada touro possuísse progênie, censuradas e não censuradas, na base de treinamento e validação.

A habilidade de predição dos modelos foi testada utilizando informações de eventos reprodutivos esperados e observados de progênie dos touros com mais de

10 filhas não censuradas (184 reprodutores). Portanto, a característica dias para o segundo parto foi categorizada em fêmeas com informação fenotípica e matrizes censuradas. Diferenças esperadas na progênie (DEP) foram preditas no treinamento e, probabilidades de o reprodutor possuir filhas em cada uma das classes foram obtidas a partir da regressão logística dos eventos reprodutivos observados no treinamento sobre as DEP dos touros. Posteriormente, as probabilidades estimadas no treinamento foram utilizadas para a predição da frequência esperada das filhas de cada touro nas classes de DPP no banco de validação. Logo, a soma da estatística χ^2 sobre os touros foi calculada como se segue:

$$\chi^2 = \sum_{i=1}^{nt} \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3.6)$$

em que: O_{ij} e E_{ij} são o número de fenótipos observados e esperados, respectivamente, das progênies do touro i serem da classe de DPP j (1 ou 2); nt é o número total de touros, variando de 1 até 184 animais.

3.3 RESULTADOS E DISCUSSÃO

As estimativas de herdabilidade de dias para o segundo parto (Tabela 3.3), utilizando os modelos lineares Gaussianos e modelos linear-limiar, foram inferiores às reportadas em trabalhos na raça Nelore, em que, este parâmetro genético apresentou estimativas variando entre 0,04 a 0,16 (PEREIRA; ELER; FERRAZ, 2000; MERCADANTE et al., 2002; PEREIRA; ELER; FERRAZ, 2002; FORNI; ALBUQUERQUE, 2005; FORNI; ALBUQUERQUE, 2006; BOLIGON; ALBUQUERQUE; RORATO, 2008)

Similaridade entre estimativas de herdabilidade obtidas para DPP foram reportadas por Donoghue, Rekaya e Bertrand (2004a), Donoghue, Rekaya e Bertrand (2004b). Estes autores, trabalhando com simulação e dados reais de dias para o parto, obtiveram $\hat{h}^2$ semelhantes entre os modelos ML, MLP e MLC. Da mesma forma, Urioste, Misztal e Bertrand (2007a) obtiveram h^2 próximas entre MLP e MLC, avaliando DPP entre o primeiro-segundo, segundo-terceiro e terceiro-quarto partos.

Por outro lado, a estimativa de herdabilidade para DPP, obtida na escala logarítmica, utilizando o modelo de risco proporcionais Weibull de 0,06 foi inferior a reportada por Pereira (2008), avaliando reconcepção em fêmeas primíparas da raça Nelore, sendo esta igual a 0,08. A estimativa de ρ , 1,10 ($\pm 0,007$), indica um aumento na taxa de prenhez das matrizes com o avançar da estação de monta, entretanto a magnitude foi inferior a obtida por Pereira (2008), valor igual a 1,69. Esta diferença pode estar

Tabela 3.3 – Média e desvio padrão, entre parênteses, a posteriori para a variância genética (σ_u^2), variância residual (σ_e^2) e herdabilidade ($\hat{h}^2$) de dias para o segundo parto, obtidas utilizando diferentes modelos de avaliação genética

Parâmetro	Modelo					
	ML	MLP	MLC	MLL	MLLP	MW
$\hat{\sigma}_u^2$	27,63 (7,145)	74,520 (18,610)	37,210 (11,100)	11,67 (12,645)	41,30 (10,488)	0,10 (0,024)
$\hat{\sigma}_e^2$	497,820 (8,227)	907,360 (18,210)	567,150 (10,930)	560,59 (39,131)	563,548 (10,584)	- -
$\hat{h}^2$	0,03 (0,014)	0,08 (0,018)	0,06 (0,018)	0,02 (0,017)	0,07 (0,017)	0,06 (0,011)

^a ML: modelo linear Gaussiano;
 MLP: modelo linear Gaussiano com penalização;
 MLC: modelo linear Gaussiano com censura;
 MLL: modelo linear-limiar;
 MLLP: modelo linear-limiar com penalização;
 MW: modelo de riscos proporcionais Weibull

relacionada ao modo com que a DPP foi calculada, sendo expressa como desvio do menor valor observado em cada grupo de contemporâneos.

As correlações de Pearson entre os valores genéticos preditos utilizando os modelos Gaussianos com penalização e censura (MLP e MLC), linear-limiar com penalização (MLLP) e de riscos proporcionais Weibull (MW) foram de alta magnitude nas duas situações delineadas, todos os touros e touros com mais de 10 progênes (Tabela 3.4). Logo, uma forte associação entre as predições destes modelos foi observada. Resultados semelhantes foram reportados por Donoghue, Rekaya e Bertrand (2004b), trabalhando com dias para o parto em gado de corte, em que obtiveram correlações de alta magnitude entre os modelos MLP-MLC (0,99). Da mesma forma, Hou et al. (2009), avaliando dias entre o parto e a primeira inseminação e a característica período de serviço em gado Holandês, observaram correlações próximo a unidade entre os modelos MLP, LMC e MLLP e, valores em torno de 0,90 entre estes modelos e MW.

Menores associações entre as predições do modelo linear Gaussiano com os demais modelos utilizados foram obtidas, variando de -0,60 (MW) a 0,84 (MLL) e -0,55 (MW) a 0,88 (MLL), respectivamente, nas situações em que todos os touros e reprodutores com mais de 10 progênes foram adotados. Por outro lado, valores de correlações superiores aos anteriormente descritos, foram estimados entre os valores

Tabela 3.4 – Correlações de Pearson entre os valores genéticos preditos de todos os touros (acima da diagonal) e de touros com mais de 10 progênie não censuradas (abaixo da diagonal) utilizando distintos modelos de avaliação genética para dias para o parto^a

Modelos ^b	ML	MLP	MLC	MLL	MLLP	MW
ML		0,69	0,80	0,84	0,68	-0,60
MLP	0,66		0,98	0,90	0,99	-0,97
MLC	0,78	0,98		0,92	0,97	-0,94
MLL	0,88	0,87	0,93		0,90	-0,84
MLLP	0,66	0,99	0,98	0,89		-0,97
MW	-0,55	-0,98	-0,93	-0,82	-0,97	

^a 921 e 184 touros;

^b ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull.

genéticos preditos por MLL e MLP, MLC, MLLP e MW, os quais variaram de -0,60 (MW) a 0,92 (MLC) e -0,84 (MW) a 0,93 (MLC), nas duas situações delineadas. Donoghue, Rekaya e Bertrand (2004b), avaliando DPP em gado de corte, reportaram resultados similares aos obtidos neste estudo, pois obtiveram correlações de Pearson, entre predições dos valores genéticos de touros, para os modelos ML-MLP e ML-MLC variando em torno de 0,77 a 0,80 e de 0,81 a 0,84, respectivamente. Uma importante observação é em relação à diferença de interpretação entre as DEP preditas via modelos lineares e de análise de sobrevivência, sendo estas, respectivamente, a diminuição esperada nos dias para emprenhar na estação de monta e a chance das progênie de um touro apresentarem dias para o parto inferiores. Por isso, a correlação entre as predições dos modelos lineares e riscos proporcionais Weibull são negativas.

A porcentagem de coincidência na classificação dos 10% melhores touros foram semelhantes e de alta magnitude entre os modelos MLP, MLC, MLLP e MW (Tabela 3.5). Donoghue, Rekaya e Bertrand (2004b) e Hou et al. (2009), reportaram resultados semelhantes ao obtido neste estudo, pois estes autores observaram, respectivamente, uma grande correspondência na classificação de touros para os modelos MLP e MLC e MLP, MLC e MLLP. Da mesma forma, González-Recio et al. (2006) e Pereira et al. (2007) investigando, respectivamente, dias de serviço em gado Holandês e idade à primeira concepção em novilhas Nelore, reportaram elevada correspondência na classificação de reprodutores para os modelos MLC e MW.

Tabela 3.5 – Porcentagem de coincidência na classificação dos 10% melhores touros (acima da diagonal) e touros com mais de 10 progênes não censuradas (abaixo da diagonal) utilizando distintos modelos de avaliação genética para dias para o parto^a

Modelos ^b	ML	MLP	MLC	MLL	MLLP	MW
ML		58,70	66,30	63,04	55,43	54,35
MLP	50,00		91,30	69,57	90,22	89,13
MLC	61,11	88,89		72,83	85,87	81,52
MLL	55,56	72,22	66,67		67,39	61,96
MLLP	50,00	94,44	88,89	66,67		90,22
MW	50,00	88,89	88,89	66,67	88,89	

^a 92 e 19 touros;

^b ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull.

A correspondência entre a classificação dos 10% melhores touros, nas duas estratégias delineadas, foram inferiores entre os modelos linear gaussiano e linear-limiar e, destes com os modelos MLP, MLC, MLLP e MW (Tabela 3.5). Resultados similares foram reportados por Donoghue, Rekaya e Bertrand (2004b) e Silva et al. (2010), os quais observaram reclassificação de touros entre os modelos ML-MLP, trabalhando, respectivamente, com dias para o parto na raça Angus e idade ao primeiro parto em fêmeas Nelore.

O modelo linear Gaussiano com censura apresentou a maior habilidade de predição, com menor média de χ^2 e o maior número de vezes em que possuiu o menor valor desta estatística (Tabela 3.6). Os modelos linear Gaussiano com penalização, linear Gaussiano e de riscos proporcionais Weibull obtiveram os menores valores de χ^2 em 2, 2 e 1 das 10 repetições realizadas, respectivamente, e além disso, médias de χ^2 que os classificaram como o segundo, quarto e quinto modelos. Já os modelos MLLP e MLL não apresentaram o menor valor de χ^2 em nenhuma das repetições e, obtiveram a terceira e sexta médias em habilidade de predição.

Donoghue, Rekaya e Bertrand (2004b), trabalhando com dias para o parto em gado de corte, recomendaram a utilização dos modelos MLP e MLC na avaliação genética desta característica, pois quando os registros censurados foram ignorados uma menor estimativa de variância genética foi obtida e, além disso, reclassificação de touros foram observadas entre estes modelos e o modelo gaussiano. Entretanto,

Tabela 3.6 – Média da estatística χ^2 , eficiência e o número de vezes em que os distintos modelos de avaliação genética de dias para o parto apresentaram o menor valor de χ^2 (NM χ^2), num esquema de validação cruzada com 10 repetições

Modelo ^a	Média χ^2	Eficiência ^b	NM χ^2
ML	533,619	1,025	2
MLP	521,901	1,003	2
MLC	520,337	1,000	5
MLL	560,655	1,077	0
MLLP	524,304	1,008	0
MW	528,304	1,015	1

^a ML: modelo linear Gaussiano;

MLP: modelo linear Gaussiano com penalização;

MLC: modelo linear Gaussiano com censura;

MLL: modelo linear-limiar;

MLLP: modelo linear-limiar com penalização;

MW: modelo de riscos proporcionais Weibull

^b Eficiência calculada como a razão da média de cada modelo em relação a menor média de χ^2 observada

neste estudo nenhuma estratégia de validação cruzada foi utilizada.

Avaliando dias para o parto entre o primeiro-segundo, segundo-terceiro e terceiro-quarto partos, Urioste, Misztal e Bertrand (2007a) reportaram resultado distinto ao deste estudo, pois relataram que o modelo MLL seria o mais recomendado para avaliação genética de DPP. O modelo linear-limiar apresentou maiores estimativas de correlações de Pearson entre predições de valores genéticos de touros em duas partições do banco de dados de análise, cada uma com aproximadamente 50% das informações. Porém, esta estratégia de validação foi realizada com apenas uma repetição.

Os modelos linear Gaussiano com penalização, linear Gaussiano com censura, linear-limiar com penalização e modelo de riscos proporcionais Weibull apresentaram associações das predições e classificação de touros semelhantes, porém distintas aos dos modelos linear Gaussiano e linear-limiar. Em termos de habilidade de predição de eventos reprodutivos das fêmeas o modelo MLC apresentou a melhor performance no esquema de validação proposto, obtendo a menor média de χ^2 e o menor valor desta estatística em 5 das 10 repetições avaliadas. Portanto, o modelo linear Gaussiano com censura seria o mais indicado para avaliação genética de dias para o parto nesta população.

3.4 CONCLUSÕES

A inclusão de observações censuradas na análise de dias para o parto acarretaram mudanças nas estimativas de herdabilidade. Além disso, o modelo que utiliza uma densidade condicional completa *a posteriori* normal truncada na amostragem de valores para as informações censuradas apresentou melhor habilidade de predição de eventos reprodutivos das progênies dos touros. Portanto, para a avaliação de dias para o parto, nesta população, seria recomendado a utilização do modelo linear Gaussiano com censura.

CAPÍTULO 4 - AVALIAÇÃO DA LONGEVIDADE PRODUTIVA MODELANDO REGISTROS CENSURADOS

4.1 INTRODUÇÃO

O desempenho reprodutivo dos bovinos de corte está diretamente relacionado à oferta de produto final da cadeia produtiva. Portanto, a rentabilidade do sistema produtivo e os preços dos produtos ao consumidor são influenciados pela eficiência reprodutiva dos animais. Além disso, a performance reprodutiva dos rebanhos limita o ganho genético da população, visto que os baixos índices reprodutivos afetam o intervalo de gerações e a intensidade de seleção a ser adotada nos programas de melhoramento (VAN ARENDONK, 1986; VAN MELIS et al., 2010a).

A longevidade produtiva apresenta-se como um critério de seleção que pode ser utilizado visando o incremento do potencial reprodutivo das matrizes. A maximização deste critério possibilita a redução da taxa de descarte involuntário, acarretando diminuição nos custos de reposição e aumento do número de animais produtivos, e consequentemente, um potencial aumento de rentabilidade do sistema de produção (VAN ARENDONK, 1986; NEVES; CARVALHEIRO; QUEIROZ, 2012). Diferentes abordagens têm sido utilizadas para avaliação da longevidade na raça Nelore, nas quais a expressão fenotípica é avaliada na escala discreta, “stayability” ou número de partos (SILVA et al., 2003; SILVA et al., 2006; VAN MELIS et al., 2010a; VAN MELIS et al., 2010b; NEVES; CARVALHEIRO; QUEIROZ, 2012), e contínua, vida produtiva ou idade da vaca ao último parto (MERCADANTE; LÔBO; OLIVEIRA, 2000; BERTAZZO et al., 2004; CAETANO, 2011).

Quando avaliada na escala contínua, apenas a expressão parcial da longevidade de algumas matrizes é conhecida, pois uma porção das fêmeas permanecem produtivas no momento da avaliação genética e ou foram descartadas involuntariamente, por razões não relacionadas à eficiência reprodutiva. Estes registros parciais são denominados informações censuradas e, de acordo com Hou et al. (2009), a utilização destes fenótipos como verdadeiros ou excluí-los das análises genéticas podem favorecer touros com elevado número de progênie com registros censurados. Entretanto, a inclusão adequada destas observações nas avaliações genéticas poderiam evitar distorções nas inferências e melhoria na predição dos valores genéticos (GUO; GIANOLA; SHORT, 2001).

Logo, o objetivo deste trabalho foi comparar modelos que tratam a expressão fenotípica da característica longevidade produtiva como não censurada e censurada quanto a estimativa de parâmetros genéticos e habilidade de predição.

4.2 MATERIAL E MÉTODOS

Dados

Os dados utilizados no estudo são oriundos do programa de avaliação genética de bovinos de corte da base de dados Aliança Nelore. Na consistência, foram utilizadas 82,304 matrizes Nelore, pertencentes a 171 fazendas localizadas nas regiões sudeste e centro-oeste do Brasil, nascidas entre 1993 e 2007 que possuíam registros de idade ao primeiro parto.

Os animais são manejados em sistema de pastejo com suplementação mineral e o manejo reprodutivo de multíparas consiste na realização de uma estação de monta de verão, normalmente de Novembro a Janeiro. Fêmeas que apresentam diagnóstico de gestação negativo, obtido 60 dias após à estação reprodutiva, são descartadas dos rebanhos.

A longevidade produtiva (LONG) foi definida como a diferença entre as datas de desmame da última progênie e a data de nascimento da matriz. Fêmeas com informações incompletas de desmame dos seus produtos (4.818 matrizes) e LONG inferior a 30 meses (124 vacas) foram excluídas das análises. Após esta primeira etapa da consistência, o banco de dados continha 77.362 fêmeas, das quais, 61.097 apresentavam informações fenotípicas não censuradas e 16.256 matrizes com registros censurados, pois apresentavam filhos nascidos na safra de 2010.

Os grupos de contemporâneos (GC) foram determinados a partir da combinação de mês e safra de nascimento, grupos de manejo e fazenda da desmama e sobreano e precocidade, sendo consideradas matrizes precoces animais com idade ao primeiro parto inferior a 30 meses. Apenas fêmeas pertencentes a grupos de contemporâneos com mais de cinco matrizes com registros fenotípicos não censurados permaneceram nas análises. Portanto, 15.688 e 9.709 matrizes que possuíam registros conhecidos e permaneciam produtivas, respectivamente, tiveram seus registros excluídos. A descrição completa da estrutura dos dados utilizados é apresentada na Tabela 4.1.

Modelos Estatísticos

No modelo linear Gaussiano (ML) apenas as informações de matrizes não censuradas foram utilizadas na análise. Este modelo pode ser descrito como se segue:

$$y = X\beta + Zu + e \quad (4.1)$$

Tabela 4.1 – Descrição da estrutura dos dados da característica longevidade produtiva quanto ao número de registros, estatística descritiva, informações de grupos de contemporâneos e pedigree

Registros	
Censurados	6.547
Não censurados	45.409
	51.956
Estatística descritiva dos registros não censurados (meses)	
Mínimo	30,15
Máximo	207,60
Média	55,12
Desvio-padrão	23,84
Estatística descritiva dos registros censurados (meses)	
Mínimo	41,95
Máximo	212,10
Média	59,81
Desvio-padrão	31,59
Grupos de Contemporâneos	
Número de grupos	3.701
Mínimo de animais não censurados	5
Máximo de animais não censurados	91
Mínimo de animais censurados	0
Máximo de animais censurados	57
Pedigree	
Número de gerações	5
Número de touros	1.256
Número total de animais	97.073

em que: $\mathbf{y}$ é o vetor de observações de longevidade produtiva; β é o vetor de efeitos sistemáticos de grupos de contemporâneos; $\mathbf{u}$ é o vetor dos valores genéticos dos animais, com a pressuposição de que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, sendo $\mathbf{A}$ e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente; $\mathbf{e}$ o vetor de resíduos, os quais são assumidos normalidade e homocedasticidade de variância, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, em que, $\mathbf{I}$ e σ_e^2 são a matriz identidade e variância residual,

respectivamente.

O modelo linear Gaussiano com censura (MLC) é similar ao anterior, entretanto neste caso são incluídas as informações censuradas:

$$\begin{bmatrix} \mathbf{y}_{nc} \\ \mathbf{y}_c \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (4.2)$$

em que: $\mathbf{y}_{nc}$ e $\mathbf{y}_c$ representam, respectivamente, as observações de longevidade produtiva não censuradas e os registros aumentados para as fêmeas censuradas.

No modelo MLC, a técnica de aumento de dados (TANER; WONG, 1987) é utilizada na modelagem e a distribuição condicional conjunta das observações não censuradas ($\mathbf{y}_{nc}$) e censuradas ($\mathbf{y}_c$) pode ser representada como a seguir (SORENSEN; GIANOLA; KORSGAARD, 1998; GUO; GIANOLA; SHORT, 2001):

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_e^2} \left[\sum_{i=1}^m (y_{nci} - x'_i\boldsymbol{\beta} - z'_i\mathbf{u}) + \sum_{i=m+1}^n (y_{ci} - x'_i\boldsymbol{\beta} - z'_i\mathbf{u}) \right] \right\} \quad (4.3)$$

em que: $y_{ci} \geq c_i$, representando que o valor fenotípico não observado para a matriz i deve ser maior ou igual a ponto de truncamento definido como registro censurado.

Os registros de longevidade produtiva parciais ($\mathbf{y}_c$) são obtidos em cada iteração do amostrador de Gibbs utilizando uma distribuição condicional completa *a posteriori* normal truncada (SORENSEN; GIANOLA; KORSGAARD, 1998; CARDOSO et al., 2009), sendo esta denotada por:

$$p(y_{ci} | \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y}_{nc}; y_{ci} > c_i) = \frac{\phi_{y_{ci}}(x'_i\boldsymbol{\beta} + z'_i\mathbf{u}_i, \sigma_e^2)}{1 - \Phi\left(\frac{c_i - x'_i\boldsymbol{\beta} + z'_i\mathbf{u}_i}{\sigma_e}\right)} \quad (4.4)$$

em que: $\phi(\cdot)$ e $\Phi(\cdot)$ são, respectivamente, as funções densidade e acumulada da distribuição normal; c_i é o ponto de truncamento de cada animal, ou seja, a observação parcial de longevidade produtiva da matriz i .

O seguinte modelo de riscos proporcionais Weibull (MW) foi utilizado para avaliação de longevidade produtiva (DUCROCQ; CASELLA, 1996):

$$h(t) = \lambda_0(t) \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \mathbf{Z}\mathbf{s}) \quad (4.5)$$

em que: $h(t)$ é a função taxa de falha; $\lambda_0(t)$ a função taxa de falha base Weibull, $\lambda_0 = \lambda\rho(\lambda t)^{\rho-1}$ sendo λ e ρ os parâmetros de escala e forma, respectivamente; β vetor dos efeitos sistemáticos tempo-independente de grupos de contemporâneos; θ efeito tempo-dependente de período, os quais são definidos posteriormente; s efeito aleatório tempo-independente de touro, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $A_s\sigma_s^2$, sendo A_s e σ_s^2 a matriz de numerador de parentesco entre os touros e a variância de touro, respectivamente.

O modelo de touro, descrito em (4.5), foi utilizado para a estimativa da variância de touro e, este componente foi multiplicado por quatro para se obter a estimativa da variância genética aditiva. Posteriormente, este componente de variância foi considerado como conhecido para a predição dos valores genéticos utilizando o modelo animal, o qual foi assumido seguir uma distribuição normal multivariada com média zero e matriz de variância e covariância $A\sigma_u^2$, sendo A e σ_u^2 a matriz do numerador de parentesco dos animais e a variância genética aditiva, respectivamente. A consistência dos dados foi realizada de forma semelhante ao banco de dados completo e a descrição dos registros utilizados no modelo de touro de riscos proporcionais Weibull são apresentadas na Tabela 4.2.

O estudo da adequabilidade do modelo Weibull como distribuição da taxa de falha base foi realizado graficamente plotando o logaritmo natural ($\ln$) da longevidade produtiva e o $\ln \left[-\ln \left(\hat{S}(LONG) \right) \right]$, em que $\hat{S}(LONG)$ é a estimativa da função de sobrevivência obtida pelo método não-paramétrico Kaplan-Meier (Figura 4.1). Para a utilização do modelo Weibull, a relação entre $\ln(LONG)$ vs $\ln \left[-\ln \left(\hat{S}(LONG) \right) \right]$ deveria ser linear, comportamento não observado na Figura (4.1). Entretanto, é possível distinguir dois intervalos onde a linearidade pode ser considerada e, em cada período a modelagem utilizando a distribuição Weibull é aceitável. Portanto, a variável período, tempo-dependente, foi definida e os seus limites foram inicialmente determinados visualmente e, posteriormente, determinados comparando-se valores de verossimilhança quando pequenas mudanças foram realizadas (PEREIRA et al., 2005; PEREIRA et al., 2007). Logo, os períodos utilizados foram, respectivamente, de 0 a 46,38 e acima de 46,38 meses.

Os programas GIBBS1F90 e GIBBS2CEN (MISZTAL et al., 2002) foram utilizados nas análises dos modelos ML e MLC, respectivamente. Inicialmente, três cadeias, de 500.000 iterações cada, foram obtidas com valores iniciais para os componentes de variâncias que correspondiam às herdabilidades de 0,10, 0,40 e 0,80. Estas cadeias foram submetidas aos testes de Brooks, Gelman e Rubin (GELMAN; RUBIN, 1992) e Heidelberger e Welch (HELDELBERGER; WELCH, 1983) utilizando o pacote *coda*

Tabela 4.2 – Descrição da estrutura dos dados da característica longevidade produtiva, utilizada no modelo de touro de riscos proporcionais Weibull, quanto ao número de registros, estatística descritiva, informações de grupos de contemporâneos e pedigree

Registros	
Censurados	2.747
Não censurados	20.543
	23.290
Estatística descritiva dos registros não censurados (meses)	
Mínimo	30,15
Máximo	189,40
Média	55,23
Desvio-padrão	23,53
Estatística descritiva dos registros censurados (meses)	
Mínimo	42,15
Máximo	187,70
Média	89,86
Desvio-padrão	29,76
Grupos de Contemporâneos	
Número de grupos	1.896
Mínimo de animais não censurados	5
Máximo de animais não censurados	77
Mínimo de animais censurados	0
Máximo de animais censurados	57
Pedigree	
Número de gerações	5
Número de touros	1.256
Número total de animais	1.752

(PLUMMER et al., 2006) do programa R (R Core Team, 2013), sendo obtidos resultados positivos de convergência das análises. Para inferência apenas uma das cadeias foi utilizada com descarte de 50.000 e 100.000 amostras para os modelos ML e MLC, respectivamente. Para o MW, o programa Survival Kit v6.0 (DUCROCQ; SOLKNER; MÉSZÁROS, 2010) foi empregado.

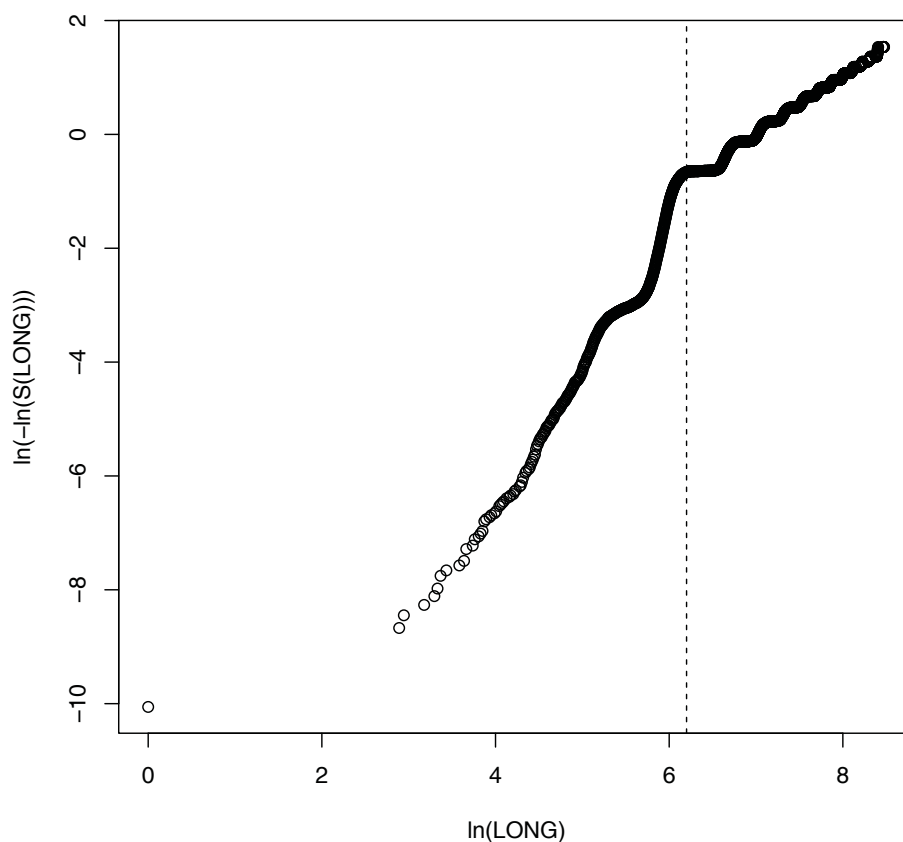


Figura 4.1 – Verificação gráfica do ajuste da distribuição Weibull como taxa de falha base na avaliação da longevidade produtiva (LONG). A estimativa da função de sobrevivência, $S(\text{LONG})$, foi obtida utilizando o estimador de Kaplan-Meier e a linha vertical indica o limite dos períodos utilizados (0 a 46,38 e acima de 46,38 meses)

Comparação dos modelos

A associação das predições de valores genéticos e similaridade na classificação dos animais quanto às diferenças esperadas na progênie, entre os três modelos utilizados, foram mensuradas por meio de correlações de Pearson e porcentagem de coincidência dos 10% melhores reprodutores (TOP10), utilizando todos os touros ou reprodutores com mais de 10 filhas com fenótipos não censurados.

O banco de dados utilizado nas análises foi particionado em treinamento e validação, sendo que em cada um destes 80% e 20% dos registros estavam presentes, respectivamente. Este processo foi aleatoriamente realizado em 10 ocasiões, seguindo as restrições de pelo menos 3 animais por grupo de contemporâneos no treinamento e que cada touro, com mais de 10 progênies com observações fenotípicas

(352 reprodutores), possuísse filhas nas bases de dados de treinamento e validação.

Os componentes de variância, estimados na base de dados completa para cada modelo, foram utilizados como verdadeiros na predição dos valores genéticos dos animais do treinamento. Posteriormente, as diferenças esperadas na progênie (DEP) foram calculadas, para os reprodutores que possuíam mais de 10 filhas não censuradas.

Os modelos foram comparados, em termos de habilidade preditiva, utilizando a estatística χ^2 entre classes de longevidade produtiva observadas e esperadas das progênies de reprodutores com mais de 10 filhas não censuradas. Para tanto, os fenótipos de longevidade produtiva das matrizes foram categorizados utilizando como ponto de truncamento 72 meses, idade normalmente utilizada na avaliação de "stayability" (SILVA et al., 2003; SILVA et al., 2006; VAN MELIS et al., 2010a). As probabilidades de o reprodutor possuir filhas em cada um das classes foram obtidas a partir da regressão logística dos eventos reprodutivos sobre as DEP dos touros, obtidas no treinamento. Posteriormente, as probabilidades estimadas foram utilizadas para prever a frequência esperada das progênies de cada touro em cada classe de LONG no banco de dados de validação. Logo, a estatística χ^2 foi mensurada da seguinte forma:

$$\chi^2 = \sum_{i=1}^{nt} \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4.6)$$

em que: O_{ij} e E_{ij} são o número de fenótipos observados e esperados, respectivamente, das progênies do touro i serem da classe de LONG j (1 ou 2); nt é o número total de touros, variando de 1 até 352 animais.

4.3 RESULTADOS E DISCUSSÃO

As estimativas de herdabilidades ($\hat{h}^2$) para os modelos linear Gaussiano e linear Gaussiano com censura foram de baixa magnitude (Tabela 4.3). Este resultado é discrepante ao reportado por Bertazzo et al. (2004), os quais obtiveram estimativa igual a 0,46, entretanto, semelhante à Mercadante, Lôbo e Oliveira (2000) que obtiveram $\hat{h}^2$ de 0,08. Similaridade entre estimativas de herdabilidade, para os modelos ML e MLC, foram obtidas por Donoghue, Rekaya e Bertrand (2004a) e Donoghue, Rekaya e Bertrand (2004b), trabalhando com simulação e dados reais, respectivamente, para a característica dias para o parto.

A estimativa de herdabilidade na escala logarítmica, utilizando o modelo de riscos proporcionais Weibull (Tabela 4.3), foi inferior a obtida por Caetano (2011), a qual reportou $\hat{h}^2$ igual a 0,25, trabalhando com longevidade de fêmeas Nelore definida

como a diferença entre a data do último parto e o nascimento da matriz. Em contrapartida, a estimativa do parâmetro de forma da distribuição taxa de falha Weibull ($\hat{\rho}$), deste estudo e o reportado por Caetano (2011), foram semelhantes, sendo estas, respectivamente, 2,63 ($\pm 0,01$) e 2,45 ($\pm 0,03$). Este parâmetro é um indicativo se a expressão fenotípica da característica diminui, permanece constante, ou aumenta com a idade dos animais, respectivamente, quando $\hat{\rho} < 1$, $\hat{\rho} = 0$ e $\hat{\rho} > 1$.

Tabela 4.3 – Média e desvio padrão, entre parênteses, a posteriori para a variância genética ($\hat{\sigma}_u^2$), variância residual ($\hat{\sigma}_e^2$) e herdabilidade ($\hat{h}^2$) de longevidade produtiva, obtidas utilizando diferentes modelos de avaliação genética

Parâmetro	Modelo ^a		
	ML	MLC	MW
$\hat{\sigma}_u^2$	16,930 (3,947)	20,280 (3,995)	0,192 (0,035)
$\hat{\sigma}_e^2$	374,40 (4,465)	534,920 (5,360)	- -
$\hat{h}^2$	0,04 (0,010)	0,04 (0,007)	0,11 (0,011)

^a ML: modelo linear Gaussiano;

MLC: modelo linear Gaussiano com censura;

MW: modelo de riscos proporcionais Weibull

A associação entre as predições dos valores genéticos dos modelos lineares Gaussianos, ML e MLC, foram 0,82 e 0,87, considerando, respectivamente, todos os touros e reprodutores com mais de 10 progênes não censuradas (Tabela 4.4). Resultados semelhantes foram reportados por Donoghue, Rekaya e Bertrand (2004b), os quais obtiveram correlações de 0,81 e 0,84, entre predições dos modelos ML e MLC, respectivamente, nas situações em que todos os reprodutores e touros com mais de 20 filhas foram utilizados.

As correlações de Pearson das predições dos modelos MLC e MW, para as duas situações delineadas, foram superiores às obtidas entre ML e MLC (Tabela 4.4). Pereira et al. (2007), estudando idade à primeira concepção de novilhas Nelore, reportaram associação superior, sendo estas -0,99 e -0,92, quando valores genéticos de todos os reprodutores e touros classificados entre os 10% superiores foram utilizados. Por outro lado, Hou et al. (2009), investigando dias entre a primeira inseminação e o parto em gado de leite, obtiveram resultados similares aos obtidos neste estudo, os quais descreveram correlações de -0,92 entre as predições dos modelos MLC e MW.

Vale ressaltar, que na análise de sobrevivência as covariáveis, que influenciam a característica em estudo, são modeladas multiplicativamente à função taxa de falha base Weibull por meio da função exponencial, expressão (4.5). Logo, quanto mais negativa for a predição de valor genético menor será a taxa de falha e, consequentemente, melhor será o mérito genético deste animal em termos de longevidade. Portanto, correlações de Pearson negativas entre as predições dos modelos lineares Gaussianos e MW são esperadas.

Os modelos linear Gaussiano e de riscos proporcionais Weibull apresentaram os menores valores de associação entre as predições dos valores genéticos dos reprodutores, sendo estas iguais a -0,68 e -0,74 (Tabela 4.4). Resultado semelhante foi reportado por González-Recio et al. (2006), investigando período de serviço em gado Holandês, em que obtiveram -0,74 de correlação de Pearson entre valores genéticos dos touros, preditos pelos modelos ML e MW.

A porcentagem de coincidência na classificação dos 10% melhores touros foram, respectivamente, em torno de 75% e 54% entre os modelos ML-MLC e ML-MW, considerando todos os reprodutores, e 75% para ML-MLC e ML-MW, quando apenas touros com mais de 10 progênes foram selecionados (Tabela 4.4). Logo, existe evidência de reclassificação dos animais quanto ao mérito genético entre os modelos ML, MLC e MW. Donoghue, Rekaya e Bertrand (2004b), reportaram resultado similar para os modelos ML e MLC, sendo o TOP10 obtido igual a 62%. Da mesma forma, González-Recio et al. (2006), relataram mudança na classificação de animais para ML-MLC, entretanto a correspondência observada entre MLC e MW foi elevada.

Tabela 4.4 – Correlações de Pearson e porcentagem de coincidência na classificação dos 10% melhores reprodutores, entre os valores genéticos preditos de todos os touros (acima da diagonal) e de touros com mais de 10 progênes não censuradas (abaixo da diagonal) utilizando distintos modelos de avaliação genética para longevidade produtiva^a

	Correlações de Pearson			Porcentagem de Coincidência		
	ML	MLC	MW	ML	MLC	MW
ML		0,82	-0,68		74,60	53,97
MLC	0,87		-0,89	74,29		71,43
MW	-0,74	-0,91		74,29	77,14	

^a 1.256 e 352 touros;

^b ML: modelo linear Gaussiano;

MLC: modelo linear Gaussiano com censura;

MW: modelo de riscos proporcionais Weibull.

O modelo de riscos proporcionais Weibull apresentou melhor habilidade de predição em termos médios e o menor valor da estatística χ^2 em todas as repetições da

validação (Tabela 4.5). Em contrapartida, o modelo linear Gaussiano obteve a menor habilidade preditiva, em torno de 4% menos eficiente do que MW, e o modelo Gaussiano com censura a segunda menor média de predição das categorias de longevidade produtiva. Resultado semelhante foi reportado por Hou et al. (2009), os quais relataram que o modelo MW apresentou menor habilidade de predição em relação ao modelo MLC, em categorias de dias entre a primeira inseminação e o parto em gado de leite. Da mesma forma, González-Recio et al. (2006), reportaram uma maior eficiência do modelo MW para prever prenhez em dois estágios iniciais de lactação, entretanto, os modelos ML e MLC, apresentaram melhores resultados para a descrição da prenhez em dois estágios finais do período de lactação.

Tabela 4.5 – Média da estatística χ^2 , eficiência e o número de vezes em que os distintos modelos de avaliação genética de longevidade produtiva apresentaram o menor valor de χ^2 (NM χ^2), num esquema de validação cruzada com 10 repetições

Modelo ^a	Média χ^2	Eficiência ^b	NM χ^2
ML	1947,743	1,037	0
MLC	1925,802	1,026	0
MW	1877,488	1,000	10

^a ML: modelo linear Gaussiano;

MLC: modelo linear Gaussiano com censura;

MW: modelo de riscos proporcionais Weibull

^b Eficiência calculada como a razão da média de cada modelo em relação a menor média de χ^2 observada

Os modelos linear Gaussiano, linear Gaussiano com censura e de riscos proporcionais Weibull apresentaram diferenças, em termos de predição dos valores genéticos dos reprodutores, denotadas pelas correlações de Pearson e TOP10, sendo que, uma menor discrepância entre MLC e MW foi observada. Em termos de habilidade de predição, o modelo MW foi o mais eficiente, em termos médios da estatística χ^2 e apresentou melhor habilidade preditiva nas 10 repetições do esquema de validação. Logo, o modelo de riscos proporcionais Weibull seria o mais indicado para a avaliação genética de longevidade produtiva nesta população.

4.4 CONCLUSÕES

A inclusão de registros censurados não acarretou mudança nas estimativas de herdabilidade nos modelos lineares, entretanto, estes modelos foram distintos em termos de eficiência preditiva das categorias de longevidade produtiva das progênes dos reprodutores. O modelo MW obteve a melhor habilidade de predição em 100%

das repetições do processo de validação. Portanto, para a avaliação genética da longevidade produtiva, nesta população, seria mais adequada a utilização do modelo de riscos proporcionais Weibull.

REFERÊNCIAS

- ALBERT, J. H.; CHIB, S. Bayesian analysis of binary and polychotomous response data. **Journal of American Statistical Association**, v. 88, n. 422, p. 669–679, 1993.
- ALLAIRE, F. R.; GIBSON, J. P. Genetic value of herd life adjusted for milk production. **Journal of Dairy Science**, v. 75, n. 5, p. 1349–1356, 1992.
- ARANGO, J.; MIZSTAL, I.; TSURUTA, S.; CULBERTSON, M.; HERRING, W. Study of codes of disposal at different parities of Large White sows using a linear censored model. **Journal of Animal Science**, v. 83, p. 2052–2057, 2005.
- ARNASON, T. Genetic evaluation of Swedish standard-bred trotters for racing performance traits and racing status. **Journal Animal Breeding and Genetics**, v. 116, p. 387–398, 1999.
- AZEVEDO, D. M. M. R.; FILHO, R. M.; LÔBO, R. N. B.; MALLHADO, C. H. M.; LÔBO, R. B.; MOURA, A. A. A.; PIMENTAFILHO, E. C. Desempenho reprodutivo de vacas nelore no norte e nordeste do Brasil. **Revista Brasileira de Zootecnia**, v. 35, n. 3, p. 988–996, 2006.
- BALDI, F.; ALENCAR, M. M.; FREITAS, A. R.; BARBOSA, R. T. Parâmetros genéticos para características de tamanho e condição corporal, eficiência reprodutiva e longevidade de fêmeas da raça canchim. **Revista Brasileira de Zootecnia**, v. 37, n. 2, p. 247–253, 2008.
- BERTAZZO, R. P.; FREITAS, R. T. F.; GONÇALVES, T. M.; PEREIRA, I. G.; ELER, J. P.; FERRAZ, J. B. S.; OLIVEIRA, A. I. G.; ANDRADE, I. F. Parâmetros genéticos de longevidade e produtividade de fêmeas da raça nelore. **Revista Brasileira de Zootecnia**, v. 33, n. 5, p. 1118–1127, 2004.
- BOLDMAN, K. G.; FREEMAN, A. E.; HARRIS, B. L. Prediction of sire transmitting abilities for herd life from transmitting abilities for linear type traits. **Journal of Dairy Science**, v. 75, n. 2, p. 552–563, 1992.
- BOLIGON, A. A.; ALBUQUERQUE, L. G.; RORATO, P. R. N. Associações genéticas entre pesos e características reprodutivas em rebanhos da raça nelore. **Revista Brasileira de Zootecnia**, v. 37, n. 4, p. 596–601, 2008.
- BOLIGON, A. A.; RORATO, P. R. N.; ALBUQUERQUE, L. G. Correlações genéticas entre medidas de perímetro escrotal e características produtivas e reprodutivas de fêmeas da raça nelore. **Revista Brasileira de Zootecnia**, v. 36, n. 3, p. 565–571, 2007.
- BOLIGON, A. A.; VOZZI, P. A.; RORATO, J. N. P. R. N.; BEZERRA, L. A. F.; LÔBO, R. B. Parâmetros genéticos para idade ao primeiro parto estimados por diferentes modelos para rebanhos da raça nelore. **Ciência Rural**, v. 38, n. 2, p. 432–436, 2008.
- CAETANO, S. L. **Estudo da idade da vaca ao último parto para avaliar longevidade em rebanhos da raça Nelore por análise de sobrevivência**. 107 p. Tese (Genética e Melhoramento Animal) — Universidade Estadual Paulista Júlio de Mesquita Filho, 2011.

CARDOSO, F. F.; ROSA, G. J. M. R.; TEMPELMAN, R. J.; TORRES JUNIOR, R. A. A. Modelos hierárquicos bayesianos para estimação robusta e análise de dados censurados em melhoramento animal. **Revista Brasileira de Zootecnia**, v. 38, n. 2, p. 72–80, 2009. Suplemento especial.

CHANG, Y. M.; ANDERSEN-RANBERG, I. M.; HERINGSTAD, B.; GIANOLA, D.; KLEMETSDAL, G. Bivariate analysis of number of services to conception and days open in Noregina Red using censored threshold-linear model. **Journal of Dairy Science**, v. 89, n. 1, p. 772–778, 2006.

COX, D. Regression models and life table (with discussion). **Journal of the Royal Statistics Society**, v. 34, p. 187–220, 1972.

DIAS, L. T.; EL FARO, L.; ALBUQUERQUE, L. G. Estimativas de herdabilidade para idade ao primeiro parto de novilhas da raça Nelore. **Revista Brasileira de Zootecnia**, v. 33, n. 1, p. 97–102, 2004.

DONOGHUE, K. A.; REKAYA, R.; BERTRAND, J. K. Comparison methods for handling censored records in beef fertility data: Simulation study. **Journal of Animal Science**, v. 82, n. 2, p. 351–356, 2004a.

DONOGHUE, K. A.; REKAYA, R.; BERTRAND, J. K. Comparison methods for handling censored records in beef fertility data: Field study. **Journal of Animal Science**, v. 82, n. 2, p. 357–361, 2004b.

DONOGHUE, K. A.; REKAYA, R.; BERTRAND, J. K.; MISZTAL, I. Threshold-linear analysis of measures of fertility in artificial insemination data and days to calving in beef cattle. **Journal of Animal Science**, v. 82, n. 2, p. 987–993, 2004c.

DUCROCQ, V. Statistical analysis of length of productive life for dairy cows of the Normande breed. **Journal of Dairy Science**, v. 77, p. 855–866, 1993.

DUCROCQ, V.; CASELLA, G. A Bayesian analysis of mixed survival models. **Genetic Selection Evolution**, v. 28, p. 505–529, 1996.

DUCROCQ, V.; QUASS, R. L.; POLLAK, E. J.; CASELLA, G. Length of productive life of dairy cows: Justification of a Weibull model. **Journal of Dairy Science**, v. 71, p. 3061–3070, 1988.

DUCROCQ, V.; SOLKNER, J.; MÉSZÁROS, G. **The Survival kit v6.0: User's Manual**. [S.l.], 2010. 83 p.

ELER, J. P.; SILVA, J. A. I. V.; FERRAZ, J. B. S.; et al. Genetic evaluation of the probability of pregnancy at 14 months for Nellore heifers. **Journal of Animal Science**, v. 80, p. 951–954, 2002.

ELER, J. P.; SILVA, J. A. I. V.; FERRAZ, J. B. S.; et al. Additive genetic relationships between heifer pregnancy at 14 months for Nellore cattle. **Journal of Animal Science**, v. 82, p. 2519–2527, 2004.

- FARIA, C. U.; MAGNABOSCO, C. U.; REYES, A.; LÔBO, R. B.; BEZERRA, L. A. F.; SAINZ, R. D. Bayesian inference on field data for genetics parameters for some reproductive traits and related traits of Nelore cattle. **Genetics and Molecular Biology**, v. 30, n. 2, p. 343–348, 2007.
- FORNI, S.; ALBUQUERQUE, L. G. Estimates of genetics correlation between days to calving and reproductive and weight traits in Nelore cattle. **Journal of Animal Science**, v. 83, p. 1511–1515, 2005.
- FORNI, S.; ALBUQUERQUE, L. G. Avaliação de fatores de ambiente e estimativas de parâmetros genéticos para a característica dias para o parto na raça Nelore. **Revista Brasileira de Zootecnia**, v. 35, n. 4, p. 1239–1335, 2006.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, p. 457–511, 1992.
- GIANLORENÇO, V. k.; ALENCAR, M. M.; MELLO, F. L. B. T. S. P.; FREITAS, A. R.; BARBOSA, P. F. Herdabilidades e correlações genéticas de características de machos e fêmeas, em um rebanho bovino da raça Canchim. **Revista Brasileira de Zootecnia**, v. 32, n. 6, p. 1587–1593, 2003.
- GIANOLA, D.; FERNANDO, R. L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v. 63, n. 1, p. 217–244, 1986.
- GONZÁLEZ-RECIO, O.; CHANG, Y. M.; GIANOLA, D.; WEIGEL, K. A. Comparison of models using different censoring scenarios for days open in spanish holstein cows. **British Society of Animal Science**, v. 82, p. 223–239, 2006.
- GUO, S. F.; GIANOLA, D.; SHORT, R. R. T. Bayesian analysis of lifetime performance and prolificacy in Landrace sows using a linear mixed model with censoring. **Livestock Production Science**, v. 72, p. 243–252, 2001.
- HELDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, v. 31, p. 1109–1144, 1983.
- HOU, Y.; MADSEN, P.; LABOURIAU, R.; ZHANG, Y.; LUND, M. S.; SU, G. Genetic analysis of days from calving to first insemination and days open in Danish Holstein using different models and censoring scenarios. **Journal of Dairy Science**, v. 92, p. 1229–1239, 2009.
- JOHNSTON, D. J.; BUNTER, K. L. Days to calving in Angus cattle: Genetic and environmental effects, and covariances with other traits. **Livestock Production Science**, v. 45, n. 1, p. 13–22, 1996.
- KORSGAARD, I. R.; LUND, M. S.; SORENSEN, D.; GIANOLA, D.; JENSEN, P. M. J. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. **Genetics Selection Evolution**, v. 35, p. 159–183, 2003.
- MCCULLAGH, P.; NELDER, J. A. **Generalized linear models**. Second. London New York: Chapman and Hall, 1989. 511 p.

MELLO, S. P.; ALENCAR, M. M.; TORAL, F. L. B.; GIANLORENÇO, V. K. Estimativas de parâmetros genéticos para características de crescimento e produtivas em vacas da raça Canchim, utilizando-se inferência bayesiana. **Revista Brasileira de Zootecnia**, v. 35, n. 1, p. 92–97, 2006.

MERCADANTE, M. E. Z. **Estudo das relações genético-quantitativas entre características de reprodução, crescimento e produção em fêmeas da raça Nelore**. 95 p. Tese (Mestrado em Genética) — Universidade de São Paulo, 1995.

MERCADANTE, M. E. Z.; LÔBO, R. B.; OLIVEIRA, H. N. Estimativas de (co)variância entre características de reprodução e crescimento em fêmeas de um rebanho Nelore. **Revista Brasileira de Zootecnia**, v. 29, n. 4, p. 997–1004, 2000.

MERCADANTE, M. E. Z.; PARCKER, I. U.; RAZZOK, A. G.; CYRILLO, J. N. S. G.; FIGUEIREDO, L. A. Dias ao parto de fêmeas Nelore de um experimento de seleção para crescimento: I - Modelo de repetibilidade. **Revista Brasileira de Zootecnia**, v. 31, n. 4, p. 1715–1725, 2002.

MEYER, K.; HAMMOND, K.; PARNELL, P. F. Estimates of heritability and repeatability for reproductive trait in Australian beef cattle. **Livestock Production Science**, v. 25, p. 15–30, 1990.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; et al. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 2002, Montpellier. **Proceeding...** Montpellier, 2002.

NEVES, H. H. R.; CARVALHEIRO, R.; QUEIROZ, S. A. Genetic parameters for an alternative criterion to improve productive longevity of Nellore cows. **Journal of Animal Science**, v. 90, p. 4209–4216, 2012.

NEWMAN, S.; MORRIS, C. A.; BAKER, R. L. Genetic improvement of beef cattle in New Zealand: breeding objectives. **Livestock Production Science**, v. 32, p. 111–130, 1992.

NOTTER, D. R. Evaluating and reporting reproductive traits. In: BEEF IMPROVEMENT FEDERATION, 1988, Albuquerque. **Proceeding...** Albuquerque, 1988. p. 21–42.

PEREIRA, E.; ELER, J. P.; FERRAZ, J. B. S. Correlações genéticas entre perímetro escrotal e algumas características reprodutivas na raça Nelore. **Revista Brasileira de Zootecnia**, v. 29, n. 6, p. 1676–1683, 2000.

PEREIRA, E.; ELER, J. P.; FERRAZ, J. B. S. Análise genética de características reprodutivas na raça Nelore. **Pesquisa Agropecuária Brasileira**, v. 37, n. 5, p. 703–708, 2002.

PEREIRA, E.; OLIVEIRA, H. N.; ELER, J. P.; SILVA, J. A. I. V.; VAN MELIS, M. H. Use of survival analysis as a tool for the genetic improvement of age at first conception in Nellore cattle. **Journal Animal Breeding and Genetics**, v. 123, p. 64–71, 2005.

PEREIRA, E.; OLIVEIRA, H. N.; ELER, J. P.; ; SILVA, J. A. I. V.; VAN MELIS, M. H. Comparison among three approaches for evaluation of sexual precocity in Nellore cattle. **The Animal Consortium**, v. 1, p. 411–418, 2007.