

Universidade Federal do Piauí-UFPI
Pós-Graduação em Ciência Animal

**MODELAGEM DE FRAGILIDADE PARA AVALIAÇÃO DA
LONGEVIDADE DE BOVINOS DA RAÇA CURRALEIRO
PÉDURO NO ESTADO DO PIAUÍ**

Keliny Martins de Melo Sousa Soares

Teresina
2021

Keliny Martins de Melo Sousa Soares

**MODELAGEM DE FRAGILIDADE PARA AVALIAÇÃO DA
LONGEVIDADE DE BOVINOS DA RAÇA CURRALEIRO PÉ-
DURO NO ESTADO DO PIAUÍ**

Tese apresentada ao Programa de Pós-Graduação em Ciência Animal, na área de Produção Animal, como parte dos requisitos para obtenção do título de Doutor.

Área de concentração: Produção Animal.

Orientador: Prof. Dr. Severino Cavalcante de Sousa Junior.

Co-orientadora: Prof(a). Dra. Vera Lúcia Damasceno Tomazella.

Co-orientador: Prof. Dr. Geraldo Magela Cortes Carvalho.

**Teresina
2021**

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Serviço de Processamento Técnico

S676m Soares, Keliny Martins de Melo Sousa.
Modelagem de fragilidade para avaliação da longevidade de bovinos da raça Curraleiro Pé-Duro no Estado do Piauí / Keliny Martins de Melo Sousa Soares. – 2021.
101 f.

Tese (Doutorado) – Programa de Pós-Graduação em Ciência Animal da Universidade Federal do Piauí, 2021.
“Orientador: Prof. Dr. Severino Cavalcante de Sousa Junior.
Co-orientadora: Prof(a). Dra. Vera Lúcia Damasceno Tomazella.
Co-orientador: Prof. Dr. Geraldo Magela Cortes Carvalho.”

1. Gado Bovino – Raça Pé-Duro. 2. Fragilidade Gaussiana inversa 3. Ciência Animal. 4. Bovinos – Estudo da Longevidade.
I. Título.

CDD 636.2

Bibliotecária: Milane Batista da Silva – CRB-3/1005

MODELAGEM DE FRAGILIDADE PARA AVALIAÇÃO DA LONGEVIDADE
DE BOVINOS DA RAÇA PÉ-DURO DO ESTADO DO PIAUÍ

KELINY MARTINS DE MELO SOUSA SOARES

Tese aprovada em: 10/09/2021

Banca Examinadora:



Prof. Dr. Severino Cavalcante de Sousa Junior (Presidente) / UFDPAr



Profa. Dra. Vera Lúcia Damasceno Tomazella (Externa) / UFSCar



Pesquisador Dr. Geraldo Magela Cortes Carvalho (Exam. Externo) / EMBRAPA



Profa. Dra. Karina Rodrigues dos Santos (Examinadora Externa) / UFDPAr



Prof. Dr. Marcelino Alves Rosa de Páscoa (Examinador Externo) / UFMT



Prof. Dr. Marcelo Richelly Alves de Oliveira (Externo) / IESM

DEDICATÓRIA

Dedico este trabalho primeiramente a Deus, por ser essencial em minha vida, autor do meu destino, meu guia, socorro presente na hora da angústia, ao meu esposo Clay Revson, minha filha Letice e aos meus pais Waldinar e Edenir pela paciência, amor e por serem colunas em minha vida.

AGRADECIMENTOS

Agradeço, dedico, ofereço, em primeiro lugar a Deus, por ter me dado força, saúde, perseverança, humildade e sabedoria ao longo dessa caminhada.

Ao meu esposo, Clay Revson, por tanto amor, contribuição emocional, paciência, cumplicidade e companheirismo dedicado de forma integral. Você foi essencial em todo essa jornada. Te amo! A minha filha Letice Martins, pelo amor, renovação diária de força, minha fonte de inspiração e por tudo que você representa para mim. Tudo isso é por você filha! Vocês são as minhas colunas de sustentação. Amor incondicional.

Àos meus pais, Waldinar de Sousa e Edenir Martins, por serem o meu porto seguro em muitos momentos da minha caminhada. Ao meu pai, uma pessoa simples e perseverante, que sempre incentivou e colocou a nossa formação profissional (minha e das minhas irmãs Keylla e Kennya e irmão Joabe) em primeiro lugar. À minha mãe, meu espelho de vida, de mãe, mulher, amiga, filha, por sua incansável generosidade me possibilitando aprender a cada dia a concretude do amor. Palavras me faltam para expressar toda a minha gratidão.

Aos meus avós, Nilo Martins e Joelina Melo (in memoriam) que estariam vibrando com mais essa minha conquista. Como vocês fazem falta, mas sei que tudo que tenho hoje é fruto da oração de vocês. Muitas saudades!!!

Às minhas irmãs, Keylla Martins (minha pastora), Kennya Martins (minha psicóloga), Simone Rodrigues e irmão Joabe Martins por tantas histórias e pelas diversas demonstrações de amor em cada palavra, gesto, silêncio, carinho, confiança, preocupação e dedicação.

As minhas sobrinhas Raquel Yasmim, Bianca Vida, Sofia Melo, Pietra Martins e meu sobrinho Nilo Melo, por me proporcionar tantos momentos ímpares de alegria.

Ao meu orientador Severino Junior, obrigada pela compreensão, paciência, disponibilidade, e orientação na elaboração deste trabalho.

A minha co-orientadora Vera Tomazella, por toda a experiência, serenidade, sabedoria, amizade e paciência. Sua participação foi imprescindível para a realização deste trabalho.

Ao meu co-orientador Geraldo Magela por toda experiência, disponibilidade e paciência no entendimento dos bovinos da raça Curraleiro Pé-Duro.

Aos professores da pós-graduação em Ciência Animal e em Genética e Melhoramento pelas valiosas orientações e contribuições na minha formação;

Ao secretário Luis Gomes do PPGCA, pela competência, dedicação e atenção.

À minha colega de trabalho e pós-graduação Cleide Mayra, pela amizade, carinho, paciência e por sempre estar disposta a me ajudar. A sua ajuda foi imprescindível para a realização deste trabalho.

Ao Grupo de estudos em genética e melhoramento animal (GEMA), pelos momentos de aprendizado e descontração.

Aos colegas de trabalho do curso de Estatística da Universidade Federal do Piauí, que torceram para realização desta conquista.

À minha colega de profissão, Anísia que me acolheu prontamente em São Carlos. Obrigada pela amizade, carinho e hospitalidade. Você foi um presente que o doutorado me deu.

A todos aqueles que de alguma maneira contribuíram ou estiveram na torcida pela realização deste trabalho.

Muito Obrigada!!!

SUMÁRIO

LISTA DE FIGURAS	9
LISTA DE TABELAS	10
RESUMO	11
ABSTRACT	12
1 INTRODUÇÃO	14
2 REVISÃO DE LITERATURA	19
2.1 Conceitos básicos em análise de sobrevivência	19
2.2 Alguns Modelos Paramétricos de Sobrevivência.....	26
2.3 A função de verossimilhança em análise de sobrevivência	31
2.4 Modelos de Riscos Proporcionais de Cox	32
2.5 O Modelo de Fragilidade	40
3 REFERÊNCIAS	42
4 CAPÍTULO 1	47
Modelo de Cox- Log Normal para avaliar a longevidade em bovinos da raça Curraleiro Pé-Duro.....	48
Abstract	48
Resumo	48
Introdução	49
Material e Métodos	51
Resultados.....	55
Discussões	61
Conclusão	62
Referências	63
5 CAPÍTULO 2	65
Avaliação da longevidade de bovinos da raça Curraleiro Pé-Duro usando o modelo de fragilidade Gaussiana inversa	66
Resumo	66
Introdução	66
Material e Métodos	69
Resultados.....	75
Discussões	79
Conclusão	81

Referências	81
6 CAPÍTULO 3	84
Modelagem de fragilidade Gaussiana inversa compartilhada aplicada ao estudo da longevidade de bovinos da raça Curraleiro Pé-Duro	85
Resumo	85
Introdução	85
Material e Métodos	88
Resultados e Discussões	94
Conclusão	99
Referências	100
7 CONSIDERAÇÕES GERAIS E PROPOSTAS FUTURAS	
7.1 Considerações gerais	101
7.2 Propostas Futuras	101

LISTA DE FIGURAS**INTRODUÇÃO**

Figura 1	Curraleiro Pé-Duro	15
----------	--------------------------	----

REFERÊNCIAL TEÓRICO

Figura 1	(a) Dados completos, (b) Dados com censura tipo I, (c) Dados com censura tipo II, (d) Dados com censura aleatória.....	18
Figura 2	Gráficos ilustrativos de algumas curvas TTT	21
Figura 3	Formas das funções de densidade, sobrevivência e de risco da distribuição Log Normal para alguns valores de μ e σ	25
Figura 4	Formas das funções de densidade, sobrevivência e de risco da distribuição Gaussiana inversa para alguns valores de a e b.....	26

CAPITULO 1

Figura 1	Curva de sobrevivência geral estimada pelo método de Kaplan-Meier.....	52
Figura 2	Curvas de sobrevivências estimadas pelo método de Kaplan-Meier para as possíveis covariáveis: Estação de Nascimento (EN), Sexo (S), Peso ao Nascer (PN), a Desmama (PD), aos 365 dias (P365) e aos 550 dias (P550) dos bovinos da raça Curraleiro Pé-Duro.....	53
Figura 3	Logaritmo da taxa de falha versus tempo para as covariáveis EN, Sexo e P365... ..	55
Figura 4	Resíduos de Cox-Snell , Martingal e Deviance do modelo de Cox-Log Normal para dados de bovinos da raça Curraleiro Pé-Duro.....	56

CAPITULO 2

Figura 1	Fragilidade Gaussiana inversa univariada com risco base log-normal.....	73
Figura 2	Resíduos de Cox-Snell, Martingal e Deviance do modelo de fragilidade Gaussiana inversa- Log Normal para dados de bovinos da raça Curraleiro Pé-Duro.....	74

CAPITULO 3

Figura 1	Resíduos de Cox-Snell , Martingal e Deviance do modelo de fragilidade Gaussiana Inversa-Log Normal para dados de bovinos da raça Curraleiro Pé-Duro.....	95
----------	--	----

LISTA DE TABELAS

CAPITULO 1

Tabela 1	Funções de densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_o(t)$ das distribuições Gompertz, Weibull e log-Normal.....	50
Tabela 2	Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), p -valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de Cox- Log Normal	54
Tabela 3	Estimativas para os parâmetros do modelo de riscos proporcionais de Cox- Log Normal, Cox-Log logístico e Cox-Gompertz	56

CAPITULO 2

Tabela 1	Número e porcentagem de bovinos que falharam ou foram censurados por variável	66
Tabela 2	Funções de densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_o(t)$ das distribuições Log Normal, Log-logístico e Gompertz.....	69
Tabela 3	Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), p -valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de fragilidade univariada Gaussiana inversa- Log Normal	72
Tabela 4	Estimativas para os parâmetros dos modelos de fragilidade Gaussiana inversa- Log Normal e Gussiana inversa-Log-logístico	75

CAPITULO 3

Tabela 1	Número e porcentagem de bovinos por variável estudada.....	85
Tabela 2	Funções de densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_o(t)$ das distribuições Log Normal, Log-logístico e Gompertz	88
Tabela 3	Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), p -valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de fragilidade compartilhada Gaussiana inversa- Log Normal	92
Tabela 4	Estimativa das fragilidades (Z_i) associadas a grupos de bovinos que nasceram no mesmo ano (2005 a 2014)	93
Tabela 5	Resultados para os parâmetros do modelo de fragilidade compartilhada Gaussiana inversa com funções de risco base Log Normal e Log-logístico	94

RESUMO

SOARES, M. M. S. Keliny. Modelagem de fragilidade para avaliação da longevidade de bovinos da raça Curraleiro Pé-Duro no Estado do Piauí. 2021. 97p. Tese de Doutorado (Programa de Pós Graduação em Ciência Animal - PPGCA-UFPI/CCA).

O Curraleiro Pé-Duro foi selecionado na região do estado do Piauí em meados do ano 1674. São animais gradativamente ambientados às pastagens nativas de baixa qualidade, às condições de estiagem, às temperaturas elevadas entre outros fatores adversos, resultando na formação de rebanhos bovinos adaptados e resistentes às condições desfavoráveis do sertão brasileiro. Nesse contexto, o objetivo geral desse trabalho, foi apresentar a metodologia de análise de sobrevivência como alternativa para análises em estudos genéticos relacionados a bovinos da raça Curraleiro Pé- Duro, no contexto em que a variável resposta está sendo influenciada por fatores não observáveis (latentes). Esta metodologia pode combinar registros completos de animais que não estão mais no rebanho ou que completaram seu ciclo de vida útil e dados censurados de animais ainda em reprodução. Para esse estudo, foram consideradas os registros de de bovinos no período de 2005 a 2014, cedidos pelo núcleo de conservação *in situ* de bovinos da raça Curraleiro Pé-Duro pertencentes a Embrapa Meio-Norte, em Teresina-Piauí, com campo experimental situado na fazenda Otávio Domingues, em São João do Piauí, no Semiárido piauiense. Foram avaliados 102 bovinos (58 Machos e 44 fêmeas) desde o nascimento até aos 550 dias. As características avaliadas, presentes no banco de dados foram: estação de nascimento, sexo, peso ao nascer, a desmama, aos 365 dias e aos 550 dias. O modelo de riscos proporcionais de Cox-Log Normal foi utilizado para verificar o efeito de fatores que possam influenciar no tempo permanência de bovinos no rebanho sendo a falha definida pela inatividade ocasionada por morte ou venda. Esse modelo considera que os tempos de vida dos animais são independentes e que o riscos de falhas são proporcionais. Para considerar a heterogeneidade não observável (como fatores genéticos ou ambientais) entre os animais e a

correlação entre os grupos de animais, foi utilizado o modelo de fragilidade Gaussiana inversa univariada e compartilhada. Para as análises estatísticas foi utilizado o *software* estatístico *R*.

Palavras-chave: Adaptabilidade, Cox- Log Normal, fragilidade Gaussiana inversa, Sobrevivência.

ABSTRACT

SOARES, M. M. S. Keliny. Frailty modeling to assess the longevity of Curraleiro Pé-Duro cattle in the state of Piauí. 2021. 97p. Doctoral Thesis (Postgraduate Program in Animal Science - PPGCA-UFPI/CCA).

Curraleiro Pé-Duro was selected in the region of the State of Piauí in the middle of the year 1674. They are animals gradually adapted to low quality native pastures, dry conditions, high temperatures and other adverse factors, resulting in the formation of adapted bovine herds and resistant to the unfavorable conditions of the Brazilian hinterland. In this context, the general objective of this work was to present the survival analysis methodology as an alternative for analyzes in genetic studies related to Curraleiro Pé-Duro cattle, in the context in which the response variable is being influenced by unobservable factors (latent). This methodology can combine complete records from animals that are no longer in the herd or that have completed their life cycle and censored data from animals still in breeding. For this study, the records of cattle from 2005 to 2014 were considered, provided by the nucleus of in situ conservation of cattle of the Curraleiro Pé-Duro breed belonging to Embrapa Meio-Norte, in Teresina-Piauí, with an experimental field located in Otávio Domingues farm, in São João do Piauí, in the semiarid region of Piauí. 102 cattle (58 males and 44 females) were evaluated from birth to 550 days. The characteristics evaluated, present in the database are: season of birth, sex, birth weight, weaning, 365 days and 550 days. The Cox-Log Normal proportional hazards model was used to verify the effect of factors that may influence the length of stay of cattle in the herd, the failure being defined

by inactivity caused by death or sale. This model considers that the lifespans of animals are independent and that the risk of failure is proportional. To consider the unobservable heterogeneity (such as genetic or environmental factors) among animals and the correlation between groups of animals, the univariate and shared inverse Gaussian frailty model was used. For statistical analysis, the R statistical software was used.

Keywords: Adaptability, Cox- Log Normal, inverse Gaussian fragility, survival.

1 INTRODUÇÃO

A raça Curraleiro Pé-Duro, reconhecida pelo Ministério da Agricultura, Pecuária e Abastecimento em 2012, é a primeira raça selecionada naturalmente no Brasil, formada de bovinos trazidos de Portugal a partir do século XV. Essa raça local, também denominada nativa, naturalizada ou crioula, é patrimônio histórico e cultural do Piauí e auxiliou o desenvolvimento econômico do país como meio de transporte, fonte de proteína e comércio de charque e couro (Salles et al., 2011). A atenção dos pesquisadores pela referida raça ocorreu a partir da observação de como estes bovinos gradativamente se ambientaram às pastagens nativas de baixa qualidade, às condições de estiagem, às temperaturas elevadas entre outros fatores adversos à criação de bovinos no Brasil. Esses mecanismos de adaptação ocorreram ao longo dos séculos resultando na formação de rebanhos bovinos adaptados e resistentes às condições desfavoráveis do sertão brasileiro (Carvalho et al., 2001). A habilidade dos animais em adaptar-se às condições climáticas extremas em ambiente natural proporciona a minimização de perdas no desempenho produtivo e reprodutivo, auxilia na resistência a doenças, conseqüentemente reduz a mortalidade e aumenta a longevidade quando expostos a condições de estresse (Baccari Júnior, 1990).

A Análise de Sobrevivência (AS), por sua vez, é considerada metodologia estatística mais adequada para lidar com dados de tempo até a ocorrência de um evento de interesse (tempo de falha), na presença de censura (Colosimo e Giolo, 2006) que é sua principal característica. Na área de produção animal, de acordo com Carvalho Dias (1997), as características de sobrevivência são: longevidade (idade do animal ao descarte); duração da vida produtiva (intervalo do primeiro parto ao descarte); número de lactações durante a vida e habilidade de permanência no rebanho ou *stayability*. A metodologia de análise de sobrevivência, já vem sendo utilizada por alguns autores no estudo da longevidade em bovinos leiteiros. Bonetti et al. (2009) estimaram parâmetros genéticos numa avaliação genética para a longevidade de touros Italianos da raça Pardo-Suíça, por meio do modelo de riscos proporcionais de Weibull. Os autores consideraram o método satisfatório

para o uso e a inclusão dos touros nos programas de melhoramento genético. Caetano et al. (2013) propuseram a idade da vaca ao último parto (IVUP) como medida para avaliar a habilidade de permanência da vaca no rebanho. Os autores concluíram que a variável é relevante para avaliar a habilidade de permanência das vacas no rebanho e que o modelo de análise de sobrevivência foi o que estimou maior proporção da variabilidade genética para a característica estudada. Outros autores que utilizaram esta metodologia foram: Giolo (2003), Van Melis et al. (2010) e Kern et al. (2016).

Uma característica especial associada aos dados de sobrevivência é a possibilidade de que, para alguns indivíduos, o tempo completo até a ocorrência do evento de interesse não seja observado, devido a diversas causas. Por exemplo, um indivíduo pode permanecer em acompanhamento até o final do tempo de estudo sem ter experimentado o evento de interesse. Em experimentos médicos, a informação sobre o tempo de vida de um paciente pode ser incompleta porque este simplesmente decide abandonar o acompanhamento antes do término da pesquisa, ou então, falha devido a outras causas não relacionadas ao evento de interesse em estudo. A não consideração destes indivíduos com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes (Bolfarine, 1991). Portanto, a importância da introdução de uma variável na análise que indique se o tempo de sobrevivência foi observado (Louzada-Neto e Pereira, 2000). Essa variável é definida na literatura como variável indicadora de censura ou simplesmente “censura”.

A censura é dita ser do tipo I quando ocorre devido ao término do estudo após um período de tempo pré-determinado; do tipo II, quando ocorre devido ao término do estudo após um número de falhas fixado previamente; ou pode ser aleatória, sendo a mais comum em situações práticas. As censuras previamente citadas são conhecidas como censura à direita, pois a falha ocorre sempre à direita do tempo registrado. Existem ainda outros mecanismos de censura, como, por exemplo, as censuras à esquerda, em que o tempo registrado é maior que o tempo de falha, e censura

intervalar, na qual não se sabe o tempo exato de falha, sendo que a única informação disponível é que o tempo de falha ocorreu em um certo intervalo de tempo (Figura 1). Nesta pesquisa, entretanto, será adotado apenas o mecanismo de censura à direita que será denominado por “censura”.

Em estudos recentes há situações em que a variável resposta, tempo de falha, pode estar sendo influenciado por fatores não observáveis, denominado fatores latentes. Modelos de sobrevivência com variáveis latentes ou modelos com fragilidade, são caracterizados pela inclusão de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou que não foram observadas; como fatores ambientais, genéticos ou informações que por algum motivo não foram considerados no planejamento. Uma das formas encontradas para incorporar esse efeito aleatório, denominado de variável de fragilidade, é introduzi-lo na modelagem da função de risco, com o objetivo de controlar a heterogeneidade não observável das unidades em estudo. A fragilidade pode ser inserida no modelo de forma aditiva ou multiplicativa, com o objetivo de avaliar a heterogeneidade entre as unidades na função de risco ou a dependência para dados multivariados. Em estudos com animais, as associações aparecem devido a influências genéticas ou ambientais compartilhadas e se forem ignoradas, inferências incorretas poderão ser obtidas.

Diversos autores estudaram a aplicação dos modelos de fragilidade multiplicativos, que representam uma generalização dos modelos de riscos proporcionais de Cox (Cox, 1972). Andersen et al. (2012) e Hougaard (1995) apresentaram uma revisão dos modelos de fragilidade multiplicativos numa perspectiva clássica, enquanto, Sinha e Dey (1997) elaboraram uma revisão completa desses modelos sob o ponto de vista Bayesiana. O termo de fragilidade foi introduzido por Vaupel et al. (1979) em modelos de sobrevivência com dados univariados. Clayton (1978) e Oakes (1982) trabalharam com modelos de sobrevivência multivariados. Devido as características da fragilidade no modelo de fragilidade multiplicativo, as candidatas naturais à distribuição da

fragilidade, supostamente contínuas e não dependentes do tempo, são as distribuições gama, log-normal, Gaussiana inversa e Weibull. Hougaard (1984) foi um dos primeiros autores a abordar o impacto do uso de diferentes distribuições para a variável de fragilidade.

1.1 Motivação

O Curraleiro Pé-Duro foi a primeira raça bovina a ser efetivamente formada no Brasil a partir de raças portuguesas trazidas por colonizadores nos séculos XV (SALLES et al., 2011). Foi introduzida na região do estado do Piauí, a partir do Rio São Francisco por Domingos Afonso Mafrense, em meados do ano 1674 (Nogueira Neto, 1980), resultando posteriormente em adaptação dos bovinos em condições ambientais da região (Carvalho, 1984). Esta raça, se apresenta como um taurino tropicalmente adaptado para ser usado pelo agronegócio em regiões quentes do Brasil. É indicado como raça pura para produção de sêmen e embriões para uso em reprodução e cruzamentos industriais com raças especializadas para produção de leite e carne macia, comercializados com denominação de origem protegida. A resistência natural a ecto e endoparasitas e adaptação às nossas gramíneas e leguminosas são as grandes armas desses bovinos, que foram selecionados naturalmente por séculos para enfrentarem as adversidades locais. A tudo isso vem somar a grande amplitude térmica em que podem ser criados e grande longevidade, vivendo por mais de 20 anos. Entretanto o grande mérito dessa raça é converter alimentos de baixa qualidade em alimentos nobres e possibilitar a convivência do homem em regiões semiáridas.

Dentre as vantagens de potencial de uso dos bovinos da raça Curraleiro Pé-Duro está o fato desses animais serem criados em sistema extensivo de produção, uma vez que as regiões onde eles são encontrados possuem grandes extensões de pastagem nativa. Segundo Fioravanti et al. (2011), os sistemas extensivos de produção de carne orgânica tendem a equilibrar a oferta e a utilização dos recursos naturais, alcançando um equilíbrio entre produção e conservação, e esta atividade pode ser uma alternativa para o fortalecimento do turismo rural. Outra vantagem no uso

da raça Curraleiro Pé-Duro é o interesse dos produtores em conservar a raça, pois quando questionados sobre os motivos que os levam a manter esses animais, descrevem que o fazem pela tradição, pelo sabor e qualidade da carne, ficando evidente o caráter cultural e familiar da atividade (Fioravanti et al., 2011). Além disso, criadores afirmam que a carne do gado Curraleiro Pé-Duro é de excelente sabor, o que somado à rusticidade da raça (em que o uso de produtos químicos e medicamentos, é reduzido) pode ser criado um mercado diferenciado para essa carne. Vale notar que há uma tendência mundial de valorização cada vez maior dos produtos naturais (Carvalho, 2002).



Figura 1. Curraleiro Pé-Duro.

1.2 Objetivo Geral e Específicos

No contexto apresentado, o objetivo geral com este trabalho é apresentar a metodologia de Análise de Sobrevivência como alternativa para análises em estudos genéticos relacionados a bovinos da raça Curraleiro Pé- Duro, no contexto em que a variável resposta está sendo influenciada por fatores não observáveis (latentes).

Os objetivos específicos estão listados abaixo:

- Aplicar o modelo Cox- Log Normal, para avaliar o tempo de permanência no rebanho Curraleiro Pé-Duro, relacionado a saída do rebanho por morte ou venda;

- Analisar os efeitos de fatores não observados, em estudo sobre o tempo de permanência no rebanho, de bovinos da raça Curraleiro Pé-Duro utilizando-se do modelo de fragilidade Gaussiana inversa- Log Normal univariada e multivariada.

1.3 Organização do Trabalho

Este trabalho foi dividido em capítulos, onde no capítulo 1 foi feita uma introdução, com a devida motivação e a apresentação dos objetivos, geral e específicos.

No Capítulo 2, é apresentado uma revisão de literatura contendo os conceitos básicos em análise de sobrevivência com suas respectivas funções básicas, o estimador de Kaplan-Meier, a função de verossimilhança em análise de sobrevivência e o modelo de fragilidade.

No Capítulo 3, é apresentado o primeiro artigo intitulado por “Modelo de Cox- Log Normal para avaliar a longevidade em bovinos da raça Curraleiro Pé-Duro”;

No Capítulo 4, é apresentado o segundo artigo intitulado por “Avaliação da longevidade de bovinos da raça Curraleiro Pé-Duro usando o modelo de fragilidade Gaussiana inversa”;

No Capítulo 5, é apresentado o terceiro artigo intitulado por “Modelagem de fragilidade Gaussiana inversa compartilhada aplicada a bovinos da raça Curraleiro Pé- duro”;

No Capítulo 6, apresenta-se as conclusões gerais e propostas futuras deste trabalho de tese.

2 REVISÃO DE LITERATURA

2.1 Conceitos básicos em análise de sobrevivência

A Análise de Sobrevivência (AS), por sua vez, é considerada metodologia estatística mais adequada para lidar com dados de tempo até a ocorrência de um evento de interesse (tempo de falha), na presença de censura (Colosimo e Giolo, 2006) que é sua principal característica. Com a presença de censuras, torna-se impossível a aplicação de técnicas estatísticas convencionais. Na área animal, temos por exemplo, o experimento conduzido por Elliott et al. (2000), onde 50 gatos

domésticos foram observados. Os tempos de falha foram considerados como o tempo (em dias) a partir do diagnóstico de insuficiência renal até a morte. Por outro lado, se os animais não morreram até o final do estudo ou se morreram por outras causas, os tempos foram considerados tempos censurados. Uma outra situação que podemos destacar na produção animal, é a medida em dias, meses, ou anos entre o nascimento, ou o primeiro parto até a morte e/ou descarte da vaca. Por outro lado, se analisarmos a longevidade usando a metodologia da análise de sobrevivência, já podemos observar a presença de algumas particularidades, como a presença de censura (Ducrocq et al., 1988). Além dessa característica, há também a influência das variáveis explicativas, que devem ser levadas em consideração.

Uma característica especial associada aos dados de sobrevivência é a possibilidade de que, para alguns indivíduos, o tempo completo até a ocorrência do evento de interesse não seja observado, devido a diversas causas. Por exemplo, um indivíduo pode permanecer em acompanhamento até o final do tempo de estudo sem ter experimentado o evento de interesse. A não consideração destes indivíduos com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes (Bolfarine, 1991). Portanto, a importância da introdução de uma variável na análise que indique se o tempo de sobrevivência foi observado (Louzada-Neto e Pereira, 2000). Essa variável é definida na literatura como variável indicadora de censura ou simplesmente “censura”.

A censura pode ser de três tipos (Colosimo e Giolo, 2006):

1. Censura à direita: quando o tempo de ocorrência do desfecho está à direita do tempo registrado. Um exemplo para esse tipo de censura é em bovinos leiteiros no estudo da longevidade, sendo a duração da vida produtiva (DVP) a variável resposta mais comumente utilizada (Forabosco et al. (2009); Imbayarwo-Chikosi et al. (2015)). A DVP mede o intervalo de tempo, em dias ou em meses, entre a data ao primeiro parto e a data de descarte ou morte da vaca. A data ao primeiro parto é conhecida, mas a data de descarte e/ou morte nem sempre está presente, visto que no

momento do estudo a vaca ainda pode estar viva e/ou ter sido vendida a outro rebanho que não faz controle leiteiro, não conhecendo, assim, a data de ocorrência do descarte do animal (Ducrocq et al., 1988). Desta forma, a observação desta vaca é dita como censurada à direita. Os mecanismos de censura à direita pode ser do tipo I, que é aquela em que o estudo será terminado após um período preestabelecido de tempo; A do tipo II, onde o estudo será terminado após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos; e a do tipo aleatória, que ocorre quando o indivíduo é retirado do estudo sem ter ocorrido a falha, ou por exemplo, o indivíduo morrer por uma razão diferente da estudada (Figura 1).

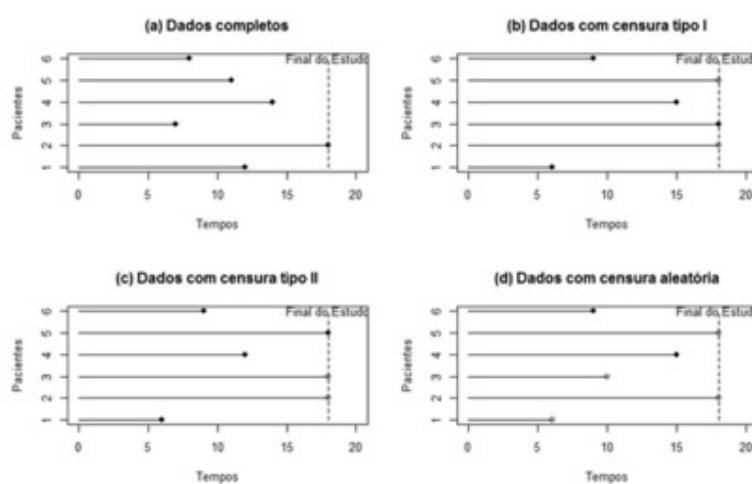


Figura 1. (a) Dados completos, (b) Dados com censura tipo I, (c) Dados com censura tipo II, (d) Dados com censura aleatória.

2. **Censura à esquerda:** não se conhece o momento de ocorrência do desfecho, mas sabemos que ele ocorreu antes do tempo observado. Por exemplo, um estudo sobre a determinação do tempo que crianças de um certo local levaram para aprender a escrever, pode ser que antes do início do estudo, algumas dessas crianças já tivessem aprendido, e portanto, não havia como identificar quando isso ocorreu.

3. **Censura intervalar:** ocorre quando não se sabe o tempo exato da ocorrência da falha, no entanto, sabe-se que ela ocorreu dentro de um intervalo conhecido. Exemplo, tempo entre exames feitos semestralmente em animais para prevenir uma certa doença e a ocorrência da doença sobre eles,

assim não se sabe o tempo exato do evento apenas o intervalo entre o exame e a ocorrência da doença.

Os dados de sobrevivência para o indivíduo i , $i = 1, \dots, n$, são representados, em geral, por (t_i, δ_i) , sendo t_i o tempo de falha ou de censura e δ_i a variável indicadora que acusa se há presença de falha ou de censura, ou seja,

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ for tempo até a falha} \\ 0, & \text{se } t_i \text{ for tempo até a censura} \end{cases}$$

Em estudos recentes há situações em que a variável resposta, tempo de falha, pode estar sendo influenciado por fatores não observáveis, denominado fatores latentes. Modelos de sobrevivência com variáveis latentes ou modelos com fragilidade, são caracterizados pela inclusão de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou que não foram observadas; como fatores ambientais, genéticos ou informações que por algum motivo não foram considerado no planejamento. Uma das formas encontradas para incorporar esse efeito aleatório, denominado de variável de fragilidade, é introduzi-lo na modelagem da função de risco, com o objetivo de controlar a heterogeneidade não observável das unidades em estudo.

2.1.1 Funções Básicas de Sobrevivência

A descrição do Tempo de Sobrevivência representado pela variável aleatória não-negativa T , será representada pela função densidade de probabilidade $f(t)$, pela função de sobrevivência $S(t)$ e pela função de falha $h(t)$. A função de densidade, $f(t)$, é definida por

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t},$$

em que $F(t) = P[T \leq t] = \int_0^t f(u) du$ é a função de distribuição acumulada (f.d.a) de T . No contexto da análise de sobrevivência, a função $f(t)$ pode ser interpretada como a probabilidade de

um indivíduo sofrer o acontecimento de interesse no intervalo de tempo Δt e possui duas propriedades:

$$f(t) \geq 0 \quad \text{e} \quad \int_0^{\infty} f(t)dt = 1$$

A função de sobrevivência, $S(t)$, é definida como sendo a probabilidade de ocorrência do acontecimento de interesse após o instante t . A função de sobrevivência é definida da seguinte forma:

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u)du = 1 - F(t)$$

em que $\lim_{t \rightarrow 0} S(t) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

A função risco, $h(t)$, representa o risco instantâneo no instante t condicional à sobrevivência até o tempo t e é definida por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

Graficamente a função de risco pode apresentar comportamento constante, crescente, decrescente e até mesmo formas não monótonas como a “curva da banheira” que, geralmente, representa a função de risco da vida de um ser humano. A função de risco acumulada é definida por:

$$H(t) = \int_0^t h(u)du = -\log S(t)$$

sendo importante em análises gráficas para averiguar a adequação de modelos estatísticos.

Uma forma empírica de determinar o comportamento da função risco se dá por meio da construção do gráfico do tempo total em teste (curva TTT), proposto por Aarset (1987). A curva TTT é obtida construindo um gráfico de

$$G\left(\frac{r}{n}\right) = \frac{\sum_{i=n}^r T_{i:n} + (n-r)T_{r:n}}{\sum_{i=n}^r T_{i:n}} \quad \text{por} \quad \frac{r}{n}$$

em que n é o tamanho da amostra, $r = 1, \dots, n$ e $T_{i:n}, i = 1, \dots, n$ são estatísticas de ordem da amostra (Mudholkar et al., 1996).

Este gráfico apresenta uma reta diagonal se o risco for constante (reta A), uma curva convexa se a função risco é decrescente (curva B) e côncava se o risco é crescente (curva C), uma curvatura primeiramente convexa e depois côncava (curva D) se o risco é em forma de “U”, e no caso reverso (curva E) é unimodal. A Figura 2 ilustra as diversas formas que a curva TTT pode apresentar.

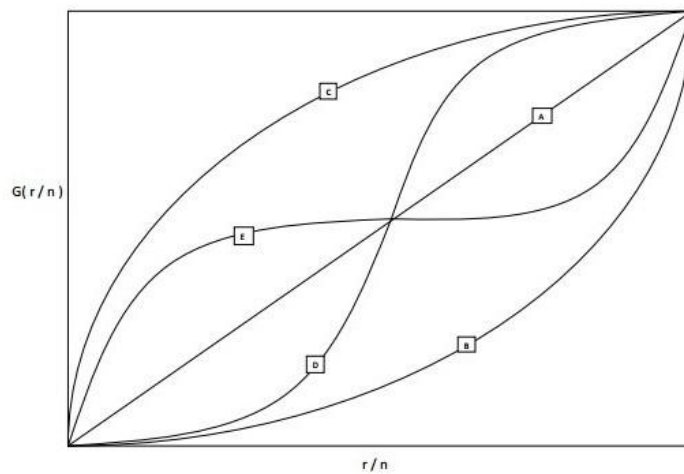


Figura 2. Gráficos ilustrativos de algumas curvas TTT

Algumas relações importantes entre as funções definidas anteriormente, e que são amplamente utilizadas na prática, são:

$$S(t) = 1 - F(t)$$

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

$$H(t) = -\log S(t)$$

$$S(t) = \exp\{-H(t)\}$$

2.1.2 O estimador de Kaplan-Meier e o teste Log-Rank

Kaplan e Meier (1958) propuseram um estimador não-paramétrico para a função de sobrevivência, quando se tem a presença de uma amostra censurada. Esse estimador é designado estimador de Kaplan-Meier ou estimador produto-limite. O estimador considera na sua construção um número de intervalos de tempo igual ao número de falhas distintas, sendo os limites dos intervalos definidos pelos tempos de falha da amostra. Dessa forma, o estimador de Kaplan-Meier pode ser definido como:

$$S(t) = \prod_{j:t_{(j)} \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

em que $t_{(1)}, t_{(2)}, \dots, t_{(k)}$ representam os k tempos de falha distintos e ordenados, d_j é o número de falhas em $t_{(j)}$, $j = 1, \dots, k$, e n_j é o número de indivíduos sob risco em $t_{(j)}$, ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a $t_{(j)}$.

A curva para a função de sobrevivência estimada, considerando o estimador de Kaplan-Meier, possui a forma de uma escada, em que os degraus são os intervalos de tempo entre duas falhas distintas observadas. Quando se pretende comparar a função de sobrevivência de dois ou mais grupos, a estimativa de Kaplan-Meier estratificada por grupo é útil para analisar o comportamento das diferentes curvas de sobrevivência. Porém, para verificar se existe diferença estatística significativa entre várias curvas existem vários testes não paramétricos, entre eles, o teste de Log-rank. Proposto por Mantel e Haenszel (1959) o teste Log-rank é o mais usado em análise de sobrevivência. É particularmente apropriado quando a razão das funções de risco dos grupos a serem comparados é aproximadamente constante. Isto é, as populações tem a propriedade de riscos proporcionais. A estatística deste teste é a diferença entre o número de falhas observado em cada grupo e uma quantidade que pode ser pensada como o correspondente número esperado de falhas sob a hipótese nula. (Colosimo e Giolo, 2006).

Considere, inicialmente, o teste de igualdade de duas funções de sobrevivência $S_1(t)$ e $S_2(t)$. Sejam $t_1 < t_2 < \dots < t_k$ os tempos de falha distintos da amostra formada pela combinação das duas amostras individuais. Suponha que no tempo t_j acontecem d_j falhas e n_j indivíduos estão sob risco em um tempo imediatamente inferior a t_j na amostra combinada e, respectivamente, d_{ij} e n_{ij} na amostra $i = 1, 2$ e $j = 1, \dots, k$. Em cada tempo de falha t_j , os dados podem ser dispostos em forma de uma tabela de contingência 2×2 com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i . A estatística de teste é dada por:

$$T = \frac{\left[\sum_{j=1}^k (d_{2j} - w_{2j}) \right]^2}{\sum_{j=1}^k (V_j)_2}$$

que, sob a hipótese nula $H_0 : S_1(t) = S_2(t)$ para todo t no período de acompanhamento, tem distribuição qui-quadrado com 1 grau de liberdade para grandes amostras. Ainda, d_{2j} é o número observado de falhas no grupo 2 no tempo t_j , $w_{2j} = n_{2j}d_jn_j^{-1}$ é a média de d_{2j} , $(V_j)_2 = n_{2j}(n_j - n_{2j})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}$ é a variância de d_{2j} .

2.2 Alguns Modelos Paramétricos de Sobrevivência

2.2.1 Distribuição Gompertz

A distribuição Gompertz foi desenvolvida a partir da lei de mortalidade formulada pelo matemático britânico Benjamin Gompertz em seus estudos em demografia (Gompertz, 1825). Áreas científicas como biologia e gerontologia consideram a distribuição Gompertz para análise de sobrevivência.

Para a variável aleatória T a função densidade de probabilidade $f(t)$, com parâmetros λ e γ é dado por:

$$f(t) = \lambda e^{\gamma t} \exp \left\{ \left(-\frac{\lambda}{\gamma} \right) (\exp(\gamma t) - 1) \right\}$$

As funções de Sobrevivência $S(t)$, de risco $h(t)$ e risco acumulado $H(t)$ são dadas respectivamente por:

$$S(t) = \exp \left\{ \left(-\frac{\lambda}{\gamma} \right) (\exp(\gamma t) - 1) \right\}$$

$$h(t) = \lambda \exp(\gamma t)$$

$$H(t) = \frac{\lambda}{\gamma} (\exp(\gamma t) - 1)$$

2.2.2 Distribuição Log-Logística

A distribuição log-logística, em muitas situações práticas, tem se apresentado como uma alternativa à de Weibull e à log-normal. Para uma variável aleatória T com esta distribuição, a função de densidade é expressa por:

$$f(t) = \frac{\exp(\alpha)\gamma t^{\gamma-1}}{[1 + \exp(\alpha)t^\gamma]^2}$$

sendo $\alpha \in \mathfrak{R}$ o parâmetro de forma e $\gamma > 0$ o de escala. As funções de sobrevivência $S(t)$ e de risco $h(t)$ são expressas, respectivamente, por:

$$S(t) = \frac{1}{1 + \exp(\alpha)t^\gamma} \quad \text{e} \quad h(t) = \frac{\exp(\alpha)\gamma t^{\gamma-1}}{1 + \exp(\alpha)t^\gamma}$$

2.2.3 Distribuição Log Normal

É muito utilizada para caracterizar tempos de vida de produtos e indivíduos. A função densidade de uma variável aleatória T com distribuição Log Normal é dada por:

$$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right], \quad t > 0$$

em que μ é a média do logaritmo do tempo de falha, assim como σ é o desvio padrão.

As funções de sobrevivência $S(t)$ e de risco $h(t)$ de uma variável Log Normal não apresentam uma forma analítica explícita e são, desse modo, representadas, respectivamente, por:

$$S(t) = 1 - \Phi \left(\frac{\log(t) - \mu}{\sigma} \right) \quad \text{e} \quad h(t) = \frac{f(t)}{S(t)}$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão.

A Figura 3 mostra a forma de algumas funções de densidade, sobrevivência e de risco da distribuição Log Normal para alguns valores de μ e σ .

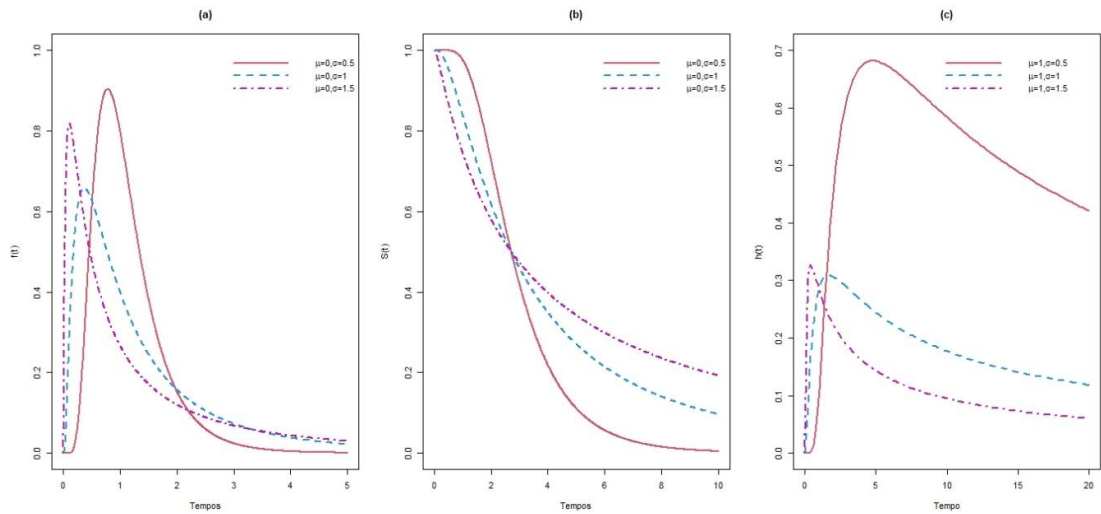


Figura 3. Formas das funções de densidade, sobrevivência e de risco da distribuição Log Normal para alguns valores de μ e σ .

A média e a variância de uma distribuição Log Normal são dadas, respectivamente, por:

$$E(T) = \exp\{\mu + \sigma^2/2\} \quad \text{e} \quad \text{Var}(T) = \exp\{2\mu + \sigma^2\}(\exp\{\sigma^2\} - 1).$$

2.2.4 Distribuição Gaussiana Inversa

A função densidade de probabilidade de uma variável aleatória T , não negativa, com distribuição Inversa-Gaussiana é dada por:

$$f(t, a, b) = \frac{1}{\sqrt{2b\pi t^3}} \exp\left[-\frac{1}{2bt}(1-at)^2\right]$$

em que $a > 0$, $b > 0$ e $t > 0$. A correspondente função de sobrevivência $S(t)$, é dada por

$$S(t, a, b) = 1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + \exp^{2a/b} \Phi\left(\frac{-1-at}{\sqrt{bt}}\right) \right]$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão. A função risco é dada por

$$h(t, a, b) = \frac{f(t, a, b)}{S(t, a, b)} = \frac{\frac{1}{\sqrt{2b\pi t^3}} \exp\left[-\frac{1}{2bt}(1-at)^2\right]}{1 - \left[\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + \exp^{2a/b} \Phi\left(\frac{-1-at}{\sqrt{bt}}\right) \right]}$$

A Figura 4 mostra a forma de algumas funções de densidade, sobrevivência e de risco da distribuição Gaussiana inversa para alguns valores de a e b .

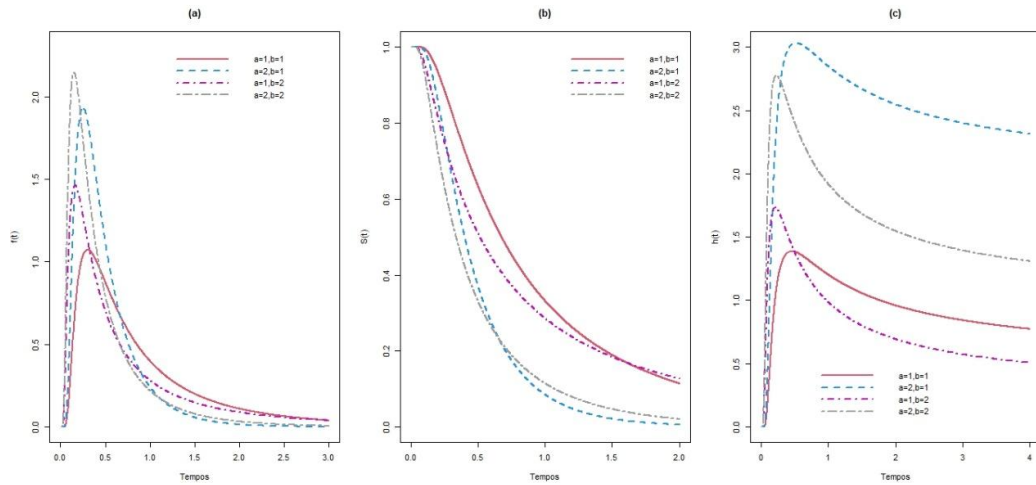


Figura 4. Formas das funções de densidade, sobrevivência e de risco da distribuição Gaussiana inversa para alguns valores de a e b .

2.2.5 Distribuição Gama

A função de densidade da distribuição gama, que é caracterizada por dois parâmetros, k e α , em que $k > 0$ é chamado parâmetro de forma e $\alpha > 0$ de escala, é expressa por:

$$f(t) = \frac{1}{\Gamma(k)\alpha^k} t^{k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, \quad t > 0$$

com $\Gamma(k)$ a função gama definida por $\Gamma(k) = \int_0^\infty x^{k-1} \exp(-x) dx$. Para $k > 1$, esta função de densidade apresenta um único pico em $t = (k - 1)/\alpha$.

A função de sobrevivência desta distribuição, é dada por:

$$S(t) = \int_t^\infty \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp\left\{-\left(\frac{u}{\alpha}\right)\right\} du$$

A função taxa de falha, obtida da relação $h(t) = f(t)/S(t)$, apresenta um padrão crescente ou decrescente convergindo, no entanto, para um valor constante quando t cresce de 0 a infinito. A média e a variância da distribuição gama são dadas, respectivamente, por $k\alpha$ e $k\alpha^2$.

2.2.6 Distribuição Gama Generalizada

A distribuição gama generalizada, foi introduzida por Stacy et al. (1962) e despertou o interesse de diversos pesquisadores pelo fato de representar uma família paramétrica que possui como casos particulares a distribuição exponencial quando $\tau = k = 1$ na equação (1), se $\tau = 1$ obtém-se a distribuição gama. Para $k = 1$ tem-se a distribuição Weibull, para $\alpha = 1/\lambda$, $\tau = 2$ e $k = 1$ obtém-se a distribuição Rayleigh; outros sub-modelos podem ser vistos por exemplo em Lawless (2011).

A função densidade de probabilidade da distribuição Gama Generalizada proposta por Stacy et al. (1962) é dada por:

$$f(t) = \frac{\tau}{\alpha \Gamma(k)} \cdot \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right], \quad t > 0 \quad (1)$$

em que $\alpha > 0$ é o parâmetro de escala, $\tau > 0$ e $k > 0$ são os parâmetros de forma $\Gamma(k)$ é a função gama, definida por:

$$\Gamma(k) = \int_0^\infty t^{k-1} \exp(-t) dt$$

Se T é uma variável aleatória positiva com distribuição Gama Generalizada com parâmetros α, τ e k , então denota-se que $T \sim GG(\alpha, \tau, k)$.

A média e a variância da distribuição Gama Generalizada são dadas por:

$$E(T) = \frac{\alpha \Gamma\left(\frac{\tau k + 1}{\tau}\right)}{\Gamma(k)} \quad \text{e} \quad V(T) = \frac{\alpha^2}{\Gamma(k)} \left\{ \Gamma\left(\frac{\tau k + 2}{\tau}\right) - \frac{[\Gamma\left(\frac{\tau k + 1}{\tau}\right)]^2}{\Gamma(k)} \right\}$$

A função de distribuição acumulada $F(t)$, função de sobrevivência $S(t)$ e função taxa de falha $h(t)$ são expressas, respectivamente, por:

$$F(t) = P[T \leq t] = \frac{\gamma(k, (t/\alpha)^\tau)}{\Gamma(k)} = \frac{1}{\Gamma(k)} \int_0^{(t/\alpha)^\tau} w^{k-1} \exp(-w) dw = \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]$$

$$S(t) = 1 - F(t) = 1 - \gamma_1\left[k, \left(\frac{t}{\alpha}\right)^\tau\right]$$

em que $\gamma(k; x) = \int_0^x w^{k-1} e^{-w} dw$ é a função gama incompleta e $\gamma_1(k; x)$ é a razão da função

gama incompleta, definida por $\gamma_1(k; x) = \frac{\gamma(k; x)}{\Gamma(k)}$ e

$$h(t) = \frac{f(t)}{S(t)} = \frac{t^{\tau k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\tau\right]}{\int_0^\infty x^{\tau k-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\tau\right] dx}$$

2.3 A função de verossimilhança em análise de sobrevivência

Em análise de sobrevivência, o método utilizado para estimação dos parâmetros de modelos paramétricos é o método da máxima verossimilhança, pois permite incorporar a informação de dados censurados (Lawless, 2011). Suponha T uma variável aleatória com distribuição de probabilidades $f(t; \theta)$, em que θ é um único parâmetro desconhecido. Sejam t_1, t_2, \dots, t_n os valores observados de uma população de interesse. Se não houver censura temos que a função de verossimilhança para θ é expressa por

$$L(\theta, t) = \prod_{i=1}^n f(t_i, \theta)$$

Quando houver censura a função de verossimilhança para θ é expressa por

$$L(\theta) = \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i}$$

em que δ_i é a variável indicadora de falha ou censura, θ é o vetor de parâmetros desconhecidos, $f(t_i)$ e $S(t_i)$ são as funções densidade de probabilidade e de sobrevivência para cada tempo de sobrevivência t_i , respectivamente. Observa-se que a contribuição de cada observação não censurada é dada pela sua função de densidade e que cada observação censurada contribui com a função de sobrevivência.

Assim, os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$, ou, equivalentemente, $\log(L(\theta))$. Para encontrar os valores de θ que maximizam o logaritmo de $L(\theta)$, ou seja, o estimador de máxima verossimilhança, temos que resolver o seguinte sistemas:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0$$

em que $U(\theta)$ é a função escore.

2.4 Modelos de Riscos Proporcionais de Cox

O modelo de regressão de Cox foi proposto por Cox (1972), onde permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse, ajustando por covariáveis (Colosimo e Giolo, 2006). Esse modelo passou a ser o mais utilizado na análise de dados de sobrevivência por sua versatilidade. Como em análise de sobrevivência o interesse também pode estar no risco de falha, o modelo proposto por Cox modela diretamente a função de risco. Ainda, permite incorporar facilmente covariáveis dependentes do tempo, que ocorrem com frequência em várias áreas de aplicação. O princípio básico deste modelo para estimar o efeito das covariáveis é a proporcionalidade dos riscos ao longo de todo o tempo de observação.

O modelo de riscos proporcional, é dado por:

$$h(t_i|x_i) = h_0(t_i)\exp\{g(\beta'x_i)\} \quad (2)$$

em que $h_0(t)$ é a função de risco de base, isto é o risco de um animal com covariáveis $x = 0$; $\beta = (\beta_1, \dots, \beta_k)$ é o vetor de dimensão k de coeficientes de regressão desconhecidos e $x = (x_1, \dots, x_k)$ é o vetor de dimensão k de covariáveis para a i -ésima observação.

O modelo (2) é composto pelo produto de dois componentes: $\exp\{\beta'x\}$ que mede o efeito das covariáveis e $h_0(t)$ que pode ser paramétrico ou não. No caso semiparamétrico o componente $h_0(t)$ é considerado ser uma função não negativa do tempo quando paramétrico, pode assumir distribuições como Gompertz, Log Normal, Gama entre outras.

O componente paramétrico é frequentemente usado na seguinte forma multiplicativa:

$$g(\beta'x) = \exp\{\beta'x\} = \exp\{\beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k\}$$

em que β é o vetor de parâmetros associado às covariáveis x .

Este modelo é também denominado modelo de risco proporcionais, pois a razão das taxas de falha de dois indivíduos diferentes é constante no tempo. Isto é, a razão das funções de taxa de falha para os indivíduos i e j dada por:

$$\frac{h(t|x_i)}{h(t|x_j)} = \frac{h_0(t)g(\beta'x_i)}{h_0(t)g(\beta'x_j)} = \exp\{\beta'x_i - \beta'x_j\}$$

não dependente do tempo. Por exemplo, se um indivíduo no início do estudo tem um risco de morte igual a duas vezes o risco de um segundo indivíduo, então, esta razão de riscos será a mesma para todo o período de acompanhamento (Colosimo e Giolo, 2006). A suposição básica para o uso do modelo de regressão de Cox é, portanto, que as taxas de falha sejam proporcionais ou, de forma equivalente para este modelo, que as taxas acumuladas sejam também proporcionais.

2.4.1 Estimação do Modelo de Cox

Para fazer a inferência quanto aos parâmetros β do modelo, é preciso fazer o uso de uma técnica de estimação, ou seja, o método da máxima verossimilhança parcial, que estima o vetor dos parâmetros β . Devido a presença do componente não aleatório $h_0(t)$ não é viável usar o método da máxima verossimilhança. Ou seja, sabe-se que,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i)^{1-\delta_i} \\ &= \prod_{i=1}^n [h(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i) \end{aligned} \quad (3)$$

No modelo de Cox, a função de sobrevivência é definida por:

$$S(t_i|x_i) = \exp \left\{ - \int_0^{t_i} h_0(u) \exp\{x_i'\beta\} du \right\} = [S_0(t_i)]^{\exp\{x_i'\beta\}}$$

Aplicando esse resultado em (3), temos que:

$$L(\beta) = \prod_{i=1}^n [h_0(t_i) \exp\{x_i' \beta\}]^{\delta_i} [S_0(t_i)]^{\exp\{x_i' \beta\}}$$

que é função do componente não-paramétrico $h_0(t)$.

Cox (1975) propôs então uma solução razoável que consiste em condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras para eliminar esta função de perturbação da verossimilhança conhecida por função de verossimilhança parcial.

Considere que uma amostra de n indivíduos existam $k < n$ falhas distintas nos tempos $t_1 < t_2 < \dots < t_k$. A função de Verossimilhança é dado por:

$$L(\beta) = \prod_{i=1}^n \left(\frac{\exp\{x_i' \beta\}}{\sum_{i \in R(t_i)} \exp\{x_i' \beta\}} \right)^{\delta_i} \quad (4)$$

onde δ_i é o indicador de censura e $R(t_i)$ é o conjunto de índices das observações sob risco no tempo t_i . Observe que condicional à história de falhas e censuras até o tempo t_i , o componente não-paramétrico $h_0(t)$ desaparece. A função de verossimilhança a ser utilizada para fazer inferências no modelo de Cox é, então, formada pelo produto de todos os termos associados aos tempos distintos de falha. Os valores de β que maximizam a função de verossimilhança parcial, $L(\beta)$, são obtidos resolvendo o sistema de equações definido por $U(\beta) = 0$, em que $U(\beta)$ é o vetor de derivadas de primeira ordem da função $\log(L(\beta))$, isto é

$$U(\beta) = \sum_{i=1}^n h_0 \left[x_i - \frac{\sum_{j \in R(t_i)} x_j \exp\{x_j' \hat{\beta}\}}{\sum_{j \in R(t_i)} \exp\{x_j' \hat{\beta}\}} \right] \quad (5)$$

A função de verossimilhança parcial (4) assume que os tempos de sobrevivência são contínuos e, conseqüentemente, não pressupõe a possibilidade de empates nos valores observados. Na prática, podem ocorrer empates nos tempos de falhas ou censuras devido a escala de medida. Quando isso ocorre, usa-se a convenção de que a censura ocorreu após a falha, o que define as observações a serem incluídas no conjunto de risco em cada tempo de falha. A função de verossimilhança parcial (4) deve ser modificada para incorporar as observações empatadas quando

estas estão presentes. Diante disso, usa-se a aproximação proposta por Breslow (1972) e Peto (1972), quando o número de observações empatadas for pequena para qualquer tempo.

Seja s_i o vetor formado pela soma das correspondentes p covariáveis para indivíduos que falham no mesmo tempo t_i ($i = 1, \dots, k$) e d_i o número de falhas nesse mesmo tempo. Assim,

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{s_i' \beta\}}{[\sum_{j \in S(t_i)} \exp\{x_j' \beta\}]^{d_i}} \quad (6)$$

A expressão (5) é adequada quando o número de observações empatadas em qualquer tempo não é grande. Naturalmente, a expressão (4) se reduz a (6) quando não houver empates. Para testar as hipóteses sobre os parâmetros do modelo é possível utilizar as estatísticas de Wald, Teste da razão de verossimilhança e Escore.

2.4.2 Intervalo de Confiança e Teste de Hipóteses

A construção de intervalos de confiança diz respeito à distribuição assintótica do estimador de máxima verossimilhança $\hat{\theta}$. Para grandes amostras, sob certas condições de regularidade, a distribuição do vetor $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)'$ é normal multivariada de média θ e matriz de variância-covariância $Var(\hat{\theta})$, isto é,

$$\hat{\theta} \approx N_k(\theta, Var(\hat{\theta}))$$

Na construção de intervalos de confiança é necessário uma estimativa para o erro padrão de $\hat{\theta}$, isto é, para $[Var(\hat{\theta})]^{1/2}$. No caso especial em que θ é um escalar, um intervalo aproximado de $(1-\alpha)100\%$ de confiança para θ é dado por

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta})}$$

Para um modelo com um vetor $\theta = (\theta_1, \dots, \theta_p)'$ de parâmetros, muitas vezes há o interesse em testar hipóteses relacionadas a este vetor ou a um subconjunto dele. Três testes são em geral utilizados para esta finalidade: o de Wald, o da Razão de Verossimilhança (TRV) e o Escore. Uma breve descrição dos testes é apresentado a seguir (Colosimo e Giolo, 2006):

i) Teste de Wald

O teste de Wald é baseado na distribuição assintótica de $\hat{\theta}$ e é uma generalização do teste t de Student (Wald, 1943). É usado para testar hipóteses relativas a um único parâmetro θ_j . Considerando-se a hipótese nula

$$H_0 : \theta = \theta_0$$

A estatística para esse teste é dada por:

$$W = (\hat{\theta} - \theta_0)' F(\theta_0) (\hat{\theta} - \theta_0) \quad (7)$$

que, sob H_0 , tem aproximadamente distribuição qui-quadrado com p graus de liberdade (χ^2_p). A um nível $100\alpha\%$ de significância, valores de W superiores ao valor tabelado da distribuição $\chi^2_{1,1-\alpha}$ indicam a rejeição de H_0 . No caso em que θ é um escalar, a expressão (7) se reduz a:

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\widehat{Var}(\hat{\theta})}$$

ii) Teste da Razão de Verossimilhança

É baseado na função de verossimilhança e envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição e sob H_0 , ou seja, a comparação de $\log L(\hat{\theta})$ e $\log L(\hat{\theta}_0)$. A estatística para este teste é dada por:

$$TRV = -2 \log \left[\frac{\log L(\hat{\theta}_0)}{\log L(\hat{\theta})} \right] = 2[\log L(\hat{\theta})] - \log L(\hat{\theta}_0)$$

que, sob $H_0 : \theta = \theta_0$, segue aproximadamente uma distribuição qui-quadrada com p graus de liberdade. Para amostras grandes, H_0 é rejeitada, a um nível de $100\alpha\%$ de significância, se $TRV > \chi^2_{p,1-\alpha}$.

ii) Teste Escore

Este teste é obtido a partir da função escore sendo, a sua estatística de teste, dada por:

$$S = U'(\theta_0)[F(\theta_0)]^{-1}U(\theta_0)$$

em que $U(\theta_0)$ é a função escore $U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta}$ avaliada em θ_0 , e $F(\theta_0)$ a matriz de variância-covariância de observada de $\hat{\theta}$ também avaliada em θ_0 . Para amostras grandes, H_0 é rejeitada, a um nível 100% de significância, se $S > X_{p, 2}^2_{1-\alpha}$.

2.4.3 Adequação e Seleção do modelo

O modelo de regressão de Cox é bastante flexível devido a presença do componente não-paramétrico. Mesmo assim, ele não se ajusta a qualquer situação clínica e, como qualquer outro modelo estatístico, requer o uso de técnicas para avaliar a sua adequação. A violação da suposição básica, que é a de riscos proporcionais, pode acarretar em sérios vícios na estimação dos coeficientes do modelo (Struthers e Kalbfleisch, 1986). Diversos métodos para avaliar a adequação desse modelo encontram-se disponíveis na literatura e baseiam-se, essencialmente, nos resíduos de Cox-Snell, de Schoenfeld, Martingal e Deviance.

i) Resíduos de Cox-Snell

Os resíduos de Cox e Snell (1968) auxiliam a examinar o ajuste global do modelo. Estes resíduos são quantidades determinadas por:

$$\hat{e}_i = \hat{H}(t_i | x_i)$$

em que $\hat{H}(\cdot)$ é a função de risco acumulado obtida no modelo ajustado.

Os resíduos \hat{e}_i vem de uma população homogênea e devem seguir uma distribuição exponencial padrão se o modelo for adequado (Lawless, 2011). Assim, o gráfico \hat{e}_i versus $\hat{H}(\hat{e}_i)$ deve ser aproximadamente uma reta com inclinação 1, quando o modelo exponencial for adequado, uma vez que $\hat{H}(\hat{e}_i) = -\log(\hat{S}(\hat{e}_i))$, onde $\hat{S}(\hat{e}_i)$ é a função de sobrevivência dos (\hat{e}_i) 's obtida pelo estimados de Kaplan-Meier. O gráfico das curvas de sobrevivência desses resíduos, obtidas por Kaplan-Meier e pelo modelo exponencial padrão, também auxiliam na verificação da qualidade do modelo ajustado, sendo que, quanto mais próximas elas se apresentarem, melhor é considerado o ajuste do modelo aos dados.

ii) Resíduos de Schoenfeld

É um método com coeficiente dependente do tempo e tem uma proposta adicional de análise da suposição de riscos proporcionais. Para o indivíduo i com covariáveis $x_i = (x_{1i}; \dots; x_{pi})'$ e que foi observado falhar, o vetor de resíduos Schoenfeld $r_i = (r_{1i}; \dots; r_{pi})'$ tem cada componente r_{iq} definido por:

$$r_{iq} = x_{iq} - \frac{\sum_{j \in S(t_i)} x_{jq} \exp\{x_j' \beta\}}{\sum_{j \in S(t_i)} \exp\{x_j' \beta\}}$$

Os resíduos padronizados de Schoenfeld são dados por:

$$s_i^* = [L(\hat{\beta})]^{-1} x r_i$$

com $L(\hat{\beta})$ a matriz de informação observada.

Com a restrição de que $\beta(t) = \beta$, como uma forma alternativa de representar o modelo de Cox. A restrição $\beta(t) = \beta$, implica na proporcionalidade dos riscos. Quando interpretamos o gráfico dos resíduos de Schoenfeld versus o tempo de vida (ou as ordens dos tempos de vida), se os dados se dispuserem numa nuvem de pontos aleatória, centrada em zero, então estamos perante um modelo adequado para os dados. Os resíduos de Schoenfeld são, em geral, usados para verificar a suposição de riscos proporcionais.

iii) Resíduos de Martingal

O resíduo Martingal é a diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, o tempo de seguimento e o percurso observado de quaisquer covariáveis tempo dependentes. Assim,

$$\widehat{m}_i = \delta_i - \widehat{e}_i$$

em que δ_i é a variável indicadora de falha e \widehat{e}_i os resíduos de Cox-Snell.

Esses resíduos podem ser usados para verificar a adequação do modelo mas, na prática, devido à sua distribuição assimétrica o que torna sua interpretação difícil, eles são usados para

verificação da forma funcional das covariáveis, isto é se estas deveriam ser usadas no modelo como $\log(x_i)$, x_i^2 e assim por diante, em vez de x_i .

iv) Resíduos Deviance

Os resíduos deviance nos modelos de regressão paramétricos são definidos por:

$$\hat{d}_i = \text{sign}(\hat{m}_i) [-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{1/2}$$

Esses resíduos, que são uma tentativa de tornar os resíduos Martingal mais simétricos em torno de zero, facilitam, em geral, a detecção de pontos atípicos (*outliers*). Se o modelo for apropriado, estes resíduos deveriam apresentar um comportamento aleatório em torno de zero. Gráficos dos resíduos Martingal, ou Deviance, versus o tempo fornecem assim, uma forma de verificar a adequação do modelo ajustado bem como auxiliam na detecção de observações atípicas.

Para a seleção do melhor modelo que se adequa ao conjunto de dados, critério de seleção de modelos como o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC), são frequentemente utilizados em diversas áreas. Segundo esses critérios, o melhor modelo será aquele que apresentar menor valor de AIC e BIC. O critério de informação de Akaike (AIC) proposto por Akaike (1974) é uma medida relativa da qualidade de ajuste de um modelo paramétrico, estimado pelo método de máxima verossimilhança. A estatística AIC é dada por:

$$AIC = -2\log l(\hat{\theta}) + 2(p)$$

em que $l(\hat{\theta})$ denota o logaritmo da função de verossimilhança e p o número de parâmetros do modelo. Já o Critério de Informação Bayesiano (BIC) proposto por Schwarz et al. (1978) é dado por:

$$BIC = -2\log l(\hat{\theta}) + p \log n$$

em que p é o número de parâmetros a serem estimados e n é o número de observações da amostra. Os dois critérios comparam modelos que não são encaixados ou mesmo com número diferentes

de parâmetros, pois consideram o número de parâmetros e penalizam a verossimilhança de modelos com muitos parâmetros.

2.5 O Modelo de Fragilidade

O modelo de fragilidade é caracterizado pela utilização de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou não foram observadas tais como: fatores ambientais, genéticos, informações que, por algum motivo, não foram consideradas no planejamento. A fragilidade é introduzida na função de risco de forma multiplicativa ou aditiva (Wienke, 2010). O modelo de fragilidade engloba duas fontes de variações que geram a heterogeneidade entre as observações causada por covariáveis individuais não observadas que não foram incluídas no planejamento do estudo por circunstâncias práticas, ou por não serem conhecidas como sendo fatores de risco e aquela proveniente das covariáveis comuns a indivíduos de um mesmo grupo ou família que, quando não são observadas, geram dependência entre os tempos de eventos (Tomazella, 2003). As fragilidades podem ser introduzidas de forma multiplicativa ou de forma aditiva na função de risco, visando assim responder as diferentes formas de avaliar a influência da heterogeneidade entre as unidades na função de risco, ou intensidade nos processos de contagem.

Existem duas categorias de modelos de fragilidade, que são os modelos de fragilidades para dados univariadas e multivariadas. Os modelos de fragilidade para dados de sobrevivência univariado leva em conta que a população é não homogênea. A heterogeneidade pode ser explicada por covariáveis, mas quando importantes covariáveis não são incorporadas no modelo, isso conduz a heterogeneidade não observada. Já os modelos de fragilidade para dados de sobrevivência multivariado é muito comum para indivíduos com eventos repetidos ou dados de eventos recorrentes ou causas competitivas (Colosimo e Giolo, 2006). O modelo de fragilidade compartilhado remonta a Clayton (1978) e leva em consideração a associação entre os tempos de

sobrevivência dos indivíduos dentro de cada grupo. A fragilidade representa, nesses casos, um efeito aleatório que descreve o risco comum, isto é, a fragilidade compartilhada por indivíduos dentro de um mesmo grupo ou família.

O modelo de fragilidade multiplicativo é uma extensão do modelo de Cox (Cox, 1972), onde o risco individual depende de uma variável aleatória não observável, não negativa, v , a qual age de forma multiplicativa na função de risco de base. A função de risco sem a presença de covariáveis no instante t para o i -ésimo indivíduo é dada por

$$h_i(t) = h_i(t|v) = v_i h_0(t), \quad i = 1, \dots, n,$$

em que $h_0(t)$ é a função de risco base comum a todos os indivíduos e v_i é uma variável aleatória não-negativa independente e identicamente distribuída. A correspondente função de sobrevivência é dada por

$$S_i(t|v_i) = [S_0(t)]^{v_i}$$

em que $S_0(t)$ é a função de sobrevivência comum à população, isto é, representa a probabilidade do indivíduo esteja vivo em t dado o efeito aleatório v_i .

3 REFERÊNCIAS

- Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1):106–108.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Andersen, P. K., Borgan, O., Gill, R. D., e Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Baccari Júnior, F. (1990). Métodos e técnicas de avaliação da adaptabilidade dos animais às condições tropicais. *Simpósio Internacional de Bioclimatologia Animal nos Trópicos: pequenos e grandes ruminantes*, 1:9–17.
- Bolfarine, H., R. J. A. J. (1991). *Análise de Sobrevivência*. Rio de Janeiro, 2a Escola de Modelos de Regressão.
- Bonetti, O., Rossoni, A., e Nicoletti, C. (2009). Genetic parameters estimation and genetic evaluation for longevity in italian brown swiss bulls. *Italian Journal of Animal Science*, 8(sup2):30–32.
- Breslow, N. (1972). Contribuição à discussão do artigo de dr cox. *Journal of the Royal Statistical Society, Series B*, 34:216–7.
- Caetano, S., Rosa, G., Savegnago, R., Ramos, S., Bezerra, L., Lôbo, R., De Paz, C., e Munari, D. (2013). Characterization of the variable cow's age at last calving as a measurement of longevity by using the kaplan–meier estimator and the Cox model. *Animal*, 7(4):540–546.
- Carvalho, J. (1984). Relatório de atividades do núcleo de preservação do gado pé-duro ou curraleiro. *EMBRAPA/UEPAE. Teresina*.
- Carvalho, J., Monteiro, F. d. C., e Girão, R. N. (2001). Conservação do bovino pé-duro ou curraleiro: situação atual. *Embrapa Meio-Norte-Documentos (INFOTECA-E)*.

- Carvalho, J. H. (2002). Potencial econômico do bovino pé-duro. *Embrapa Meio-Norte- Documentos (INFOTECA-E)*.
- Carvalho Dias, A. (1997). Estudo de características produtivas e de permanência em um rebanho de bovinos da raça caracu. *Jaboticabal: Universidade Estadual Paulista*.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.
- Colosimo, E. A. e Giolo, S. R. (2006). Análise de sobrevivência aplicada. *São Paulo: Editora Edgard Blucher*.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Cox, D. R. e Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.
- Ducrocq, V., Quaas, R., Pollak, E., e Casella, G. (1988). Length of productive life of dairy cows. 2. variance component estimation and sire evaluation. *Journal of Dairy Science*, 71(11):3071–3079.
- Elliott, J., Rawlings, J., Markwell, P., e Barber, P. (2000). Survival of cats with naturally occurring chronic renal failure: effect of dietary management. *Journal of Small Animal Practice*, 41(6):235–242.
- Fioravanti, M. C. S., Juliano, R. S., Costa, G. L., Abud, L. J., Cardoso, V. S., Carpio, M. G., Oliveira, M. F., et al. (2011). Conservación del bovino curraleiro: cuantificación del censo y caracterización de los criadores. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, 48:109–116.

- Forabosco, F., Jakobsen, J., e Fikse, W. (2009). International genetic evaluation for direct longevity in dairy bulls. *Journal of dairy science*, 92(5):2338–2347.
- Giolo, S. R. (2003). *Variáveis latentes em análise de sobrevivência e curvas de crescimento*. PhD thesis, Universidade de Sao Paulo.
- Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.
- Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71(1):75–83.
- Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, 1(3):255–273.
- Imbayarwo-Chikosi, V., Dzama, K., Halimani, T., Van Wyk, J., Maiwashe, A., e Banga, C. (2015). Genetic prediction models and heritability estimates for functional longevity in dairy cattle. *South African Journal of Animal Science*, 45(2):105–121.
- Kaplan, E. L. e Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Kern, E. L., Cobuci, J. A., Costa, C. N., e Ducrocq, V. (2016). Survival analysis of productive life in brazilian holstein using a piecewise weibull proportional hazard model. *Livestock Science*, 185:89–96.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, vol.362, John Wiley e Sons.
- Louzada-Neto, F. e Pereira, B. d. B. (2000). Modelos em análise de sobrevivência. *Cad. saúde colet.,(Rio J.)*, pages 9–26.
- Mantel, N. e Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, 22(4):719–748.

- Mudholkar, G. S., Srivastava, D. K., e Kollia, G. D. (1996). A generalization of the weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, 91(436):1575–1583.
- Nogueira Neto, A. (1980). Aspectos da pecuária piauiense. *Sociedade de Medicina Veterinária do Piauí. Teresina*.
- Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(3):414–422.
- Peto, R. (1972). Contribuição à discussão do artigo de dr Cox. *Journal of the Real Statistics Society B*, 34:205–207.
- Salles, P., Medeiros, G., Costa, R. G., Ramos, C. T. C., Borburema, J., Rocha, M., Rocha, L., e Mathias, W. (2011). Programa de conservação e melhoramento de uma raça bovina brasileira: Curraleiro Pé-Duro. *AICA-Actas Iberoamericana de Conservacion animal, Córdoba*, 1(1):453–6.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Sinha, D. e Dey, D. K. (1997). Semiparametric bayesian analysis of survival data. *Journal of the American Statistical Association*, 92(439):1195–1212.
- Stacy, E. W. et al. (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, 33(3):1187–1192.
- Struthers, C. A. e Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, 73(2):363–369.
- Tomazella, V. L. D. (2003). *Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade*. PhD thesis, Universidade de São Paulo.

- Van Melis, M., Eler, J., Rosa, G., Ferraz, J., Figueiredo, L., Mattos, E., e Oliveira, H. N. d. (2010). Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in nellore cattle. *Journal of Animal Science*, 88(12):3809–3813.
- Vaupel, J. W., Manton, K. G., e Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.
- Wienke, A. (2010). *Frailty models in survival analysis*. Chapman and Hall/CRC.

4 CAPÍTULO 1

MODELO DE COX-LOG NORMAL PARA AVALIAR A LONGEVIDADE EM BOVINOS DA RAÇA CURRALEIRO PÉ-DURO*

* Segundo as normas da revista

Modelo de Cox-Log Normal para avaliar a longevidade em bovinos da raça Curraleiro Pé-Duro

Cox-Log Normal model to assess longevity in Curraleiro Pé-Duro cattle breed

Abstract

Cattle of the Curraleiro Pé Duro breed are characterized as very docile and exceptionally rustic animals. They are animals gradually adapted to low quality native pastures, drought conditions, high temperatures and other adverse factors, resulting in the formation of bovine herds adapted and resistant to the unfavorable conditions of the Brazilian hinterland. In this context, the objective of this work was to use the survival analysis methodology with a Cox-lognormal proportional hazards model, to assess the stayability in the herd, considering the failure as the inactivity of the bovine being caused by death or sale, using information from 102 Curraleiro Pé-Duro cattle born in the period from 2005 to 2014 in an experimental field of Embrapa Meio-Norte located in São João do Piauí. The Cox-lognormal proportional hazards model considered, among the analyzed covariates, the season of birth, sex and weight at 365 days as significant. There was a predominance of birth in the dry season (july to december). It was observed that the cattle that longer remained in the herd were born in the dry season, were male and weighed less than 95 kg at 365 days, justified by the absence of parasites, availability of pasture for the mother during the pregnancy period, carcass evaluation, sale, slaughter and reproduction. The Cox-lognormal proportional hazards model proved to be adequate in the adjustment of the statistical model to estimate the stayability in the herd, with censorship related to the inactivity of the Curraleiro Pé Duro breed.

Keywords: Adaptability, survival analysis, censorship, failure.

Resumo

Os bovinos da raça Curraleiro Pé-Duro são caracterizados como animais muito dóceis e de excepcional rusticidade. São animais gradativamente ambientados às pastagens nativas de baixa qualidade, às condições de estiagem, às temperaturas elevadas entre outros fatores adversos, resultando na formação de rebanhos bovinos adaptados e resistentes às condições desfavoráveis do sertão brasileiro. Nesse contexto, objetivou-se com esse trabalho, utilizar a metodologia de análise de sobrevivência com modelo de riscos

proporcionais de Cox-Log Normal, para avaliação do tempo permanência no rebanho, considerando a falha como sendo a inatividade do bovino sendo ocasionada por morte ou venda, utilizando-se informações de 102 bovinos da raça Curraleiro Pé-Duro nascidas no período de 2005 a 2014 em um campo experimental da Embrapa Meio-Norte situado em São João do Piauí. O modelo de riscos proporcionais de Cox-Log Normal, considerou dentre as covariáveis analisadas, a estação de nascimento, sexo e peso aos 365 dias como sendo significativas ($p\text{-valor}<5\%$). Houve uma predominância do nascimento na estação seca (julho a dezembro). Foi observado que os bovinos que mais permaneceram no rebanho nasceram na estação seca, eram machos e tiveram um peso menor que 95 kg aos 365 dias justificado pela ausência de parasitas, disponibilização de pastagens para a mãe no período de prenhez, avaliação de carcaça, venda, abate e reprodução. O modelo de riscos proporcionais de Cox-Log Normal se mostrou adequado no ajuste do modelo estatístico para estimar o tempo de permanência no rebanho, com censura relacionada a inatividade do bovino Curraleiro Pé-Duro.

Palavras-chave: Adaptabilidade, análise de sobrevivência, censura, falha.

1. Introdução

A raça Curraleiro Pé-Duro, reconhecida pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) em 2012, é a primeira raça selecionada no Brasil. Essa raça local, também denominada nativa, naturalizada ou crioula, é patrimônio histórico e cultural no Piauí e auxiliou o desenvolvimento econômico do país como meio de transporte, fonte de proteína e comércio de charque e couro. A atenção dos pesquisadores pela referida raça ocorreu a partir da observação de como estes bovinos gradativamente se ambientaram às pastagens nativas de baixa qualidade, às condições de estiagem, às temperaturas elevadas entre outros fatores adversos à criação de bovinos no Brasil. Esses mecanismos de adaptação ocorreram ao longo dos séculos resultando na formação de rebanhos bovinos adaptados e resistentes às condições desfavoráveis do sertão brasileiro. A habilidade dos animais em adaptar-se às condições climáticas extremas (altas temperaturas, seca) em ambiente natural favoreceu a minimização de perdas no desempenho produtivo e reprodutivo, auxiliando na resistência a doenças, consequentemente reduzindo a mortalidade e aumentando a longevidade quando expostos a condições de estresse (Baccari Júnior, 1990).

A Análise de Sobrevivência, por sua vez, é considerada metodologia estatística mais adequada para lidar com dados de tempo até a ocorrência de um evento de interesse (tempo de falha), na presença de censura (Colosimo e Giolo, 2006) que é sua principal característica. Na área de produção animal, as características de sobrevivência são: longevidade (idade do animal ao descarte); duração da vida produtiva (intervalo do primeiro parto ao descarte); número de lactações durante a vida e habilidade de permanência no rebanho ou *staybility*. A metodologia de análise de sobrevivência, já vem sendo utilizada por alguns autores no estudo da longevidade em bovinos leiteiros. Uma característica especial associada aos dados de sobrevivência é a possibilidade de que, para alguns indivíduos, o tempo completo até a ocorrência do evento de interesse não seja observado, devido a diversas causas. Por exemplo, um indivíduo pode permanecer em acompanhamento até o final do tempo de estudo sem ter experimentado o evento de interesse. A não consideração destes indivíduos com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes. Portanto, a importância da introdução de uma variável na análise que indique se o tempo de sobrevivência foi observado (Louzada-Neto e Pereira, 2000). Essa variável é definida na literatura como variável indicadora de censura. A censura é dita ser do tipo I quando ocorre devido ao término do estudo após um período de tempo pré-determinado; do tipo II, quando ocorre devido ao término do estudo após um número de falhas fixado previamente; ou pode ser aleatória, sendo a mais comum em situações práticas. As censuras previamente citadas são conhecidas como censura à direita, pois a falha ocorre sempre à direita do tempo registrado. Existem ainda outros mecanismos de censura, como, por exemplo, as censuras à esquerda, em que o tempo registrado é maior que o tempo de falha, e censura intervalar, na qual não se sabe o tempo exato de falha, sendo que a única informação disponível é que o tempo de falha ocorreu em um certo intervalo de tempo. Nesta pesquisa, entretanto, será adotado apenas o mecanismo de censura à direita que será denominado por “censura”.

O objetivo principal com esse estudo foi avaliar o tempo de sobrevivência ou permanência no rebanho, do bovino raça Curraleiro Pé-Duro, relacionado a saída do rebanho por morte ou venda, usando a metodologia dos modelos de Cox paramétrico com função de risco base Log Normal.

2. Material e Métodos

2.1 Description of the data

Os dados deste estudo, foram cedidos pelo núcleo de conservação de bovinos da raça Curraleiro Pé-Duro (CPD) pertencentes a Embrapa Meio-Norte, em Teresina-Piauí-Brazil, com campo experimental situado na fazenda Otávio Domingues, em São João do Piauí (entre 8° 26' e 8° 54' latitude Sul e entre 42° 19' e 42° 45' longitude Oeste), no semiárido piauiense pertencente a um rebanho de conservação *in situ*. Os bovinos são criados extensivamente com o fornecimento apenas de sal, mineral e água, o que justifica seu baixo peso se comparado a bovinos da raça CPD criado em outras regiões e de outras raças. Ainda, devido a falta de suplementação alimentar, existe um aumento da incidência de efeitos ambientais, com a presença de muitas plantas tóxicas como o barbatimão (*Stryphnodendron coriaceum*) e a erva-de-rato (*Palicourea marcgravii*), carrapatos, verminoses, dentre outras.

Foram avaliados 102 bovinos (58 Machos e 44 fêmeas) da raça CPD desde o nascimento até aos 550 dias de idade. Os dados dos animais foram coletados entre o período de 2005 a 2014. Para modelagem do tempo de sobrevivência (em meses), após início de vida reprodutiva dos bovinos até a ocorrência da falha (inatividade ocasionada pela morte ou venda), em relação ao bovino permanecer ativo (vivo), considerou-se o tempo como variável resposta, e a data de nascimento do bezerro como o início do estudo. A variável T (tempo) foi obtida a partir da diferença entre a data de nascimento e a data do descarte. A data do último descarte no rebanho foi considerada como período final de observação para animais que ainda não haviam sido descartado (19/08/2016). A falha foi definida como inatividade do bovino (morte ou venda), enquanto a censura foi definida para bovinos que permaneceram vivos no rebanho. O tipo de censura utilizado foi a direita. Foi registrado, para cada animal observado, um correspondente indicador de censura, denominado status ($\delta = 1$ se falha e $\delta = 0$ se censura) indicando se o animal está ativo ou inativo no rebanho.

A variável estação foi considerada como sendo chuvosa os meses que compreenderam de janeiro a junho e seca os meses de julho a dezembro. Todos os 102 bezerros foram acompanhados desde o nascimento. As covariáveis consideradas como possíveis fatores de riscos no tempo de permanência dos bovinos CPD no rebanho, foram: EN: estação de nascimento (0 – chuvosa, 1 – seca), S: sexo (0 – macho, 1 – fêmea), PN: peso ao nascer (0 – $< 20\text{kg}$, 1 – $\geq 20\text{kg}$), PD: peso a desmama (0 – $< 66\text{kg}$, 1 – $\geq 66\text{kg}$),

P365: peso aos 365 dias ($0 - < 95\text{kg}$, $1 - \geq 95\text{kg}$), P550: peso aos 550 dias ($0 - < 131\text{kg}$, $1 - \geq 131\text{kg}$) categorizados de acordo com o peso médio em cada fase (Carvalho et al., 2015).

2.2 Statistical Analysis

O modelo de regressão de Cox foi proposto por Cox (1972), onde permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse, ajustando por covariáveis (Colosimo e Giolo, 2006). O modelo proposto por Cox modela diretamente a função de risco e tem como princípio básico deste modelo para estimar o efeito das covariáveis, a proporcionalidade dos riscos ao longo de todo o tempo de observação.

O modelo de riscos proporcional, é dado por:

$$h(t|\mathbf{x}) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}) \quad (1)$$

em que $h_0(t)$ é a função de risco de base, isto é o risco de um animal com covariáveis iguais a zero; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$ é o vetor de dimensão k de coeficientes de regressão desconhecidos e $\mathbf{x} = (x_1, \dots, x_k)$ é o vetor de dimensão k de covariáveis para a i -ésima observação.

O modelo (1) é composto pelo produto de dois componentes: $\exp(\boldsymbol{\beta}'\mathbf{x})$ que mede o efeito das covariáveis e $h_0(t)$ que pode ser paramétrico ou não. No caso semiparamétrico o componente $h_0(t)$ é considerado ser uma função não negativa do tempo; quando paramétrico, pode assumir distribuições como Gompertz, Log Normal, gama entre outras. Este modelo é também denominado modelo de risco proporcionais, pois a razão das taxas de falha de dois indivíduos diferentes é constante no tempo. Isto é, a razão das funções de taxa de falha para os indivíduos i e j dada por:

$$\frac{h(t|x_i)}{h(t|x_j)} = \frac{h_0(t)\exp\{\boldsymbol{\beta}'x_i\}}{h_0(t)\exp\{\boldsymbol{\beta}'x_j\}} = \exp\{\boldsymbol{\beta}'(x_i - x_j)\}, \quad (2)$$

não depende do tempo. Por exemplo, se um indivíduo no início do estudo tem um risco de morte igual a duas vezes o risco de um segundo indivíduo, então, esta razão de riscos será a mesma para todo o período de acompanhamento (Colosimo e Giolo, 2006).

A interpretação dos coeficientes em um modelo de regressão de Cox é realizada através das quantidades $\exp(\beta_i)$, chamadas de razão de taxas de risco ou falhas (RTF). Um valor de β_i maior que zero, ou equivalente a uma taxa de risco maior que um, indica que, como o valor de i -ésimas covariáveis, o risco

de falha aumenta e, portanto, a duração da sobrevivência diminui. Assim, se $RTF = 1$ é dita sem efeito; se $RTF < 1$ houve uma redução do risco e se $RTF > 1$ um aumento do risco. No modelo de Cox, a função de sobrevivência é definida por:

$$S(t_i | \mathbf{x}_i) = \exp \left\{ - \int_0^{t_i} h_0(u) \exp\{\boldsymbol{\beta}' \mathbf{x}\} du \right\} = [S_0(t_i)]^{\exp\{\boldsymbol{\beta}' \mathbf{x}\}} \quad (3)$$

em que $S_0(t_i)$ é a função de sobrevivência base, ou seja, a função de sobrevivência de um animal com todas as covariáveis iguais a zero. A função de risco acumulada base, $H_0(t)$, é dada por $H_0(t) = -\log(S_0(t))$.

Neste estudo, é considerada a abordagem paramétrica para o modelo de Cox, onde os tempos de vida dos bovinos em risco seguem distribuição Log Normal. A Tabela 1 mostra as funções de densidade de probabilidade $f(t)$, de risco $h_0(t)$ e de sobrevivência $S_0(t)$ para as distribuições Log Normal, Log-logístico e Gompertz, que servem como comparação para o modelo proposto Cox-Log Normal.

Tabela 1

Funções densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_0(t)$ das distribuições Log Normal, Log-logistic Gompertz.

Distribuição	$f(t)$	$h_0(t)$	$S_0(t)$
Log Normal	$\frac{1}{\sqrt{2\pi} t\theta} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right]$	$\frac{1}{\sqrt{2\pi} t\theta} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right] \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)$	$\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)$
Log-logistic	$\frac{\gamma}{\lambda^\gamma} t^{\gamma-1} \left(1 + \left(\frac{t}{\lambda} \right)^\gamma \right)^{-2}$	$\frac{\gamma \left(\frac{t}{\lambda} \right)^{\gamma-1}}{\lambda \left[1 + \left(\frac{t}{\lambda} \right)^\gamma \right]}$	$\frac{1}{\left[1 + \left(\frac{t}{\lambda} \right)^\gamma \right]}$
Gompertz	$\lambda \exp(\gamma t) \exp \left\{ -\left(\frac{\lambda}{\gamma} \right) (e^{\gamma t} - 1) \right\}$	$\lambda \exp(\gamma t)$	$\exp \left\{ -\left(\frac{\lambda}{\gamma} \right) (e^{\gamma t} - 1) \right\}$

Assim, substituindo as funções de risco base $h_0(t)$ e sobrevivência base $S_0(t)$, da Log Normal, respectivamente em (1) e (3), temos o modelo de riscos proporcionais de Cox- Log Normal cuja função de risco fica assim definida:

$$h(t|x) = \frac{\frac{1}{\sqrt{2\pi} t\theta} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right]}{\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)} \exp(\boldsymbol{\beta}' \mathbf{x}) \quad (4)$$

e a função de sobrevivência é reescrita por

$$S(t_i|\mathbf{x}_i) = \left[\Phi \left(\frac{-\log(t)+\mu}{\sigma} \right) \right]^{exp\{\boldsymbol{\beta}'\mathbf{x}_i\}} \quad (5)$$

Para inferência dos parâmetros foi utilizado o método da máxima verossimilhança. Supondo que os dados sejam independentes e identicamente distribuídos, a função de verossimilhança, com dados censurados, segundo Lawless (2011) é definida por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [h(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i) \quad (6)$$

em que δ_i é o indicador de censura. Assim, o conjunto de dados $G = (n, t, \delta, X)$ é formado pelos tempos de falha observados, tempos de falha censurados e a matriz $n \times k$ contendo as covariáveis, respectivamente. Substituindo (4) e (5) em (6), temos

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi} t \theta} \exp \left[-\frac{1}{2} \left(\frac{\log(t)-\mu}{\sigma} \right)^2 \right] \right]^{\delta_i} \exp(\boldsymbol{\beta}'\mathbf{x}_i) \left[\Phi \left(\frac{-\log(t)+\mu}{\sigma} \right) \right]^{exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}$$

As estimativas de máxima verossimilhança são obtidas através da maximização numérica da função log-verossimilhança ($\log[L(\boldsymbol{\theta})]$). Foi utilizado o pacote *optim* do *software* R e o método "BFGS" para maximização. A construção de intervalos de confiança diz respeito à distribuição assintótica do estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$. Para grandes amostras, sob certas condições de regularidade, a distribuição do vetor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)'$ é normal multivariada de média $\boldsymbol{\theta}$ e matriz de variância-covariância $Var(\hat{\boldsymbol{\theta}})$, isto é,

$$\hat{\boldsymbol{\theta}} \approx N_k(\boldsymbol{\theta}, Var(\hat{\boldsymbol{\theta}}))$$

Para a construção de intervalos de confiança é necessário uma estimativa para o erro padrão de $\hat{\boldsymbol{\theta}}$, isto é, para $[Var(\hat{\boldsymbol{\theta}})]^{1/2}$. No caso especial em que $\boldsymbol{\theta}$ é um escalar, um intervalo aproximado de $(1 - \alpha)100\%$ de confiança para $\boldsymbol{\theta}$ é dado por

$$\hat{\boldsymbol{\theta}} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\boldsymbol{\theta}})}$$

Para um modelo com um vetor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ de parâmetros, muitas vezes há o interesse em testar hipóteses relacionadas a este vetor ou a um subconjunto dele. Três testes são em geral utilizados para esta finalidade: o de Wald, o da Razão de Verossimilhança (TRV) e o Escore (Colosimo e Giolo, 2006). Critérios de seleção de modelos como o critério de informação de Akaike (AIC) proposto por Akaike (1974), e o critério de informação bayesiano (BIC), proposto por Schwarz et al. (1978), são frequentemente utilizados

para selecionar modelos em diversas áreas. Os melhores modelos são considerados aqueles que apresentam menores valores de AIC e BIC. Para a adequação do modelo, diversos métodos encontram-se disponíveis na literatura e baseiam-se, essencialmente, nos resíduos de Cox-Snell, que auxiliam a examinar o ajuste global do modelo, o de Schoenfeld que tem coeficiente dependente do tempo, o de Martingal, que é dado pela diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, e de Deviance, que facilitam na detecção de pontos atípicos (*outliers*).

3. Resultados

Nesse estudo, o rebanho é formado por 102 bovinos (58 machos e 44 fêmeas) da raça CPD nascidos no período de 2005 a 2014. Ao considerar a sobrevivência desses animais no rebanho, constatou-se que, 28% permaneceram vivos até o período final da observação, isto é, 19/08/2016, sendo esses considerados como censura e 72% falharam, isto é, morreram ou foram vendidos. No referido trabalho, houve predominância do nascimento na estação seca que engloba os meses de julho a dezembro, correspondendo a 65,7%. Os animais que falharam (morte ou venda) ocorreram com maior frequência em relação aos animais que permaneceram vivos, isto é, foram censurados. Em relação ao tempo de vida, foi apresentado a Figura 1 que representa a função de sobrevivência empírica estimada pelo estimador não-paramétrico de Kaplan-Meier. É possível verificar que 50% dos bovinos apresentaram o tempo de vida maior que 63,5 meses.

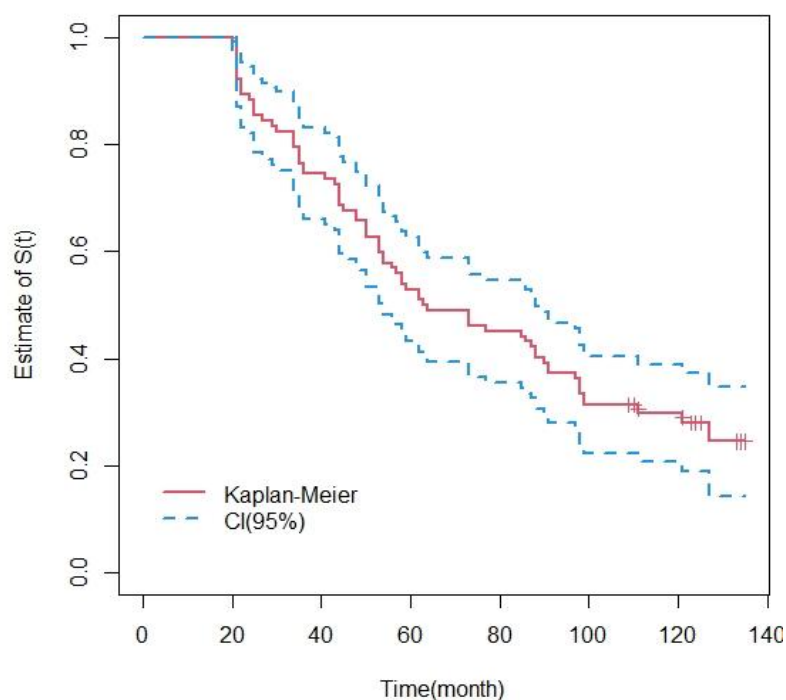


Figura 1. Curva de sobrevivência global estimada pelo método Kaplan-Meier.

A fim de analisar se existe diferença entre a sobrevivência das categorias para as possíveis covariáveis, as referidas curvas de sobrevivência estratificadas são dadas na Figura 2.

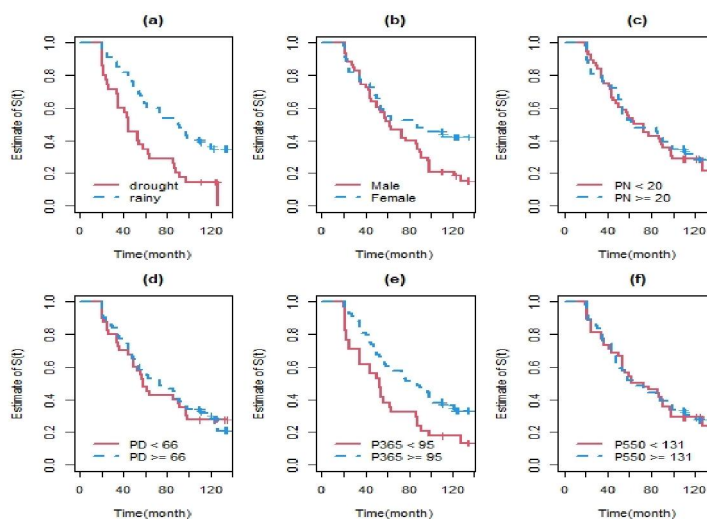


Figura 2 Curvas de sobrevivências estimadas pelo método de Kaplan-Meier para as possíveis covariáveis: Estação de Nascimento (EN), Sexo (S), Peso ao Nascer (PN), a Desmama (PD), aos 365 dias (P365) e aos 550 dias (P550) dos bovinos da raça Curraleiro Pé-Duro.

A análise empírica sugere que existe diferença entre a sobrevivência das categorias para as covariáveis EN e P365 e uma leve diferença na covariável S. Para verificar se há evidência estatística da diferença, foi realizado o teste Log-rank sob a hipótese de igualdade das curvas de sobrevivência. A partir do teste de Log-rank, adotando-se um nível de significância menor ou igual a 10%, observou-se significância para a EN (p-valor = 0,0004), S (p-valor = 0,04), e o P365 (p-valor = 0,008). Para as demais covariáveis, não se verificou efeito significativo entre as categorias das covariáveis, isto é, para o PN (p-valor = 0,7), PD (p-valor = 0,7) e P550 (p-valor = 0,9) para a estimativa das probabilidades de sobrevivência.

Neste trabalho, foi utilizado a distribuição Log Normal e para fim de comparação as distribuições Gompertz e Log-logístico. A construção do modelo de Cox- Log-Normal para analisar a influência das covariáveis sobre o tempo de permanência dos bovinos no rebanho desde o nascimento, partiu primeiramente, de um modelo sem covariável e, em seguida, avaliamos a inclusão das covariáveis e interações, usando o critério do teste da razão de verossimilhança (TRV). Entraram no modelo as variáveis que se mantiveram significativas considerando $\alpha = 10\%$. São elas: EN, S e P365.

De acordo com a Tabela 2, a taxa de falha na estação de nascimento seca foi de 0,2811 vezes, indicando que há diminuição do risco dos bovinos falhar (morte ou venda), dentre as que conseguem se manterem vivas demonstrando uma maior longevidade do bovino no rebanho. Para a variável S, a taxa de falha para as fêmeas foi 4,8602 vezes a taxa de falhas para os machos, ou seja, bezerros fêmeas tem 4,8602 vezes mais chances de falhar (morte ou venda) do que os bezerros machos. Para a interação da variável S com o P365 foi observado uma taxa de falha de 0,0466 indicando uma diminuição do risco dos bovinos virem a falhar, ou seja, morrer ou sair para venda.

Tabela 2

Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), p -valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de Cox- Log Normal

Parâmetros	EMV	EP	CI (95%)	p -valor	RTF
μ	3,9590	0,1329	[3,69; 4,21]	< 2e-16	-
σ	0,5804	0,0627	[0,45; 0,70]	< 2e-16	-
β_1 (EN)	-1,2689	0,2715	[-1,80;-0,73]	2,96e-06	0,2811
β_2 (Sexo)	1,5811	0,4055	[0,78; 2,37]	9,67e-05	4,8602
β_3 (P365)	0,7476	0,3358	[0,08; 1,40]	0,026	2,1119
β_4 (S* P365)	-3,0648	0,5640	[-4,17;-1,95]	5,50e-08	0,0466

As funções de risco em (4) e sobrevivência em (5) para o modelo final são descritas, respectivamente, por:

$$\hat{h}(t|x) = \frac{\frac{1}{\sqrt{2\pi} \cdot 0,5804t} \exp\left[-\frac{1}{2} \left(\frac{\log(t)-3,9590}{0,5804}\right)^2\right]}{\Phi\left(\frac{-\log(t)+3,9590}{0,5804}\right)} g(\hat{\beta}'x_i)$$

e

$$\hat{S}(t_i|x_i) = \left[\Phi\left(\frac{-\log(t) + 3,9590}{0,5804}\right) \right]^{g(\hat{\beta}'x_i)}$$

onde $g(\hat{\beta}'x_i) = \exp(-1,2689 \times EN + 1,5811 \times S + 0,7476 \times P365 - 3,0648 \times S * P365)$.

Para que o modelo possa está adequado ao estudo, os pressupostos da proporcionalidade para modelo de Cox precisou ser verificado. Através do método gráfico descritivo, que envolve o logaritmo da função de falha acumulada versus o tempo para as covariáveis: EN (a), S (b) e P365 (c), observou-se que as covariáveis EN e P365 atende perfeitamente a proporcionalidade, no entanto, para o sexo, embora as mesmas no início não sejam perfeitamente paralelas, não existem, em termos descritivos, violação da suposição de taxas de falhas proporcionais. (Figura 3)

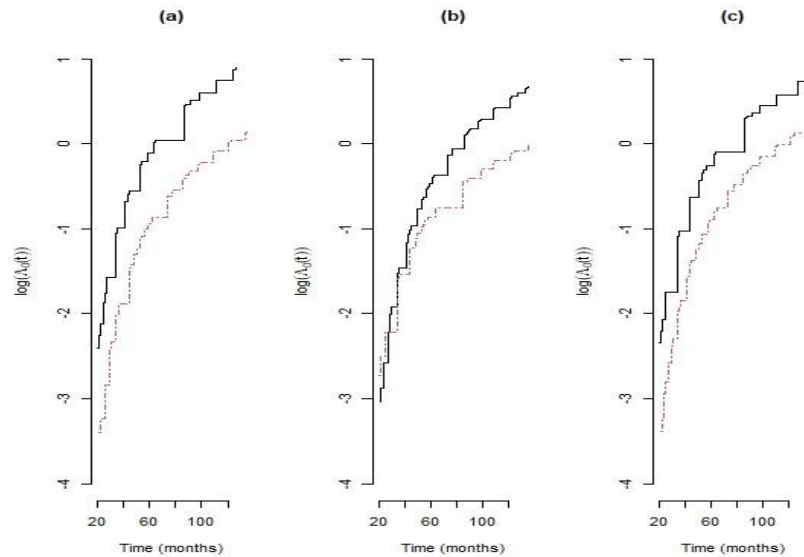


Figura 3 Logaritmo da taxa de falha versus tempo para as covariáveis EM (a), Sexo(b) e P365(c)

Para verificação da qualidade do modelo ajustado e possíveis *ouliers*, os gráficos dos resíduos de Cox-Snell, resíduos martingale e resíduos deviance são construídos (Figura 4). Os resíduos de Cox e Snell (1968) auxiliam a examinar o ajuste global do modelo paramétrico. Quanto mais próximas as curvas de sobrevivência obtidas por Kaplan-Meier e pelo modelo exponencial padrão, melhor é considerado o ajuste do modelo aos dados. Assim, as curvas de sobrevivência para os resíduos de Cox-Snell são apresentados pelos gráficos (a) e (b) indicando uma boa qualidade do ajuste global do modelo paramétrico de Cox-Log-Normal. Os Gráficos dos resíduos Martingale (c) e resíduos Deviance (d) sugeriram a ausência *ouliers*.

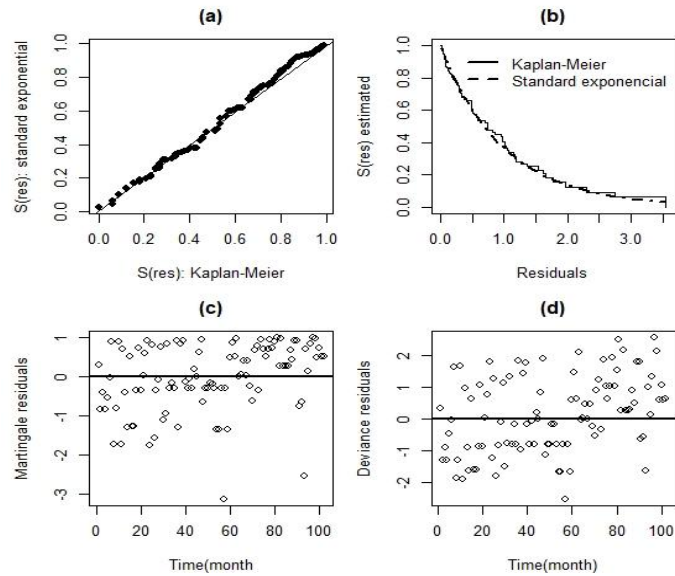


Figura 4 Resíduos de Cox-Snell , Martingal e Deviance do modelo de modelo de Cox-Log Normal para dados de bovinos da raça Curraleiro Pé-Duro

Para fins comparativo, foi realizado um estudo para o modelo de riscos proporcionais de Cox- Log Normal com os modelos de Cox-Log logístico e Cox-Gompertz. A Tabela 3 mostra as estimativas de máxima verossimilhança para os parâmetros dos modelos propostos.

Tabela 3

Estimativas para os parâmetros do modelo de riscos proporcionais de Cox- Log Normal, Cox-Log logístico e Cox-Gompertz

Parâmetros	Log Normal		Log-logístico		Gompertz	
	Estimativas	Erro Padrão	Estimativas	Erro Padrão	Estimativas	Erro Padrão
μ	3,9590	0,1329	-	-	-	-
σ	0,5804	0,0627	-	-	-	-
α	-	-	45,7043	1,3193	-	-
γ	-	-	3,1859	0,3661	0,0148	0,0036
λ	-	-	-	-	0,0092	0,0025
$\beta_1(EN)$	-1,2689	0,2715	-1,7489	0,4425	-1,4178	0,4014
$\beta_2(S)$	1,5811	0,4055	1,6139	0,3962	1,2463	0,3787
$\beta_3(P365)$	0,7476	0,3358	1,0220	0,4593	0,7954	0,4263
$\beta_4(S*P365)$	-3,0648	0,5640	-2,8626	0,5006	-2,5037	0,4872
Log-Veross	-372,9956		-374,4386		-388,4093	
AIC	757,9912		760,8772		788,8186	
BIC	773,741		776,6271		804,5685	

A seleção de modelos pelos critérios de Akaike (AIC) mostraram que apesar dos valores para a log-normal e log-logístico serem bem próximos se comparado ao Gompertz, o modelo de Cox- Log Normal foi o mais adequado, por apresentar o menor valor de AIC, o que indica que este modelo possui uma melhor qualidade de ajuste aos tempos de sobrevivência estimados.

4. Discussão

Vários são os fatores que interferem na permanência do bovino no rebanho. Podemos destacar o peso ao nascimento, que é um dos fatores mais importantes na rentabilidade do rebanho bovino e tem grande influência da sobrevivência ao desmame (Carvalho et al., 2015). O controle do peso ao nascer é de suma importância e pode ser feita de duas maneiras: através da nutrição e genética. O peso ao nascer tem influência de raça, sexo, ano (pluviosidade que afeta a disponibilidade de pastagens, ocorrência de doenças, mudanças de manejo e progresso genético) e mês de nascimento (época seca ou úmida, fornecimento de alimento pela mãe e pelo ambiente), efeitos paterno (tamanho) e materno (idade), além de possíveis interações entre variáveis. As estações têm grande influência no nascimento dos bezerros, sendo a seca considerada um ótimo período por ter menos doenças para os bezerros, além da oferta de nutrientes no pasto na época de prenhez para a mãe. Animais pesados no nascimento, geralmente são pesados na desmama devido a influência materna, isto é, bovinos com muito leite, tem bezerros mais pesados, com influência se estendendo até 1 ano. Depois desse período já é observada a genética do animal, quando é feita a seleção e avalia-se a adaptabilidade. A adaptabilidade, ou capacidade de se adaptar, pode ser avaliada pela habilidade do animal em se ajustar às condições ambientais médias, assim como aos extremos climáticos. Animais bem adaptados caracterizam-se pela manutenção ou mínima redução no desempenho produtivo, pela elevada eficiência reprodutiva, resistência às doenças, longevidade e baixa taxa de mortalidade durante a exposição ao estresse (Baccari Júnior, 1990).

No presente estudo, houve uma predominância de 65,7% de bezerros nascidos na estação seca, indicando que mais da metade dos bezerros nasceram em um período favorável para o seu desenvolvimento. Ainda, 50% dos bovinos apresentaram um tempo de vida maior que 63,5 meses.

A perspectiva abordada nessa pesquisa teve como suporte a longevidade dos bovinos e a saída ocasionada pela morte ou venda. Através da Figura 2, observou-se empiricamente, diferenças entre as

categorias das covariáveis EN, S e P365. Foi realizada a verificação da evidência estatística dessa diferença, através do teste log-rank ao nível de significância menor ou igual a 10%, sob a hipótese de igualdade das curvas de sobrevivência, onde as covariáveis EN, S e P365 se mantiveram significativas, isto é, existe diferença entre a estação seca e chuvosa, entre os machos e fêmeas e do peso aos 365 ser menor que 95 kg e igual ou maior que 95 kg.

A metodologia utilizada nesse estudo mostrou-se promissora e teve seus pressupostos de proporcionalidade atendidos (Figura 3), critério exigido para o uso dos modelos de riscos proporcionais de Cox e com resultados interessantes no estudo de avaliação de permanência dos animais no rebanho. Foi utilizado o modelo de Cox com risco base Log Normal onde foram encontradas significância (p-valor < 10%) as covariáveis: EN, S e P365 e a interação do S com P365, conforme Tabela 2. Ainda, de acordo com a Tabela 2, foi observado através da razão da taxa de falha, que para a covariável EN, houve uma diminuição na taxa de falha (morte ou venda) para os bovinos nascidos na estação seca que pode ser explicado por ter menos doenças para os bezerros, além da oferta de nutrientes no pasto na época de prenhez para a mãe, demonstrando uma maior longevidade para os bovinos. Para a variável S, foi encontrado uma taxa de falha maior para as fêmeas, explicada pela maternidade, ou seja, como as fêmeas dão cria anualmente, na ausência de suplemento, o organismo das mesmas direcionam o cálcio dos ossos para o leite durante a amamentação, tornando a vaca mais magra e mais sujeita a mortalidade. Para a variável P365, foi encontrado uma taxa de falha maior para os bovinos que tem um peso igual ou maior que 95 kg pois bovinos mais pesados saem do rebanho para venda, abate ou reprodução.

Para fins comparativo, foi realizado o estudo para o modelo de Cox com função de risco base, dada pelas distribuições Log-logístico e Gompertz, que apesar de ter estimativas bem próximas da Log Normal, tiveram um AIC e BIC maior, concluindo que o modelo que melhor se ajusta é o modelo de Cox- Log Normal por possuir AIC e BIC menor. A verificação da qualidade do ajuste foi realizada através dos testes Cox-Snell, Martingal e Deviance (Figura 4), indicando a adequação do modelo e a ausência de *outliers*.

4. Conclusões

O modelo de riscos proporcionais de Cox- Log Normal, se mostrou adequado na verificação de fatores observados que influenciam na longevidade desses animais, considerando-se a saída do bovino do

rebanho seja por morte ou venda. Os bezerros que nascem na estação seca, tendem a ser mais longevos. Em relação ao sexo, as fêmeas mostram serem menos longevos do que os machos. No entanto, bovinos que pesam aos 365 dias acima do peso médio (95 kg), tendem a ser menos longevos. Vale ressaltar que, outros fatores de difícil mensuração, não puderam ser avaliados com a aplicação do modelo, sendo que, modelos de fragilidades, que são uma extensão dos modelos de riscos proporcionais de Cox são usados para tal finalidade.

Declaração de Conflito de Interesse

Os autores declaram não haverem conflitos de interesse.

Financiamento

Esta pesquisa não recebeu nenhuma bolsa específica de agências de fomento nos setores público, comercial ou sem fins lucrativos.

Referências

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Baccari Júnior, F. (1990). Métodos e técnicas de avaliação da adaptabilidade dos animais às condições tropicais. *Simpósio Internacional de Bioclimatologia Animal nos Trópicos: pequenos e grandes ruminantes*, 1:9-17.
- Carvalho, G. M. C., Lima Neto, A., Da Frota, M. N. L., de Sousa, V. R., Carneiro, M. d. S., Vêras, G. M. C., Costa, R. N. M., e Azevedo, D. M. M. R. (2015). O uso de bovinos Curraleiro Pé-Duro em cruzamentos para produção de carne de boa qualidade no trópico quente-fase 1. In *Embrapa MeioNorte-Artigo em anais de congresso (Alice)*. In: Congresso nordestino de produção animal, 10., 2015, Teresina.
- Colosimo, E. A. e Giolo, S. R. (2006). *Análise de sobrevivência aplicada*. São Paulo: Editora Edgard Blucher.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187-202.

- Cox, D. R. e Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B(Methodological)*, 30(2):248–265.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.
- Louzada-Neto, F. e Pereira, B. d. B. (2000). Modelos em análise de sobrevivência. *Caderno de saúde coletiva (Rio Janeiro)*, pages 9–26.
- R Core Team 2016. R: a language and environment for statistical computing. Retrieved on 9 September 2019 from <https://www.R-project.org/>
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

5 CAPÍTULO 2

AVALIAÇÃO DA LONGEVIDADE DE BOVINOS DA RAÇA CURRALEIRO PÉ-DURO USANDO O MODELO DE FRAGILIDADE GAUSSIANA INVERSA *

* Segundo as normas da revista

Avaliação da longevidade de bovinos da raça Curraleiro Pé-Duro usando o modelo de fragilidade Gaussiana inversa

Resumo. O bovino da raça Curraleiro Pé-Duro possui uma grande habilidade em adaptar-se às condições climáticas extremas em ambiente natural proporcionando a minimização de perdas no desempenho produtivo e reprodutivo, auxiliando na resistência a doenças, consequentemente reduzindo a mortalidade e aumentando a longevidade quando expostos a condições de estresse. Neste contexto, objetivou-se com este trabalho fazer uso da técnica de análise de sobrevivência, analisando os efeitos de fatores não observados, em estudo sobre o tempo de permanência no rebanho, de bovinos da raça Curraleiro Pé-Duro utilizando-se do modelo de fragilidade Gaussiana inversa. Foram observados dados de registros de 102 bovinos nascidos no período de 2005 a 2014, em um rebanho experimental da Embrapa Meio-Norte situado em São João do Piauí, no semiárido piauiense. Considerou-se a falha como sendo a inatividade do bovino ocasionada pela morte ou venda e a censura como sendo o animal permanecer ativo no rebanho. Foi utilizada a metodologia dos modelos de fragilidade univariada Gaussiana inversa com risco base Log Normal, onde considerou como significativas, as covariáveis estação de nascimento, sexo, peso aos 365 dias e a interação peso aos 365 dias com sexo. Houve uma predominância do nascimento na estação seca (julho a dezembro). Foi observado que os bovinos que mais permaneceram no rebanho nasceram na estação seca e eram machos, justificado pela ausência de parasitas, disponibilização de pastagens para a mãe no período de prenhez, avaliação de carcaça, venda, abate e reprodução. O uso de modelos de fragilidade mostrou-se eficiente para atender aos objetivos propostos, ganhando relevância o potencial de seu uso para contemplar a heterogeneidade não-observada, sendo uma ótima ferramenta para o melhoramento animal.

Palavras-chave: Sobrevivência, fragilidade univariada, censura, falha.

Introdução

A permanência do animal no rebanho, está ligada a aspectos de produção, reprodução, nutrição e econômicos. A permanência da matriz no rebanho é influenciada por fatores como as características de reprodução animal, tendo em vista que desempenho insuficiente do animal quanto a quesitos reprodutivos

pode determinar a redução do tempo de permanência. As características produtivas são de fundamental importância para a permanência funcional da matriz no rebanho. 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).

A Análise de Sobrevivência, por sua vez, é considerada metodologia estatística mais adequada para lidar com dados de tempo até a ocorrência de um evento de interesse (tempo de falha), na presença de censura (Colosimo e Giolo, 2006) que é sua principal característica. Na produção animal, podemos destacar Bonetti et al. (2009) que estimaram parâmetros genéticos numa avaliação genética para a longevidade de touros italianos da raça Pardo-Suíça, por meio do modelo de riscos proporcionais de Weibull. Os autores consideraram o método satisfatório para o uso e a inclusão dos touros nos programas de melhoramento genético. No entanto, Caetano et al. (2013) propuseram a idade da vaca ao último parto (IVUP) como medida para avaliar a habilidade de permanência da vaca no rebanho. Os autores concluíram que a variável é relevante para avaliar a habilidade de permanência das vacas no rebanho e que o modelo de análise de sobrevivência estimou maior proporção da variabilidade genética para a característica estudada. Uma característica especial associada aos dados de sobrevivência é a possibilidade de que, para alguns indivíduos, o tempo completo até a ocorrência do evento de interesse não seja observado, devido a diversas causas. A não consideração destes indivíduos com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes (Bolfarine, 1991). Portanto, a importância da introdução de uma variável na análise que indique se o tempo de sobrevivência foi observado (Louzada-Neto e Pereira, 2000). Essa variável é definida na literatura como variável indicadora de censura ou simplesmente “censura”.

Em estudos recentes há situações em que a variável resposta, tempo de falha, pode estar sendo influenciado por fatores não observáveis, denominado fatores latentes. Modelos de sobrevivência com variáveis latentes ou modelos de fragilidade, são caracterizados pela inclusão de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou que não foram observadas; como fatores ambientais, genéticos ou informações que por algum motivo não foram considerados no planejamento. Uma das formas encontradas para incorporar esse efeito aleatório, denominado de variável de fragilidade, é introduzi-lo na modelagem da função de risco, com o objetivo de

controlar a heterogeneidade não observável das unidades em estudo (Tomazella, 2003). A fragilidade pode ser inserida no modelo de forma aditiva ou multiplicativa, com o objetivo de avaliar a heterogeneidade entre as unidades na função de risco ou a dependência para dados multivariados. Em estudos com animais, as associações aparecem devido a influências genéticas ou ambientais compartilhadas e se forem ignoradas, inferências incorretas poderão ser obtidas.

O termo de fragilidade foi introduzido por Vaupel et al. (1979) em modelos de sobrevivência com dados univariados. Devido as características da fragilidade no modelo de fragilidade multiplicativo, as candidatas naturais à distribuição da fragilidade, supostamente contínuas e não dependentes do tempo, são as distribuições gama, log-normal, Gaussiana inversa e Weibull. Hougaard (1984) foi um dos primeiros autores a abordar o impacto do uso de diferentes distribuições para a variável de fragilidade.

O Curraleiro Pé-Duro foi a primeira raça bovina selecionada no Brasil a partir de raças portuguesas trazidas por colonizadores a partir do século XV (Salles et al., 2011). Foi introduzida na região do estado do Piauí, a partir do Rio São Francisco por Domingos Afonso Mafrense, em meados do ano 1674 (Nogueira Neto, 1980), resultando posteriormente em adaptação dos bovinos em condições ambientais da região (Carvalho, 1984). É indicado como raça pura para produção de sêmen e embriões para uso em reprodução e cruzamentos industriais com raças especializadas para produção de leite e carne macia, comercializados com denominação de origem protegida. A resistência natural a ecto e endoparasitas e adaptação às nossas gramíneas e leguminosas são as grandes armas desses bovinos, que foram selecionados naturalmente por séculos para enfrentarem as adversidades locais. A tudo isso vem somar a grande amplitude térmica em que podem ser criados e grande longevidade, vivendo por mais de 20 anos. Entretanto o grande mérito dessa raça é converter alimentos de baixa qualidade em alimentos nobres e possibilitar a convivência do homem em regiões semiáridas.

Neste contexto, o objetivo com este trabalho foi analisar os efeitos de fatores não-observados, em estudo sobre o tempo de permanência no rebanho, de bovinos da raça Curraleiro Pé-Duro utilizando o modelo de fragilidade univariada Gaussiana inversa.

Material e Métodos

Os dados deste estudo, foram cedidos pelo núcleo de conservação de bovinos da raça Curraleiro Pé-Duro (CPD) pertencentes a Embrapa Meio-Norte, em Teresina-Piauí-Brasil, com campo experimental situado na fazenda Otávio Domingues, em São João do Piauí (entre 8° 26' e 8° 54' latitude Sul e entre 42° 19' e 42° 45' longitude Oeste), no semiárido piauiense pertencente a um rebanho de conservação in situ.

O manejo desses bovinos, foi feita de forma extensiva com o fornecimento apenas de sal, mineral e água, o que justifica seu baixo peso se comparado a bovinos da raça CPD criado em outras regiões e de bovinos de outras raças. Vale destacar, a falta de suplementação alimentar, o aumento da incidência de efeitos ambientais com a presença de muitas plantas tóxicas, carrapatos, babesia, verminoses, que contribuem para o baixo peso do CPD.

Foram avaliados 102 bovinos (58 Machos e 44 fêmeas) da raça CPD desde o nascimento até aos 550 dias. Os dados dos animais foram coletados entre o período de 2005 a 2014. Para modelagem do tempo de sobrevivência (em meses), após início de vida reprodutiva dos bovinos até a ocorrência da falha (inatividade ocasionada pela morte ou venda), em relação ao bovino permanecer ativo (vivo), considerou-se o tempo como variável resposta, e a data de nascimento do bezerro como o início do estudo. A variável T (tempo) foi obtida a partir da diferença entre a data de nascimento e a data do descarte. A data do último descarte no rebanho foi considerada como período final de observação para animais que ainda não haviam sido descartado (19 de agosto de 2016). A falha foi definida como inatividade do bovino (morte ou venda), enquanto a censura foi definida para bovinos que permaneceram vivos no rebanho. O tipo de censura utilizado foi a direita. Foi registrado, para cada animal observado, um correspondente indicador de censura, denominado status ($\delta = 1$ se falha e $\delta = 0$ se censura) indicando se o animal está ativo ou inativo no rebanho.

Nesse estudo, as covariáveis consideradas como possíveis fatores de riscos no tempo de permanência dos bovinos CPD no rebanho, foram: EN: estação de nascimento, S: sexo, PN: peso ao nascer, PD: peso a desmama, P365: peso aos 365 dias, P550: peso aos 550 dias categorizados de acordo com o peso médio em cada fase (Carvalho et al., 2015) e descritos de acordo com a Tabela 1.

Tabela 1. Número e porcentagem de bovinos que falharam ou foram censurados por variável

Variável	Categoria	n	Falhas (%)	Censuras (%)
Estação de nascimento	0 - Chuvosa	35	31(88,6)	4(11,4)
	1 - Seca	67	42(62,7)	25(37,3)
Sexo	0 - Machos	58	48(82,7)	10(17,3)
	1 - Fêmeas	44	25(56,8)	19(43,2)
Peso ao Nascer	0 - < 20 kg	56	41(73,2)	15(26,8)
	1 - ≥ 20 kg	46	32(69,6)	14(30,4)
Peso a desmama	0 - < 66 kg	40	29(72,5)	11(27,5)
	1 - ≥ 66 kg	62	44(71,0)	18(29,0)
Peso aos 365 dias	0 - < 95 kg	34	29(85,3)	5(14,7)
	1 - ≥ 95 kg	68	44(64,7)	24(35,3)
Peso aos 550 dias	0 - < 131 kg	48	35(72,9)	13(27,1)
	1 - ≥ 131 kg	54	38(70,4)	16(29,6)

Formulação do modelo

Para análise da longevidade dos bovinos no rebanho, as técnicas de análise de sobrevivência foram utilizadas. Devido à presença de censuras em dados de sobrevivência, os mesmos são resumidos com estimativas da função de sobrevivência e da função de risco (Moore, 2016). Para a estimativa dessas funções, foi utilizado o método não paramétrico de Kaplan e Meier (Kaplan e Meier, 1958). Para analisar a influência das covariáveis na longevidade dos bovinos no rebanho foi utilizado o modelo de fragilidade univariada, que é uma extensão do modelo de Cox. Todas as análises estatísticas foram realizadas com o uso do software estatístico livre (R, versão 3.3.1., Team, 2013) e o pacote parfm (Journal of Statistical Software, 2012).

Modelo de fragilidade univariada

Os modelos de fragilidade para dados de sobrevivência univariada, leva em conta que a população é não homogênea. A heterogeneidade pode ser explicada por covariáveis, mas quando importantes covariáveis não são incorporadas no modelo, isso conduz a heterogeneidade não observada (Colosimo e Giolo, 2006). O modelo de fragilidade multiplicativo é uma extensão do modelo de Cox (Cox, 1972), onde

o risco individual depende de uma variável aleatória Z não observável, não negativa, a qual age de forma multiplicativa na função de risco de base. A função de risco com a presença de covariáveis no instante t para o i -ésimo indivíduo é dada por:

$$h(t|X, z_i) = z_i h_0(t) \exp\{\mathbf{X}'\boldsymbol{\beta}\}$$

em que \mathbf{X} é o vetor de covariáveis e $\boldsymbol{\beta}$ o vetor de parâmetros associados a \mathbf{X} . Como z_i representa um valor da variável aleatória não observável, o risco individual cresce quando $z_i > 1$, decresce se $z_i < 1$ e para $z_i = 1$ o modelo de fragilidade (1) reduz ao modelo de risco proporcional de Cox (Cox, 1972). O fato da variável de fragilidade atuar de forma multiplicativa na função de risco, implica quanto maior for o valor da variável de fragilidade, maior será a chance de ocorrer a falha. Dessa forma, quanto maior for z_i , mais “frágeis” as observações pertencentes ao indivíduo i estão para falhar, daí o nome de fragilidade. Portanto, é esperado que o evento de interesse ocorra para os indivíduos mais “frágeis” (Tomazella, 2003).

Um problema importante em modelos de fragilidade é na escolha da distribuição para o efeito aleatório. Devido à forma como o termo de fragilidade atua na função de risco, as candidatas à distribuição de fragilidade são supostamente não negativas, usualmente contínuas e não dependentes do tempo como, por exemplo, as distribuições gama, log-normal, Gaussiana inversa e weibull (Hougaard, 1995). Já para a função risco base podem ser atribuídas distribuições tradicionalmente utilizadas para representar tempos de vida, como a exponencial, log-normal, weibull, gama, dentre outras (Wienke, 2010).

Distribuição de fragilidade Gaussiana inversa

A distribuição Gaussiana inversa foi introduzida por Hougaard (1984). Assim, seja Z uma variável aleatória que segue uma distribuição de fragilidade Gaussiana inversa com $E(Z) = 1$ e $\text{Var}(Z) = \theta$. A função densidade de probabilidade é dado por

$$f(z) = \frac{1}{\sqrt{2\pi\theta}} u^{-3/2} \exp\left(-\frac{(z-1)^2}{2\theta z}\right), \theta > 0$$

Uma ferramenta útil para análise de modelo de fragilidade é a transformada de Laplace. Dada uma função $g(x)$, a transformada de Laplace considerada com uma função real com argumento s é definida por (Wienke, 2010)

$$L(s) = \int_0^{\infty} g(s)e^{-sx} ds$$

A razão para isso ser útil em nosso contexto é que a transformada de Laplace tem exatamente a mesma forma que a função de sobrevivência não condicional. A função de sobrevivência não condicional, integrando o termo de fragilidade, é dada por:

$$S(t) = \int_0^{\infty} [S_0(t)]^z g(z) dz = \int_0^{\infty} e^{-H_0(t)z} g(z) dz = L[H_0(t)]$$

em que $g(z)$ é a função densidade de probabilidade da variável de fragilidade e $L[H_0(t)]$ denota a transformação de Laplace da função $g(z)$ considerando a função de risco acumulada, $H_0(t)$.

Consequentemente, a transformada de Laplace da distribuição Gaussiana inversa é dada por

$$L(z) = \exp\left[\frac{1}{\theta}(1 - \sqrt{1 + 2\theta z})\right], z \geq 0$$

A função de risco e função de sobrevivência não condicional da variável de fragilidade Gaussiana inversa é dada, respectivamente, por

$$h(t) = \frac{h_0(t)}{(1 + 2\sigma^2 H_0(t))^{1/2}} \quad \text{e} \quad S(t) = \exp\left[\frac{1}{\sigma^2}(1 - \sqrt{1 + 2\sigma^2 H_0(t)})\right]$$

onde h_0 e H_0 são as funções de risco base e risco acumulado base.

Na presença de covariáveis, as funções de risco e sobrevivência não-condicional, são dadas respectivamente, por:

$$h(t, X) = \frac{h_0(t)e^{\beta'X}}{(1 + 2\sigma^2 H_0(t)e^{\beta'X})^{1/2}} \quad (1)$$

e

$$S(t, X) = \exp\left[\frac{1}{\sigma^2}(1 - \sqrt{1 + 2\sigma^2 H_0(t)e^{\beta'X}})\right] \quad (2)$$

em que X é o vetor de covariáveis e β o vetor de parâmetros associados a X .

Modelo de Fragilidade Gaussiana Inversa-Log Normal

Diferentes formas paramétricas podem ser assumidas para a função de risco base $h_0(t)$, por exemplo: Log Normal, Log-logístico, Gompertz e etc. A Tabela 2, mostra as funções de densidade de probabilidade e funções de risco e sobrevivência base para estas distribuições.

Tabela 2. Funções de densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_0(t)$ das distribuições Log Normal, Log-Logístico e Gompertz.

Distribuição	$f(t)$	$h_0(t)$	$S_0(t)$
Log Normal	$\frac{1}{\sqrt{2\pi} t\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right]$	$\frac{1}{\sqrt{2\pi} t\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right] \Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)$	$\Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)$
Log-logístico	$\frac{\gamma}{\lambda^\gamma} t^{\gamma-1} \left(1 + \left(\frac{t}{\lambda}\right)^\gamma\right)^{-2}$	$\frac{\gamma \left(\frac{t}{\lambda}\right)^{\gamma-1}}{\lambda \left[1 + \left(\frac{t}{\lambda}\right)^\gamma\right]}$	$\frac{1}{\left[1 + \left(\frac{t}{\lambda}\right)^\gamma\right]}$
Gompertz	$\lambda \exp(\gamma t) \exp\left\{-\left(\frac{\lambda}{\gamma}\right)(e^{\gamma t} - 1)\right\}$	$\lambda \exp(\gamma t)$	$\exp\left\{-\left(\frac{\lambda}{\gamma}\right)(e^{\gamma t} - 1)\right\}$

Neste estudo, foi considerada a abordagem paramétrica para o modelo de fragilidade univariada Gaussiana inversa, onde os tempos de vida dos bovinos em risco segue distribuição Log Normal. Assim, substituindo as funções de risco base $h_0(t)$, sobrevivência base $S_0(t)$, e risco base acumulada $H_0 = -\log(S_0(t))$, da distribuição Log Normal, respectivamente em (1) e (2), com a presença de covariáveis, o modelo de fragilidade univariada Gaussiana inversa- Log Normal com função de sobrevivência e risco não condicionada dada, respectivamente, por:

$$S(t, X) = \exp\left[\frac{1}{\sigma^2} \left(1 - \sqrt{1 - 2\sigma^2 \log\left\{\Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)\right\}} e^{\beta'X}\right)\right] \quad (3)$$

e

$$h(t, X) = \left[\frac{\frac{1}{\sqrt{2\pi} t\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right]}{\Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)} e^{\beta'X}\right] \cdot \left[\frac{1}{\left(1 - 2\sigma^2 \log\left(\Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)\right)\right)^{1/2}} e^{\beta'X}\right] \quad (4)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão.

Estimação dos Parâmetros do Modelo

Na análise de sobrevivência, o método mais amplamente utilizado para estimar o vetor de parâmetros $\tau = (\theta, \mu, \sigma, \beta)$ do modelo (4) é o método de máxima verossimilhança, uma vez que pode incorporar dados censurados. Supondo que os dados sejam independentes e identicamente distribuídos, a função de verossimilhança não-condicional, com dados censurados, é dada por:

$$L(\tau) = \prod_{i=1}^n [h_i(t)]^{\delta_i} S_i(t) \quad (5)$$

onde δ_i é o indicador de censura, h_i e S_i é a função risco e sobrevivência da distribuição de fragilidade. Substituindo (1) e (2) em (5), temos a função de verossimilhança não-condicional da fragilidade Gaussiana inversa dada por:

$$L(\tau) = \prod_{i=1}^n \left[\frac{h_0(t) e^{\beta' X}}{(1 + 2\sigma^2 H_0(t) e^{\beta' X})^{1/2}} \right]^{\delta_i} \exp \left[\frac{1}{\sigma^2} (1 - \sqrt{1 + 2\sigma^2 H_0(t) e^{\beta' X}}) \right] \quad (6)$$

Consequentemente, substituindo (3) e (4) em (6), temos para o modelo de fragilidade Gaussiana inversa com risco base Log Normal com parâmetros τ , a função de verossimilhança não-condicional dada por

$$L(\tau) = \prod_{i=1}^n \left[\frac{\frac{1}{\sqrt{2\pi} t \sigma} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right]}{\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)} e^{\beta' X} \right]^{\delta_i} \left[\frac{1}{\left(1 - 2\theta \log \left(\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right) \right) e^{\beta' X} \right)^{1/2}} \right]^{\delta_i} \times \exp \left[\frac{1}{\sigma^2} \left(1 - \sqrt{1 - 2\theta \log \left(\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right) \right) e^{\beta' X}} \right) \right] \quad (7)$$

As estimativas de máxima verossimilhança são obtidas através da maximização numérica da função log-verossimilhança descrita em (7). Para estimação dos parâmetros, foi utilizado o pacote *parfm* (Journal of Statistical Software, 2012) do *software* R. A construção dos intervalos de confiança para os parâmetros do modelo são baseados nas propriedades de normalidade assintótica dos estimadores de máxima verossimilhança. Se $\hat{\tau}$ denota os estimadores de máxima verossimilhança de τ então a distribuição de $\hat{\tau} - \tau$ é aproximada por uma distribuição normal q-variada com média zero e matriz de covariância $I^{-1}(\hat{\tau})$ onde

$I(\tau)$ é a matriz de informação observada. Assim, um intervalo de confiança assintótico com $100(1 - \alpha)\%$ para cada parâmetro é:

$$\hat{\tau} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\tau)}$$

onde $\hat{\tau}$ é o elemento da diagonal principal de $I^{-1}(\hat{\tau})$ correspondente a cada parâmetro e $z_{\alpha/2}$ é o quantile $(1 - \alpha)\%$ da distribuição normal padrão.

Para testar hipóteses relacionadas aos parâmetros $(\theta, \mu, \sigma, \beta)$ foram utilizados três testes: o de Wald, o da Razão de Verossimilhança (TRV) e o Escore (Colosimo e Giolo, 2006). Critérios de seleção de modelos como o critério de informação de Akaike (AIC) proposto por Akaike (1974), e o critério de informação bayesiano (BIC), proposto por Schwarz et al. (1978), são frequentemente utilizados para selecionar modelos em diversas áreas. Os melhores modelos são considerados aqueles que apresentam menores valores de AIC e BIC. Para a adequação do modelo, diversos métodos encontram-se disponíveis na literatura e baseiam-se, essencialmente, nos resíduos de Cox-Snell, que auxiliam a examinar o ajuste global do modelo, o de Schoenfeld que tem coeficiente dependente do tempo, o de martingal, que é dado pela diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, e de deviance, que facilitam na detecção de pontos atípicos (outliers).

Resultados

Nesse estudo, o rebanho é formado por 102 bovinos, onde 57% é formado por machos e 43% por fêmeas, da raça CPD nascidos no período de 2005 a 2014. Ao considerar a sobrevivência desses animais no rebanho, constatou-se que 28% (29) foram censurados, ou seja, permaneceram vivos até o período final da observação, e 72% (73) falharam, isto é, morreram ou foram vendidos. Foi observado para a EN, uma maior predominância (65%) do período da seca, onde tem uma menor incidência de chuvas. Em relação aos pesos padrões (nascer, desmama, ano e sobreano), foi encontrado uma frequência superior a 50% com relação ao peso médio para o PD, P365 e P550.

Na tabela 1 apresentamos o número e porcentagem de bovinos que falharam ou foram censurados por variável. Com relação a covariável EN, observou-se uma maior porcentagem de falhas (88,6%) nos bezerros nascidos na estação chuvosa. Já em relação ao S, os bezerros machos falharam mais (82,7%)

comparado as fêmeas (56,8%). Em relação as variáveis PN, PD, P365 e P550 todos tiveram uma maior porcentagem de falhas, para pesos abaixo da média.

A estimativas para os parâmetros do modelo de fragilidade univariada Gaussiana inversa- Log Normal, estão descritos na Tabela 3, onde foram apresentadas apenas as estimativas dos parâmetros das covariáveis que foram significativas ao se considerar p -valor menor que 5%. Dentre as variáveis analisadas (EN, S, PN, PD, P365 e P550) foi observado significância nas variáveis EN, S, P365 e a interação do S com P365 (p – valor < 0,05). A taxa de falha na EN seca foi de $\exp(-1,310) = 0,269$ vezes se comparado com a chuvosa, indicando que há diminuição do risco dos bovinos falhar (morte ou venda), dentre as que conseguem se manterem viva demonstrando uma adaptabilidade do bovino no rebanho. Para a variável S, a taxa de falha para as fêmeas foi $\exp(1,677) = 5,349$ vezes a taxa de falhas para os machos, ou seja, bezerros fêmeas tem 5,349 vezes mais chances de falhar (morte ou venda) do que os bezerros machos. Para a variável P365, a taxa de falha foi $\exp(0,775) = 2,170$ vezes para animais que pesam igual ou maior a 95Kg. Para a interação da variável S com o P365 foi observado uma taxa de falha de $\exp(-3,204) = 0,040$ indicando uma diminuição do risco dos bovinos virem a falhar, ou seja, morrer ou sair para venda. Quanto a variância estimada para a fragilidade θ é igual a 0,155, revelando a presença de heterogeneidade não observada, como por exemplo, fatores genéticos ou ambientais.

Tabela 3. Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), p -valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de fragilidade univariada Gaussiana inversa- Log Normal

Parâmetros	EMV	EP	CI (95%)	p -valor	RTF
θ	0,155	0,593	-	-	-
μ	3,924	0,187	-	-	-
σ	0,544	0,134	-	-	-
β_1 (EN)	-1,310	0,319	[0,144; 0,504]	<0,001	0,269
β_2 (S)	1,677	0,527	[1,904;15,028]	0,001	5,349
β_3 (P365)	0,775	0,373	[1,044; 4,515]	0,038	2,170
β_4 (S*P365)	-3,204	0,738	[0,010; 0,173]	<0,001	0,040

As funções de sobrevivência (3) e função de risco (4) não condicionado do modelo de fragilidade Gaussiana inversa com risco base log-normal é dada, respectivamente, por:

$$S(t, X) = \exp \left[\frac{1}{0,155} \left(1 - \sqrt{1 - 0,31 \log \left\{ \Phi \left(\frac{-\log(t)+3,924}{0,544} \right) \right\} g(\beta' X_{ij})} \right) \right]$$

e

$$h(t, X) = \left[\frac{\frac{1}{0,155\sqrt{2\pi} t} \exp \left[-\frac{1}{2} \left(\frac{\log(t)-3,924}{0,544} \right)^2 \right]}{\Phi \left(\frac{-\log(t)+3,924}{0,544} \right)} g(\beta' X_{ij}) \right] \cdot \left[\frac{1}{\left(1 - 0,31 \log \left(\Phi \left(\frac{-\log(t)+3,924}{0,544} \right) \right) g(\beta' X_{ij}) \right)^{1/2}} \right]$$

onde $g(\beta' X_{ij}) = \exp\{-1,310 \times EN + 1,677 \times S + 0,775 \times P365 - 3,204 \times (S * P365)\}$.

De acordo com a (Figura 1), foi possível observar que 50% dos bezerros tiveram fragilidade variando entre [0,682; 0,985] e 50% variando entre [1; 1,111], indicando a presença de uma heterogeneidade não-observada relativamente alta em metade do rebanho.

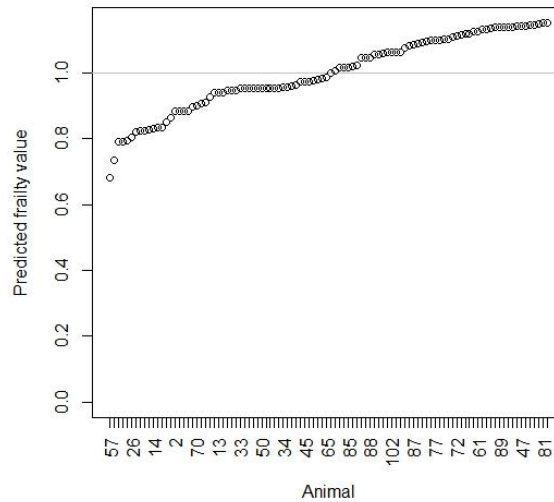


Figura 1. Fragilidade Gaussiana inversa univariada com risco base Log Normal

Para avaliar a qualidade do ajuste do modelo, na Figura 2, estão os gráficos dos resíduos de Cox-Snell, resíduos Martingal e resíduos Deviance. Nos painéis (a) e (b), observa-se que os resíduos de Cox-

Snell seguem aproximadamente uma distribuição exponencial padrão, o que indica uma qualidade de ajuste global aceitável do modelo. Os painéis (c) e (d) não sugerem a existência de pontos de *outliers*.

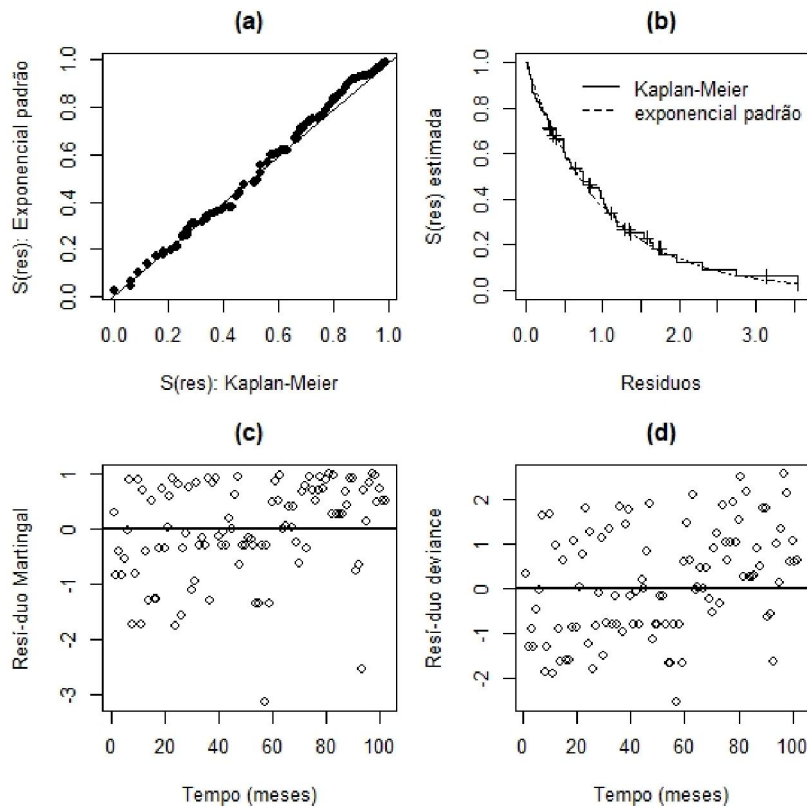


Figura 2 Resíduos de Cox-Snell , Martingal e Deviance do modelo de modelo de Cox- Log Normal para dados de bovinos da raça Curraleiro Pé-Duro.

Foi feito um estudo comparativo do modelo de fragilidade Gaussiana inversa usando as funções de risco base log-normal e log-logístico, descritos na Tabela 4, onde foi utilizado o critério de Akaike (AIC) e critério de informação bayesiano (BIC), para seleção de modelos, mostrando que o modelo de fragilidade Gaussiana inversa- Log Normal é o mais adequado por apresentar o menor valor de AIC e BIC.

Tabela 4. Estimativas para os parâmetros dos modelos de fragilidade Gaussiana inversa- Log Normal e Gaussiana inversa-log-logístico

Parâmetros	Log Normal		Log-Logístico	
	Estimativas	Erro Padrão	Estimativas	Erro Padrão
θ	0,155	0,593	0,014	0,322
μ	3,924	0,187	-	-
σ	0,544	0,134	-	-
α	-	-	-11,617	2,082
γ	-	-	2,962	0,581
$\beta_1(\text{EN})$	-1,310	0,319	-1,264	0,289
$\beta_2(\text{S})$	1,677	0,527	1,513	0,493
$\beta_3(\text{P365})$	0,775	0,373	0,695	0,355
$\beta_4(\text{S*P365})$	-3,204	0,738	-2,978	0,687
Log-Veross		-372,94		-374,79
AIC		759,88		763,59
BIC		778,25		781,97

Discussão

A adaptabilidade, ou capacidade de se adaptar, pode ser avaliada pela habilidade do animal em se ajustar às condições ambientais médias, assim como aos extremos climáticos. Animais bem adaptados caracterizam-se pela manutenção ou mínima redução no desempenho produtivo, pela elevada eficiência reprodutiva, resistência às doenças, longevidade e baixa taxa de mortalidade durante a exposição ao estresse (Baccari Júnior, 1990).

O pequeno porte do gado pé-duro é o principal motivo alegado para quase levá-lo à extinção. Entretanto, raramente são mencionadas as extremas condições a que os animais dessa raça são submetidos. Raças melhoradas, adaptadas a boas pastagens, se mantidas nas mesmas condições oferecidas ao gado pé-duro, em pouco tempo terão seu peso reduzido e o desempenho reprodutivo negativamente afetado.

Ademais, é importante acrescentar que a capacidade de produção de uma raça não depende apenas do peso individual dos animais, sendo fundamentais os índices de natalidade e de mortalidade. Um rebanho de animais mais pesados, porém menos prolíficos e com maior percentagem de mortes, resulta em menor produção de carne. Além disso, uma mesma pastagem pode garantir a alimentação de um maior número de animais de porte mais reduzido. Portanto, um índice mais adequado para avaliar o desempenho de uma raça bovina em pastagem é sua produtividade anual por unidade de área (Carvalho, 2002).

A perspectiva abordada nessa pesquisa teve como suporte a longevidade dos bovinos e a saída do rebanho ocasionada pela morte ou venda. Houve uma predominância de 66% de bezerros nascidos na estação seca, indicando que mais da metade dos bezerros nasceram em um período favorável para o seu desenvolvimento, ocasionada pela diminuição de doenças no ambiente e mortalidade dos bezerros, além da oferta de nutrientes no pasto na época de prenhes para a mãe. Ou seja, o bezerro nascido na época da seca, a mãe durante a prenha teve mais pasto disponível, logo mais leite e conseqüentemente bezerro mais gordo. Em relação aos pesos padrões (nascer, desmama, ano e sobreano), foi encontrado uma maior frequência de bovinos com peso acima da média, indicando uma melhor concentração genética e uma boa adaptabilidade.

Foi utilizado o modelo de fragilidade univariada Gaussiana inversa com risco base Log Normal onde foram significativas (p -valor < 5%) as covariáveis: EN, S e P365 e a interação do S com P365, conforme Tabela 3. Ainda, foi observado através da razão da taxa de falha, que para a covariável EN, houve uma diminuição na taxa de falha (morte ou venda) para os bovinos nascidos na estação seca que pode ser explicado por ter menos doenças para os bezerros, além da oferta de nutrientes no pasto na época de prenhes para a mãe, demonstrando uma maior longevidade para os bovinos. Para a variável S, foi encontrado uma taxa de falha maior para as fêmeas do que os machos, que pode ser explicado pela maternidade, ou seja, como as fêmeas dão cria anualmente, na ausência de suplemento, as mesmas tiram dos ossos para colocar no leite durante a amamentação, tornando a vaca mais magra e mais sujeita a mortalidade. Para a variável P365, foi encontrado uma taxa de falha maior para os bovinos que tem um peso igual ou maior que 95 kg que pode ser explicado devido a bovinos mais pesados saírem do rebanho para venda, abate ou reprodução.

A metodologia utilizada nesse estudo mostrou-se promissora com resultados interessantes no estudo de avaliação de permanência dos animais no rebanho como pode ser visto na (Figura 1), onde foi detectado as fragilidades inerentes a fatores ambientais e genéticos, e na (Figura 2) onde foi realizado os testes de

ajustes de modelos para efeito de natureza genética como aleatório, nas quais constam a indicação de adequação do modelo de fragilidade para atender ao objetivos propostos pelos testes Cox-Snell, Martingal e Deviance. Nesse estudo na modelagem do tempo de permanência do bezerro no rebanho com o modelo de fragilidade univariada Gaussiana inversa- Log Normal, onde o efeito aleatório foi associado a cada animal, a variância estimada para a fragilidade, $\hat{\theta} = 0,155$, indicou a presença de heterogeneidade não observada, ocasionada, por exemplo, por fatores genéticos ou ambientais, que está relacionada a fatores como: cor da pelagem, cor da pelame, habilidade materna, zona de termoneutralidade, condição de criação, tipo de pasto, tipo de alimentação, dentre outros (Carvalho et.al, 2010).

Conclusão

A principal conclusão deste artigo, é que o uso de modelos de fragilidade mostrou-se ser um critério de seleção adequado às situações propostas. Também permitiu incorporar um termo para a heterogeneidade não-observada que afeta a estimativa do risco, onde foi possível observar para o modelo de fragilidade univariada, a significância estatística nas covariáveis EN, S, P365 e a interação do S com o P365 afetando na longevidade do bovino. Nesse sentido, o modelo de fragilidade univariada pode ser utilizado como ferramenta de ajuda no melhoramento animal.

Declaração de Conflito de Interesse

Os autores declaram não haverem conflitos de interesse.

Financiamento

Esta pesquisa não recebeu nenhuma bolsa específica de agências de fomento nos setores público, comercial ou sem fins lucrativos.

Referências

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* **19**, 716–723.

- Baccari Júnior, F. (1990). Métodos e técnicas de avaliação da adaptabilidade dos animais às condições tropicais. *Simpósio Internacional de Bioclimatologia Animal nos Trópicos: pequenos e grandes ruminantes* **1**, 9–17.
- Bolfarine, H., R. J. A. J. (1991). ‘Análise de Sobrevivência’. (2ª Escola de Modelos de Regressão, Rio de Janeiro)
- Bonetti, O., Rossoni, A., e Nicoletti, C. (2009). Genetic parameters estimation and genetic evaluation for longevity in italian brown swiss bulls. *Italian Journal of Animal Science*, **8**, 30–32.
- Caetano, S., Rosa, G., Savegnago, R., Ramos, S., Bezerra, L., Lôbo, R., De Paz, C., e Munari, D. (2013). Characterization of the variable cow’s age at last calving as a measurement of longevity by using the kaplan–meier estimator and the Cox model. *Animal* **7**, 540–546.
- Carvalho, G. M. C., Lima Neto, A., Da Frota, M. N. L., de Sousa, V. R., Carneiro, M. d. S., Vêras, G. M. C., Costa, R. N. M., e Azevedo, D. M. M. R. (2015). O uso de bovinos Curraleiro Pé-Duro em cruzamentos para produção de carne de boa qualidade no trópico quente-fase 1. In ‘Embrapa Meio Norte- Artigo em anais de congresso (ALICE)’. (Congresso nordestino de produção animal)
- Carvalho, J. (1984). Relatório de atividades do núcleo de preservação do gado pé-duro ou curraleiro. EMBRAPA/UEPAE. (Teresina)
- Carvalho, J. H. (2002). Potencial econômico do bovino pé-duro. Embrapa Meio-Norte-Documents (INFOTECA-E).
- Carvalho, G. M. C., Almeida, M. D. O., Azevêdo, D. M. M. R., de Araújo Neto, R. B., Leal, T. M., Monteiro, F. D. C., Lima Neto, A. F. (2010). Origem, formação e conservação do gado Pé-Duro, o bovino do Nordeste brasileiro. Embrapa Meio-Norte-Documents (INFOTECA-E).
- Colosimo, E. A. e Giolo, S. R. (2006). ‘Análise de sobrevivência aplicada’. (Editora Edgard Blucher, São Paulo)
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202.
- Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* **71**, 75–83.

- Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis* **1**, 255–273.
- Kaplan, E. L. e Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**, 457–481.
- Louzada-Neto, F. e Pereira, B. d. B. (2000). Modelos em análise de sobrevivência. *Caderno de saúde coletiva*, Rio Janeiro, 9–26.
- Moore, D. F. (2016). ‘Applied survival analysis using R’. (Springer)
- Nogueira Neto, A. (1980). Aspectos da pecuária piauiense. *Sociedade de Medicina Veterinária do Piauí*, Teresina.
- Salles, P., Medeiros, G., Costa, R. G., Ramos, C. T. C., Borburema, J., Rocha, M., Rocha, L., e Mathias, W. (2011). Programa de conservação e melhoramento de uma raça bovina brasileira: Curraleiro Pé-Duro. *AICA-Actas Iberoamericana de Conservacion animal, Córdoba* **1**, 453–6.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics* **6**, 461–464.
- Tomazella, V. L. D. (2003). ‘Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade’. PhD thesis, Universidade de São Paulo.
- Van Arendonk, J.A.M.(1986) Economic importance and possibilities for improvement of dairy cow herd life. In: ‘World congress of genetic applied to livestock production’. pp. 95-100. (Lincoln)
- Vaupel, J. W., Manton, K. G., e Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- Wienke, A. (2010). ‘Frailty models in survival analysis.’ (Chapman and Hall/CRC)

6 CAPÍTULO 3

MODELAGEM DE FRAGILIDADE GAUSSIANA INVERSA COMPARTILHADA
APLICADA AO ESTUDO DA LONGEVIDADE DE BOVINOS DA RAÇA CURRALEIRO
PÉ DURO*

* Segundo as normas da revista

Modelagem de fragilidade Gaussiana inversa compartilhada aplicada ao estudo da longevidade de bovinos da raça Curraleiro Pé-Duro

Resumo – O objetivo com este trabalho foi fazer uso da técnica de análise de sobrevivência, analisando os efeitos de fatores não observados, em estudo sobre o tempo de permanência no rebanho, de bovinos da raça Curraleiro Pé-Duro utilizando-se do modelo de fragilidade compartilhada Gaussiana inversa. Foram observados dados de registros de 102 bovinos nascidas no período de 2005 a 2014, em um rebanho experimental de conservação in situ da Embrapa Meio-Norte. Considerou-se a falha como sendo a inatividade do bovino ocasionada pela morte ou venda e a censura o bovino permanecer ativo no rebanho. Foi utilizada a metodologia dos modelos de fragilidade compartilhada Gaussiana inversa com risco base Log Normal, onde foi considerado como significativas, as covariáveis estação de nascimento, sexo e a interação do peso ao ano com sexo. Foi observado que os bovinos que mais permaneceram no rebanho nasceram na estação seca e eram machos explicado pela ausência de parasitas, disponibilização de pastagens para a mãe no período de prenhez, avaliação de carcaça, venda, abate e reprodução. O uso de modelos de fragilidade compartilhada mostrou-se eficiente para atender aos objetivos propostos, ganhando relevância o potencial de seu uso para contemplar heterogeneidade não-observada, sendo uma ótima ferramenta para o melhoramento animal.

Termos para indexação: Curraleiro Pé-Duro, sobrevivência, censura, falha.

Introdução

O Curraleiro Pé-Duro possui grande resistência natural a ecto e endoparasitas e adaptação às nossas gramíneas e leguminosas são as grandes armas desses bovinos, que foram selecionados naturalmente por séculos para enfrentarem as adversidades locais. A tudo isso vem somar a alta amplitude térmica em que podem ser criados e grande longevidade, vivendo por mais de 20 anos.

Entretanto o grande mérito dessa raça é converter alimentos de baixa qualidade em alimentos nobres e possibilitar a convivência do homem em regiões semiáridas.

A Análise de Sobrevivência (AS), por sua vez, é considerada metodologia estatística mais adequada para lidar com dados de tempo até a ocorrência de um evento de interesse (tempo de falha), na presença de censura (Colosimo e Giolo, 2006) que é sua principal característica. Na produção animal, podemos destacar Bonetti et al. (2009) estimaram parâmetros genéticos numa avaliação genética para a longevidade de touros Italianos da raça Pardo-Suiça, por meio do modelo de riscos proporcionais de Weibull. Os autores consideraram o método satisfatório para o uso e a inclusão dos touros nos programas de melhoramento genético. Caetano et al. (2013) propuseram a idade da vaca ao último parto (IVUP) como medida para avaliar a habilidade de permanência da vaca no rebanho. Os autores concluíram que a variável é relevante para avaliar a habilidade de permanência das vacas no rebanho e que o modelo de análise de sobrevivência foi o que estimou maior proporção da variabilidade genética para a característica estudada. Uma característica especial associada aos dados de sobrevivência é a possibilidade de que, para alguns indivíduos, o tempo completo até a ocorrência do evento de interesse não seja observado, devido a diversas causas. A não consideração destes indivíduos com informações incompletas sobre seus tempos de vida pode levar a inferências viciadas ou menos eficientes (Bolfarine, 1991). Portanto, a importância da introdução de uma variável na análise que indique se o tempo de sobrevivência foi observado (Louzada-Neto e Pereira, 2000). Essa variável é definida na literatura como variável indicadora de censura ou simplesmente “censura”.

Em estudos recentes há situações em que a variável resposta, tempo de falha, pode estar sendo influenciado por fatores não observáveis, denominado fatores latentes. Modelos de sobrevivência com variáveis latentes ou modelos de fragilidade, são caracterizados pela inclusão de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou que não foram observadas; como fatores ambientais, genéticos ou

informações que por algum motivo não foram considerado no planejamento. Uma das formas encontradas para incorporar esse efeito aleatório, denominado de variável de fragilidade, é introduzi-lo na modelagem da função de risco, com o objetivo de controlar a heterogeneidade não observável das unidades em estudo. A fragilidade pode ser inserida no modelo de forma aditiva ou multiplicativa, com o objetivo de avaliar a heterogeneidade entre as unidades na função de risco ou a dependência para dados multivariados. Em estudos com animais, as associações aparecem devido a influências genéticas ou ambientais compartilhadas e se forem ignoradas, inferências incorretas poderão ser obtidas.

Diversos autores estudaram a aplicação dos modelos de fragilidade multiplicativos, que representam uma generalização dos modelos de riscos proporcionais de Cox (Cox, 1972). Andersen et al. (2012) e Hougaard (1995) apresentaram uma revisão dos modelos de fragilidade multiplicativos numa perspectiva clássica, enquanto, Sinha e Dey (1997) elaboraram uma revisão completa desses modelos sob o ponto de vista Bayesiana. O termo de fragilidade foi introduzido por Vaupel et al. (1979) em modelos de sobrevivência com dados univariados. Clayton (1978) e Oakes (1982) trabalharam com modelos de sobrevivência multivariados. Devido as características da fragilidade no modelo de fragilidade multiplicativo, as candidatas naturais à distribuição da fragilidade, supostamente contínuas e não dependentes do tempo, são as distribuições gama, log-normal, Gaussiana inversa e Weibull. Hougaard (1984) foi um dos primeiros autores a abordar o impacto do uso de diferentes distribuições para a variável de fragilidade.

Neste contexto, o objetivo deste trabalho foi analisar os efeitos de fatores não-observados, em estudo sobre o tempo de permanência no rebanho, de bovinos da raça Curraleiro Pé-Duro utilizando o modelo de fragilidade compartilhada Gaussiana inversa.

Material e Métodos

Os dados deste estudo, foram cedidos pelo núcleo de conservação de bovinos da raça Curraleiro Pé-Duro pertencentes a Embrapa Meio-Norte, em Teresina-Piauí, com campo experimental situado na fazenda Octávio Domingues, em São João do Piauí (entre 8° 26' e 8° 54' latitude Sul e entre 42° 19' e 42° 45' longitude Oeste), no semiárido piauiense pertencente a um rebanho de conservação in situ. O manejo foi realizado de forma extensiva com o fornecimento apenas de sal, mineral e água, o que justifica seu baixo peso se comparado a bovinos da raça Curraleiro Pé-Duro criado em outras regiões e de bovinos de outras raças. Vale destacar, a falta de suplementação alimentar, o aumento da incidência de efeitos ambientais com a presença de muitas plantas tóxicas como o barbatimão (*Stryphnodendron coriaceum*) e a erva-de-rato (*Palicouria marcgravii*), carrapatos, verminoses, que contribuem para o baixo peso do Curraleiro Pé-Duro.

Foram avaliados 102 bovinos (58 Machos e 44 fêmeas) da raça desde o nascimento até aos 550 dias. Os dados dos animais foram coletados entre o período de 2005 a 2014.

Para modelagem do tempo de sobrevivência (em dias), após início de vida reprodutiva dos bovinos até a ocorrência da falha (inatividade ocasionada pela morte ou venda), em relação ao bovino permanecer ativo (vivo), considerou-se o tempo como variável resposta, e a data de nascimento do bezerro como o início do estudo. A variável T (tempo) foi obtida a partir da diferença entre a data de nascimento e a data do descarte. A data do último descarte no rebanho foi considerada como período final de observação para animais que ainda não haviam sido descartado (19/08/2016). A falha foi definida como inatividade do bovino (morte ou venda), enquanto a censura foi definida para bovinos que permaneceram vivos no rebanho. O tipo de censura utilizado foi a direita. Foi registrado, para cada animal observado, um correspondente indicador de censura, denominado status ($\delta = 1$ se falha e $\delta = 0$ se censura) indicando se o animal está ativo ou inativo no rebanho.

A variável estação de nascimento foi considerada como sendo chuvosa os meses que compreenderam de janeiro a junho e seca os meses de julho a dezembro. Nesse estudo, as covariáveis consideradas como possíveis fatores de riscos no tempo de permanência dos bovinos Curraleiro Pé-Duro no rebanho, foram: EN: estação de nascimento, S: sexo, PN: peso ao nascer, PD: peso a desmama, P365: peso aos 365 dias, P550: peso aos 550 dias categorizados de acordo com o peso médio em cada fase (Carvalho et al., 2015) e descritos de acordo com a Tabela 1.

Tabela 1 Número e porcentagem de bovinos por variável estudada

Variável	Categoria	N (%)
Estação de nascimento	0 - Chuvosa	35 (34,3)
	1 - Seca	67 (65,7)
Sexo	0 - Machos	58 (56,9)
	1 - Fêmeas	44 (43,1)
Peso ao Nascer	0 - < 20 kg	56 (54,9)
	1 - ≥ 20 kg	46 (45,1)
Peso a desmama	0 - < 66 kg	40 (39,2)
	1 - ≥ 66 kg	62 (60,8)
Peso aos 365 dias	0 - < 95 kg	34 (33,3)
	1 - ≥ 95 kg	68 (66,7)
Peso aos 550 dias	0 - < 131 kg	48 (47,1)
	1 - ≥ 131 kg	54 (52,9)

Para análise da longevidade dos bovinos no rebanho, as técnicas de análise de sobrevivência foram utilizadas. Devido à presença de censuras em dados de sobrevivência, os mesmos são resumidos com estimativas da função de sobrevivência e da função de risco (Moore, 2016) onde para estimativa dessas funções, foi utilizado o método não paramétrico de Kaplan e Meier (Kaplan e Meier, 1958). Para analisar a influência das covariáveis na longevidade dos bovinos no rebanho, foi utilizado o modelo de fragilidade compartilhada, onde foi considerado o ano de nascimento de

cada bezerro como grupo, por apresentarem uma associação decorrente de fatores ambientais. Todas as análises estatísticas foram realizadas com o uso do *software* estatístico livre *R*, versão 3.3.1. (*Team*, 2013) e o pacote *parfm* (2017).

O modelo de fragilidade compartilhada foi sugerido por Clayton (1978) para a análise da correlação entre os dados de sobrevivência agrupados. Em um modelo de fragilidade compartilhada, a fragilidade é definida como uma medida do risco relativo que os indivíduos em um grupo compartilham. Assim, a variável fragilidade está associada a grupos de indivíduos ao invés de indivíduos. Neste modelo, bovinos do mesmo grupo compartilham covariáveis comuns, como por exemplo, se considerarmos grupos formados por ano de nascimento, teremos fatores ambientais influenciando de maneira diferente para cada período; se grupos forem formados considerando cada matriz, teremos influência dos fatores genéticos, dentre outros. Para o modelo de fragilidade compartilhada, assume-se que os tempos de sobrevivência são condicionalmente independentes para uma dada fragilidade compartilhada, isto é, a dependência entre os tempos de sobrevivência é apenas devido a covariáveis não observáveis ou fragilidade.

Suponha que existam n grupos e cada grupo i tem n_i observações associadas à fragilidade não observada Z_i ($1 \leq i \leq n$). O vetor \mathbf{X}_{ijk} ($1 \leq i \leq n; 1 \leq j \leq n_i$) contém as informações de k -ésimas covariáveis dos tempos de falha T_{ij} da j -ésima observação e i -ésimo grupo. Condicional ao termo de fragilidade Z_i os tempos de sobrevivência no grupo i ($1 \leq i \leq n$) são assumidos independentes com função de intensidade da forma dada por:

$$h(t_{ij} | \mathbf{X}_{ijk}, Z_i) = Z_i h_0(t_{ij}) \exp\{\boldsymbol{\beta}' \mathbf{X}_{ijk}\} \quad (1)$$

onde $h_0(t)$ representa a função de risco base, $\boldsymbol{\beta}$ é o coeficiente associado a k -ésima covariável e Z_i são variáveis aleatórias independentes e identicamente distribuídas com função densidade $f(z)$ que depende de parâmetros desconhecidos a serem estimados. A função de risco acumulado condicional é dada por

$$H(t_{ij} | \mathbf{X}_{ijk}, Z_i) = Z_i H_0(t_{ij}) \exp(\boldsymbol{\beta}' \mathbf{X}_{ijk}) \quad (2)$$

onde $H_0(t)$ representa a função de risco base acumulado. A função de sobrevivência condicionada ao efeito da fragilidade Z_i e ao efeito das covariáveis é definida por

$$S(t_{ij} | \mathbf{X}_{ijk}, Z_i) = (S_0(t))^{Z_i \exp(\boldsymbol{\beta}' \mathbf{x}_{ijk})} \quad (3)$$

em que $S_0(t)$ representa a função de sobrevivência base.

Assim, para valores de $Z_i > 1$, os bovinos dentro dos grupos correspondentes a esses valores tendem a apresentar sobrevivência mais curta. Por outro lado, para valores de $Z_i < 1$, tem-se que os bovinos destes grupos tendem a apresentar sobrevivência mais longa. A principal suposição de um modelo de fragilidade compartilhada é que todos os indivíduos no grupo i compartilham o mesmo valor da fragilidade Z_i ($i = 1, \dots, n$). Os tempos de vida são considerados condicionalmente independentes com relação a fragilidade compartilhada, sendo ela a causa da dependência entre os tempos de vida dentro dos grupos. Desta forma, há dependência dos tempos de falha dentro dos grupos e independência desses tempos entre os grupos.

Considerando para a variável de fragilidade Z uma distribuição gaussiana inversa, com $E(Z) = 1$ e $\text{Var}(Z) = \theta$, a função densidade de probabilidade da variável de fragilidade Z é dada por:

$$f(z_i, \theta) = \frac{1}{\sqrt{2\pi\theta}} z_i^{-3/2} \exp\left(-\frac{(z_i-1)^2}{2\theta z_i}\right) \quad z_i, \theta > 0 \quad (4)$$

denominado de modelo de fragilidade compartilhada Gaussiana inversa. A variância da variável fragilidade, quantifica a heterogeneidade presente. O modelo de fragilidade compartilhada pode assumir a forma paramétrica ou semi paramétrica. Diferentes formas paramétricas podem ser assumidas para a função de risco base $h_0(t)$, por exemplo: Log Normal, Log-logístico, Gompertz e etc. A Tabela 2, mostra as funções de densidade de probabilidade e funções de risco e sobrevivência base para estas distribuições.

Tabela 2

Funções densidade $f(t)$, sobrevivência $S_0(t)$ e risco $h_0(t)$ das distribuições Log Normal, Log-logistic Gompertz.

Distribuição	$f(t)$	$h_0(t)$	$S_0(t)$
Log Normal	$\frac{1}{\sqrt{2\pi} t\theta} \exp\left[-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right]$	$\frac{1}{\sqrt{2\pi} t\theta} \exp\left[-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right] \Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)$	$\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)$
Log-logistic	$\frac{\gamma}{\lambda^\gamma} t^{\gamma-1} \left(1 + \left(\frac{t}{\lambda}\right)^\gamma\right)^{-2}$	$\frac{\gamma \left(\frac{t}{\lambda}\right)^{\gamma-1}}{\lambda \left[1 + \left(\frac{t}{\lambda}\right)^\gamma\right]}$	$\frac{1}{\left[1 + \left(\frac{t}{\lambda}\right)^\gamma\right]}$
Gompertz	$\lambda \exp(\gamma t) \exp\left\{-\left(\frac{\lambda}{\gamma}\right) (e^{\gamma t} - 1)\right\}$	$\lambda \exp(\gamma t)$	$\exp\left\{-\left(\frac{\lambda}{\gamma}\right) (e^{\gamma t} - 1)\right\}$

Assim, substituindo as funções de risco base $h_0(t)$, sobrevivência base $S_0(t)$, e risco base acumulada $H_0 = -\log(S_0(t))$, da distribuição Log Normal, respectivamente em (1) e (2), com a presença de covariáveis, temos o modelo de fragilidade compartilhada Gaussiana inversa- Log Normal, cuja função de risco e função de sobrevivência condicional ao efeito ao efeito da fragilidade Z_i e k -ésimas covariáveis dos tempos de falha T_{ij} da j -ésima observação e i -ésimo grupo dadas, respectivamente, por

$$h(t_{ij} | \mathbf{X}_{ijk}, Z_i) = \frac{\frac{Z_i}{\sqrt{2\pi} t\sigma} \exp\left(-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right)}{\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)} \exp(\boldsymbol{\beta}' \mathbf{X}_{ijk}) \quad (5)$$

e

$$S(t_{ij} | \mathbf{X}_{ijk}, Z_i) = \left(\Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)\right)^{Z_i \exp(\boldsymbol{\beta}' \mathbf{X}_{ijk})} \quad (6)$$

Na análise de sobrevivência, o método mais amplamente utilizado para estimar o vetor de parâmetros $\boldsymbol{\tau} = (\theta, \mu, \sigma, \beta)$ do modelo (5) onde $Z_i \sim$ Gaussiana Inversa(1, θ) é o método de máxima verossimilhança, uma vez que pode incorporar dados censurados. A função de verossimilhança condicional (Wienke, 2010), no caso de n grupos de tamanho n_i (i, \dots, n) é dada por

$$L(\boldsymbol{\tau}) = \prod_{i=1}^n \int_0^\infty (z_i h_0(t_{ij}) e^{\boldsymbol{\beta}' \mathbf{X}_{ij}})^{\delta_{ij}} e^{-z_i H_0(t_{ij})} e^{\boldsymbol{\beta}' \mathbf{X}_{ij}} f(z_i; \theta) dz_i \quad (7)$$

Substituindo a função (4) e as funções de risco base $h_0(t_{ij})$ e risco acumulado base $H_0(t_{ij})$ da distribuição Log Normal em (7), a função de verossimilhança é reescrita da seguinte forma:

$$L(\boldsymbol{\tau}) = \prod_{i=1}^n \int_0^{\infty} \left(\left(\frac{1}{\sqrt{2\pi} t \theta} \exp \left[-\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right] \right) z_i e^{\beta' x_{ij}} \right)^{\delta_{ij}} e^{z_i \log \left(\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right) \right) e^{\beta' x_{ij}}} \times \left(\frac{1}{\sqrt{2\pi} \theta} z_i^{-3/2} \exp \left(-\frac{(z_i - 1)^2}{2\theta z_i} \right) \right) dz_i \quad (8)$$

onde $z_i > 1$ indica que os indivíduos do grupo i são frágeis, e $z_i < 1$ os indivíduos do grupo i são fortes e tem menor risco.

As estimativas de máxima verossimilhança são obtidas através da maximização numérica da função log-verossimilhança descrita em (8). Para estimação dos parâmetros, foi utilizado o pacote `parfm` (Journal of Statistical Software, 2012) do software R. A construção dos intervalos de confiança para os parâmetros do modelo são baseados nas propriedades de normalidade assintótica dos estimadores de máxima verossimilhança. Se $\hat{\boldsymbol{\tau}}$ denota os estimadores de máxima verossimilhança de $\boldsymbol{\tau}$ então a distribuição de $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ é aproximada por uma distribuição normal q -variada com média zero e matriz de covariância $I^{-1}(\hat{\boldsymbol{\tau}})$ onde $I(\boldsymbol{\tau})$ é a matriz de informação observada. Assim, um intervalo de confiança assintótico com $100(1 - \alpha\%)$ para cada parâmetro é dado por:

$$\hat{\boldsymbol{\tau}} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\boldsymbol{\tau})}$$

onde $\hat{\boldsymbol{\tau}}$ é o elemento da diagonal principal de $I^{-1}(\hat{\boldsymbol{\tau}})$ correspondente a cada parâmetro e $z_{\alpha/2}$ é o quantile $(1 - \alpha)\%$ da distribuição normal padrão. Para testar hipóteses relacionadas aos parâmetros $(\theta, \mu, \sigma, \beta)$ foram utilizados três testes: o de Wald, o da Razão de Verossimilhança (TRV) e o Escore (Colosimo e Giolo, 2006). Critérios de seleção de modelos como o critério de informação de Akaike (AIC) proposto por Akaike (1974), e o critério de informação bayesiano (BIC), proposto

por Schwarz et al. (1978), são frequentemente utilizados para selecionar modelos em diversas áreas. Os melhores modelos são considerados aqueles que apresentam menores valores de AIC e BIC. Para a adequação do modelo, diversos métodos encontram-se disponíveis na literatura e baseiam-se, essencialmente, nos resíduos de Cox-Snell, que auxiliam a examinar o ajuste global do modelo, o de Schoenfeld que tem coeficiente dependente do tempo, o de Martingal, que é dado pela diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, e de deviance, que facilitam na detecção de pontos atípicos (*outliers*).

Resultados e Discussão

Nesse estudo, o rebanho é formado por 102 bovinos, onde 56,9% é formado por machos e 43,1% por fêmeas, da raça Curraleiro Pé-Duro nascidos no período de 2005 a 2014. Ao considerar a sobrevivência desses animais no rebanho, constatou-se que 28,4% foram censurados, ou seja, permaneceram vivos até o período final da observação, e 71,6% falharam, isto é, morreram ou foram vendidos. Vários são os fatores que interferem na permanência do bovino no rebanho. Podemos destacar o peso ao nascimento, que é um dos fatores mais importantes na rentabilidade do rebanho bovino e tem grande influência na sobrevivência ao desmame (Carvalho et al., 2015). O controle do peso ao nascer é de suma importância e pode ser feita através da nutrição e genética sofrendo influência de raça, sexo, ano e mês de nascimento, efeitos paterno e materno, além de possíveis interações entre variáveis.

Foi observado que para a covariável estação de nascimento, o período da seca, com menor incidência de chuvas, ocorreu com maior predominância (65%), o que é visto de forma positiva, pois bezerro nascido na época da seca, a mãe durante a prenha teve mais pasto, logo mais seio e conseqüentemente bezerro mais gordo. Ainda, a diminuição das doenças no ambiente, pois no período chuvoso os bezerros tem um aumento da mortalidade. Em relação aos pesos padrões (nascer, desmama, ano e sobreano), foi encontrado uma frequência superior a 50% com relação ao

peso médio para para o peso a desmama, ao ano e 550 dias, indicando uma melhor concentração genética e uma boa adaptabilidade.

Para avaliar a influência das covariáveis observadas, descritas na Tabela 1, partiu-se de um modelo sem covariáveis e avaliou-se a inclusão de cada uma das delas, bem como a interação entre elas, usando o critério do teste da razão de verossimilhança (TRV). Na Tabela 3, estão as estimativas para os parâmetros do modelo de fragilidade compartilhada Gaussiana inversa- Log Normal, onde foram apresentadas apenas as estimativas dos parâmetros das covariáveis que foram significativas ao se considerar p-valor menor que 5%. Dentre as variáveis analisadas (EN, S, PN, PD, P365 e P550) foi observado significância nas variáveis EN, S e a interação do S com P365. A taxa de falha na estação de nascimento seca foi de $\exp(-0,743) = 0,475$ vezes se comparado com a estação de nascimento chuvosa, indicando que há diminuição do risco dos bovinos falhar (morte ou venda) na seca, dentre as que conseguem se manterem viva que pode ser explicado por ser um período com menos doenças para os bezerros como, verminoses, infecção no umbigo, ectoparasitas, carrapatos, dentre outras, além da oferta de nutrientes no pasto na época de prenhez para a mãe, demonstrando uma maior longevidade para os bovinos. Para a variável S, a taxa de falha para as fêmeas foi $\exp(1,676) = 5,344$ vezes a taxa de falhas para os machos, ou seja, bovinos fêmeas tem 5,344 vezes mais chances de falhar (morte ou venda) do que os bovinos machos influenciados por fatores como a maternidade. Como as fêmeas dão cria anualmente, na ausência de suplemento, as mesmas tiram dos ossos para colocar no leite durante a amamentação, tornando a vaca mais magra e mais sujeita a mortalidade. Para a interação da variável S com o P365, foi observado uma taxa de falha de $\exp(-2,715) = 0,066$ indicando uma diminuição do risco dos bovinos fêmeas com o peso aos 365 dias maior que 95 kg virem a falhar, ou seja, morrer ou sair para venda, influenciado por fatores como por exemplo, avaliação de carcaça, onde somente os bovinos machos vão para abate. A variância estimada para a fragilidade θ foi igual a 0,906, indicando a existência de associação significativa entre os tempos dos bovinos pertencentes a um

mesmo grupo (ano de nascimento) e também a existência de outros fatores não observados, e comum a todos os grupos, que também poderiam está influenciando na longevidade desses animais.

Tabela 3 Estimativas de máxima verossimilhança (EMV), erro-padrão (EP), intervalo de confiança (CI 95%), *p*-valor e razão de taxa de falhas (RTF) para os parâmetros do modelo de fragilidade compartilhada Gaussiana inversa- Log Normal

Parâmetros	EMV	EP	CI (95%)	<i>p</i> -valor	RTF
θ	0,906	0,749	-	-	-
μ	3,792	0,196	-	-	-
σ	0,458	0,079	-	-	-
β_1 (EN)	-0,743	0,348	[0,241; 0,941]	<0,033*	0,475
β_2 (S)	1,676	0,435	[2,279; 12,543]	<0,001*	5,344
β_4 (S*P365)	-2,715	0,590	[0,021; 0,210]	<0,001*	0,066

* ao nível de significância de 5% - sem efeito interpretativo

De acordo com os valores das fragilidades estimadas (\hat{z}_i), foi possível observar os bezerros em termos de longevidade de acordo com o ano de nascimento. Na Tabela 4 estão as estimativas para as fragilidades compartilhadas por grupos formados por ano de nascimento. Valores de \hat{z}_i maiores e menores do que 1 indicam bezerros que apresentaram, respectivamente, maiores e menores riscos de ficar inativo do rebanho. Assim, bezerros nascidos em 2008, 2010, 2011 e 2012 foram mais frágeis por ter valores maiores que 1, e conseqüentemente, uma menor longevidade. Essa fragilidade pode ser explicada pelos fatores ambientais de cada ano, como por exemplo, o índice pluviométrico, fatores genéticos, saída para leilões, dentre outros.

Tabela 4 Estimativa das fragilidades (Z_i) associadas a grupos de bovinos que nasceram no mesmo ano (2005 a 2014)

Ano	Estimativa
2005	0,282
2006	0,341
2007	0,261
2008	1,151
2009	0,936
2010	1,948
2011	1,282
2012	1,889
2014	0,642

As funções de risco (1) e sobrevivência (3) condicionado do modelo de fragilidade compartilhada Gaussiana inversa com risco base Log Normal foram dadas, respectivamente, por:

$$h(t_{ij} | \mathbf{X}_{ijk}, Z_i) = \frac{Z_i}{0,458t\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log(t)-3,792}{0,458}\right)^2\right) \Phi\left(\frac{-\log(t)+3,792}{0,458}\right) g(\boldsymbol{\beta}'\mathbf{X})$$

e

$$S(t_{ij} | \mathbf{X}_{ijk}, Z_i) = \left(\Phi\left(\frac{-\log(t) + 3,792}{0,458}\right) \right)^{Z_i g(\boldsymbol{\beta}'\mathbf{X})}$$

onde $g(\boldsymbol{\beta}'\mathbf{X}) = \exp\{-0,743EN + 1,676S - 2,715(S * P365)\}$

Foi feito um estudo comparativo do modelo de fragilidade Gaussiana inversa usando as funções de risco base log-normal e log-logístico, descritos na Tabela 5, onde foi utilizado o critério de Akaike (AIC) e critério de informação bayesiano (BIC), para seleção de modelos, mostrando que o modelo de fragilidade Gaussiana inversa- Log Normal é o mais adequado por apresentar o menor valor de AIC e BIC.

Tabela 5 Resultados para os parâmetros do modelo de fragilidade compartilhada Gaussiana inversa com funções de risco base Log Normal e log-logístico

Parâmetros	Log Normal		Log-logístico	
	Estimativas	Erro padrão	Estimativas	Erro padrão
θ	0,906	0,749	0,689	0,504
μ	3,792	0,196	-	-
σ	0,458	0,079	-	-
α	-	-	-13,500	1,642
γ	-	-	3,475	0,499
$\beta_1(\text{EN})$	-0,743	0,348	-0,741	0,350
$\beta_2(\text{S})$	1,676	0,435	1,738	0,437
$\beta_3(\text{P365})$	0,584	0,357	0,658	0,353
$\beta_4(\text{S*P365})$	-2,715	0,590	-2,773	0,591
Log-Verossimilhança	-365,793		-368,327	
AIC	745,587		750,654	
BIC	763,961		769,029	

Para avaliar a qualidade do ajuste do modelo, na Figura 1, estão os gráficos dos resíduos de Cox-Snell, resíduos Martingal e resíduos Deviance. Nos gráficos (a) e (b), observa-se que os resíduos de Cox-Snell seguem aproximadamente uma distribuição exponencial padrão, o que indica uma qualidade de ajuste global aceitável do modelo. Os gráficos (c) e (d) não sugerem a existência de pontos que possam ser considerados atípicos.

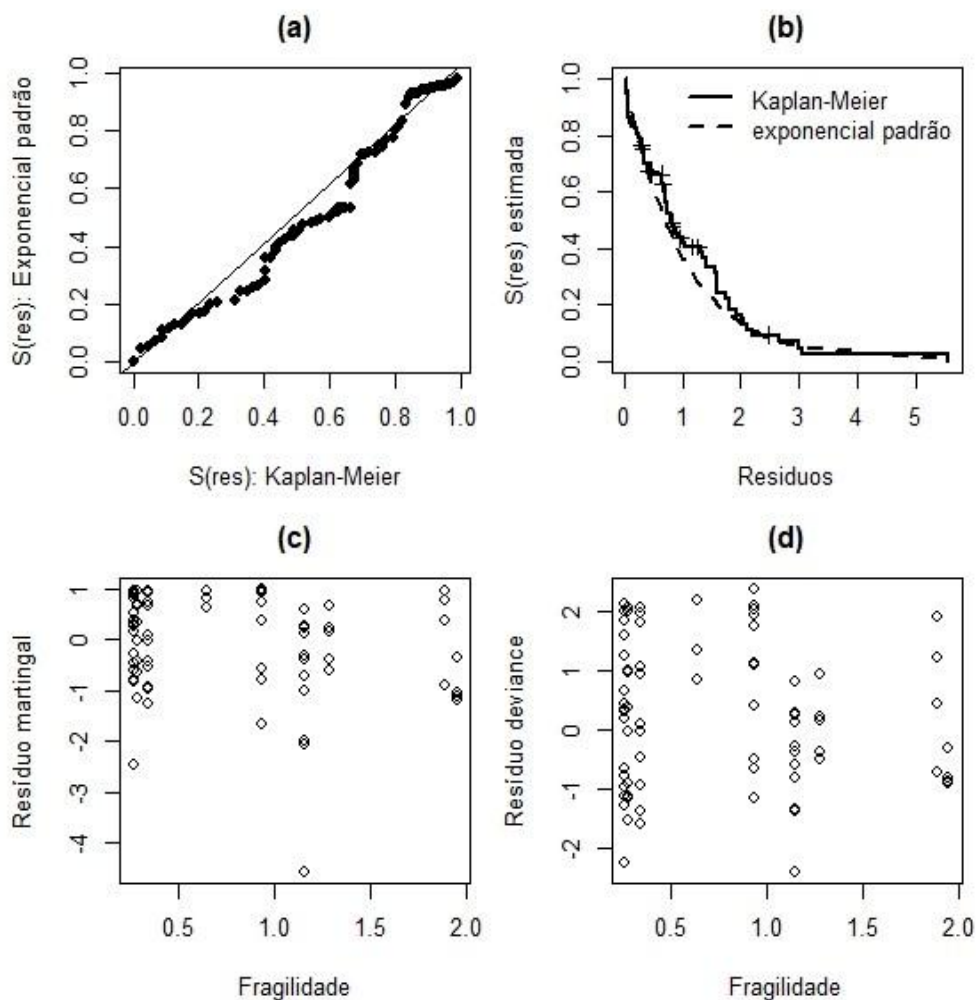


Figura 1 Resíduos de Cox-Snell , Martingal e Deviance do modelo de Fragilidade Gaussiana Inversa-Log Normal para dados de bovinos da raça Curraleiro Pé-Duro.

Conclusões

O uso de modelos de fragilidade compartilhada mostrou-se ser um critério de seleção adequado às situações propostas. Também permitiu incorporar um termo para a heterogeneidade não-observada que afeta a estimativa do risco, onde foi possível observar para os modelos de fragilidade multivariadas, a significância estatística nas covariáveis estação de nascimento, sexo e a interação do sexo com o peso aos 365 dias afetando na longevidade do bovino. Nesse sentido, o

modelo de fragilidade compartilhada pode ser utilizada como ferramenta de ajuda no melhoramento animal.

Referências

- Andersen, P. K.; Borgan, O.; Gill, R. D.; Keiding, N. **Statistical models based on counting processes**. Springer Science & Business Media, 2012.
- Caetano, S. L.; Rosa, G. J. M.; Savegnago, R. P.; Ramos, S. B., Bezerra, L. A. F.; Lôbo, R. B.; Munari, D. P. Characterization of the variable cow's age at last calving as a measurement of longevity by using the Kaplan–Meier estimator and the Cox model. **Animal**, v. 7, n. 4, p. 540-546, 2013.
- Carvalho, G. M. C., Lima Neto, A., Da Frota, M. N. L., de Sousa, V. R., Carneiro, M. d. S., Vêras, G. M. C., Costa, R. N. M., e Azevedo, D. M. M. R. (2015). O uso de bovinos Curraleiro Pé-Duro em cruzamentos para produção de carne de boa qualidade no trópico quente-fase 1. In *Embrapa MeioNorte-Artigo em anais de congresso (ALICE)*. In: CONGRESSO NORDESTINO DE PRODUÇÃO ANIMAL, 10., 2015, Teresina.
- Fioravanti, M. C. S., Juliano, R. S., Costa, G. L., Abud, L. J., Cardoso, V. S., Carpio, M. G., Oliveira, M. F., et al. (2011). Conservación del bovino curraleiro: cuantificación del censo y caracterización de los criadores. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, 48:109–116.
- Hougaard, P. (2012). *Analysis of multivariate survival data*. Springer Science & Business Media.
- Moore, D. F. (2016). *Applied survival analysis using R*. Springer.
- Salles, P., Medeiros, G., Costa, R. G., Ramos, C. T. C., Borburema, J., Rocha, M., Rocha, L., e Mathias, W. (2011). Programa de conservação e melhoramento de uma raça bovina brasileira: Curraleiro Pé-Duro. *AICA-Actas Iberoamericana de Conservacion animal, Córdoba*, 1(1):453–466.

7 CONSIDERAÇÕES GERAIS E PROPOSTAS FUTURAS

7.1 Considerações Gerais

Neste trabalho, o objetivo geral foi aplicar a metodologia de Análise de Sobrevida como alternativa para análises em estudos genéticos relacionados a bovinos da raça Curraleiro Pé-Duro, no contexto em que a variável resposta está sendo influenciada por fatores não-observáveis. Dessa forma, foram utilizados modelos que podem ser usados como critérios de seleção de bovinos.

No capítulo 4, através do modelo de riscos proporcionais de Cox- Log Normal foi possível identificar a existência de fatores que influenciaram no tempo a longevidade dos bovinos da raça Curraleiro Pé-Duro considerando a inatividade ocasionada por morte ou venda do bovino como falha e a censura a permanência no rebanho. A estação seca e os bovinos machos se mostraram mais longevos ao rebanho.

No capítulo 5 e 6, através do modelo de fragilidade univariada e compartilhada Gaussiana Inversa- Log Normal, que é uma extensão do modelo de riscos proporcionais de Cox, foi possível quantificar a heterogeneidade não-observada, devido a fatores genéticos ou ambientais não-observados, de para cada bovino e compartilhadas por grupos de bovinos. O modelo mostrou-se eficiente para atender aos objetivos propostos.

Por fim, os métodos apresentados nesse trabalho, por meio da análise de sobrevivência, podem ser considerados como ferramentas de grande uso no melhoramento animal.

7.2 Propostas Futuras

1. Analisar os fatores que interferem no ganho de peso em um dado intervalo de tempo através dos modelos de fragilidade;
2. Aplicar modelos de longa duração, para modelar o tempo que um bovino leva para atingir um peso ideal.