UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE MATEMÁTICA

MATHEUS HENRIQUE PIMENTA ZANON

ANÁLISE DE SOBREVIVÊNCIA APLICADA À DADOS PENAIS DE REINCIDÊNCIA AO CRIME DA COMARCA DE PRIMEIRO DE MAIO - PR

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO

2016

MATHEUS HENRIQUE PIMENTA ZANON

ANÁLISE DE SOBREVIVÊNCIA APLICADA À DADOS PENAIS DE REINCIDÊNCIA AO CRIME DA COMARCA DE PRIMEIRO DE MAIO - PR

Trabalho de Conclusão de Curso apresentado ao Departamento Acadêmico de Matemática da Universidade Tecnológica Federal do Paraná como requisito para obtenção do grau de "Licenciado em Matemática" – Área de Concentração: Matemática.

Orientador: Dr. Emílio Augusto Coelho Barros

CORNÉLIO PROCÓPIO

2016

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE MATEMÁTICA

FOLHA DE APROVAÇÃO

BANCA EXAMINADORA

Prof. Dr. Emílio Augusto Coelho Barros
Orientador

Prof. Dr. Roberto Molina de Souza

Prof. Me. Vinicius Araujo Peralta

CORNÉLIO PROCÓPIO

2016

A minha mãe, Maria de Lourdes, que é a grande responsável pela minha formação pessoal, pelo seu amor e dedicação à mim. Aos meus avós maternos (*in memoriam*), Artur e Maria José, que juntamente com minha mãe, dedicaram seu amor para a minha formação.

AGRADECIMENTOS

A Deus, pelo dom da vida e suas inúmeras graças durante esta caminhada.

A minha namorada e parceira, Heloiza, que durante todos os momentos sempre esteve me apoiando, motivando e ajudando com muito amor.

A minha família, que desde o início sempre apoiou minhas decisões.

Ao Emílio, meu orientador, pela orientação e atenção dispensada ao trabalho, pelas sugestões e direções tomadas.

Ao professor Roberto, pelas sugestões, conversas e paciência.

A professora Elisangela pelas contribuições.

Aos amigos da UTFPR, pela amizade, conversas e motivação durante toda esta jornada.

Ao SECAT da comarca de Primeiro de Maio, que gentilmente disponbilizou as informações para a realização do trabalho.

Aos professores Vinicius e Glaucia pelas conversas e amizade.

A UTFPR e todo seu corpo docente de excelente qualidade que realizaram seu trabalho com enorme dedicação para que os alunos tenham um ensino de qualidade.

A todos que de forma direta e indireta participaram no desenvolvimento deste trabalho.

RESUMO

PIMENTA-ZANON, M. H.. ANÁLISE DE SOBREVIVÊNCIA APLICADA À DADOS PENAIS DE REINCIDÊNCIA AO CRIME DA COMARCA DE PRIMEIRO DE MAIO - PR. 79 f. Trabalho de Conclusão de Curso – Departamento Acadêmico de Matemática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

O sistema carcerário brasileiro é ponto de grandes discussões em nosso atual momento político e social. De acordo com os últimos relatórios fornecidos pela CPI dos Deputados (2009), DEPEN (2014) e IPEA (2015) a análise sobre os indivíduos reincidentes é escassa na literatura devido à falta de dados concretos sobre o sistema carcerário brasileiro. Os números apontam que algo em torno de $24,5\%$ dos indivíduos são reincidentes no Brasil, segundo o IPEA (2015). Logo, o objetivo deste trabalho é realizar uma análise de dados do Setor de Carceragem Temporária (SECAT) da comarca de Primeiro de Maio - PR. O conjunto de dados contém informações dos indivíduos no período de dezembro de 2009 à dezembro de 2015, somando 356 observações, sendo o evento de interesse o tempo até a reincidência ao primeiro crime de cada indivíduo no período observado. A incorporação de covariáveis ao estudo é realizada com o objetivo de estudar algumas características dos indivíduos observados em relação ao seu tempo de reincidência ao crime. Inicialmente são realizados comparações entre as distribuições Burr XII e Weibull (com e sem fração de cura) no âmbito da análise de sobrevivência para avaliar qual modelo melhor se ajusta aos dados em estudo. Também são realizadas comparações entre as técnicas frequentista, via estimação de máxima verossimilhança, e Bayesiana. A incorporação de covariáveis ocorre na distribuição Burr XII, que se mostrou mais flexível se comparada com os outros modelos estudados. Para a obtenção de estimadores intervalares e pontuais para os parâmetros dos modelos propostos o software `SAS` é utilizado.

Palavras-chave: Análise de Sobrevivência, Modelos de Longa Duração, Distribuição de Weibull, Distribuição Burr XII.

ABSTRACT

PIMENTA-ZANON, M. H.. APPLIED SURVIVAL ANALYSIS OF RECIDIVISM CRIMINAL DATA CRIME OF THE DISTRICT OF PRIMEIRO DE MAIO - PR. 79 f. Trabalho de Conclusão de Curso – Departamento Acadêmico de Matemática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

The Brazilian prison system is point of great discussion in our current political and social moment. According to the latest reports provided by CPI of Deputados (2009), DEPEN (2014) and IPEA (2015) analysis of repeat offenders individuals is scarce in the literature due to lack of concrete data on the Brazilian prison system. The numbers show that around 24.5% of individuals are repeat offenders in Brazil, according to the IPEA (2015). Therefore, the objective of this paper is conduct a data analysis in Setor de Carceragem Temporária (SECAT) of the Primeiro de Maio - PR. The data set contains individuals information in the period from december 2009 to december 2015, totaling 356 observations, and the event of interest is the time to recurrence to the first crime for each individual in the observed period. The incorporation of covariates in the study is performed in order to study some characteristics of the observed individuals in relation to time of recurrence crime. Initially, comparisons are made between the Burr XII and Weibull distributions (with and without cure fraction) in the survival analysis to verify what model is best fited to the data in the study. Also comparisons are made between the frequentist inference, by the maximum likelihood estimating, and Bayesian inference. The incorporation of covariates occurs in the distribution Burr XII, which was more flexible compared to the other models. To obtain the interval and point estimates for the parameters of the proposed models the `SAS` software is considered.

Keywords: Survival Analysis, Long-term Survival Models, Weibull Distribution, Burr XII Distribution.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

DEPEN	Departamento Penitenciário Nacional
EUA	Estados Unidos da América
CPI	Comissão Parlamentar de Inquérito
IPEA	Instituto de Pesquisa Econômica Aplicada
CDP	Centro de Detenção Provisória
SECAT	Setor de Carceragem Temporária
DP	Delegacia de Polícia
EMV	Estimadores de Máxima Verossimilhança
MCMC	Monte Carlo via Cadeia de Markov
ECA	Estatuto da Criança e Adolescente
AIC	*Akaike Information Criterion*
DIC	*Deviance Information Criterion*

LISTA DE SÍMBOLOS

$\in$	Pertence
$f(t)$	Função Densidade de Probabilidade
$S(t)$	Função de Sobrevivência
$h(t)$	Função de Risco
$H(t)$	Função Acumulada de Risco
P	Probabilidade
F	Função de Distribuição Acumulada
$H(t)$	Função Acumulada de Risco
#	Número de Observações
$\approx$	Aproximado
$\propto$	Proporcional
$\times$	Multiplicação

SUMÁRIO

1 INTRODUÇÃO

O cenário penal brasileiro é fruto de uma história que tem início no século XIX, onde foi abdicado os antigos tratamentos de punição para a utilização do método de privar a liberdade dos indivíduos que praticassem crimes. A primeira constituição datada em 1824 em seu artigo 179, parágrafo XXI já determinava que as cadeias deveriam ser "seguras, limpas, e bem arejadas, havendo diversas casas para separação dos Réos, conforme suas circumstancias, e natureza dos seus crimes."BRAZIL (1824)

Desde 1935, o Código Penitenciário Federal já esclarece que os locais onde serão alojados os indivíduos privados de liberdade deverão fornecer local adequado e uma forma de reintegração à sociedade.

Com a criação da Lei de Execução Penal em 1984 foram determinados vários deveres ao Estado, dentre os quais destacam-se: assistência material, onde o Estado fornecerá alimentos, vestimentas e instalações higiênicas, assistência à saúde de caráter preventivo e curativo, fornecendo todo e qualquer amparo para a saúde física e mental dos detentos, assistência jurídica aos presos que não possuem condições financeiras para arcar com o processo, assistência educacional fornecendo instrução escolar e profissionalização dos detentos, assistência social fornecendo amparo aos detentos de modo que este consiga retornar ao convívio em sociedade, assistência religiosa, permitindo a participação e realização de cultos nos centros de detenção e ainda a possibilidade do detento possuir livros sagrados de sua crença e assistência ao egresso que consiste no amparo ao retorno à sociedade e ainda se necessário ceder moradias específicas para o egressante no prazo máximo de 2 (dois) meses.

Vários estudos, como os de Silva e Sellos-Knoerr (2015), Campos e Sousa (2013) e Marques (2013), apresentados na última década apontam que os detentos preferem locais onde há trabalho para ser realizado durante o dia, de modo que não fiquem no ócio e que estejam produzindo algo.

Diante de todo esse amparo legal, o sistema penal é deficiente. Percebe-se que tal adjetivo é real através dos dados apresentados pelo Departamento Penitenciário Nacional (DEPEN)

no primeiro semestre de 2014 no Levantamento Nacional de Informações Penitenciárias.

Os dados são estes: a população carcerária brasileira era de 607.731 indivíduos, sendo que a capacidade máxima do sistema penitenciário é de 376.669 vagas, isto é, 161% de taxa de ocupação no sistema penitenciário, de modo prático, um local que foi planejado para receber 10 indivíduos está com 16 indivíduos, podendo em casos extremos, alocar uma quantidade superior a esta apresentada.

Comparando o Brasil com outros países, no ranking de países que mais possuem presos em números absolutos, o Brasil ocupa a quarta posição. Se comparar com a taxa de ocupação a cada 100.000 habitantes o Brasil ocupa a quinta posição. No período de 1995 a 2010 o Brasil ocupou a segunda posição dos países que mais tiraram a liberdade de pessoas devido à crimes, perdendo somente para a Indonésia, que ainda assim, possui uma população carcerária bem inferior a do Brasil com 167.163 indivíduos. Ainda analisando a taxa de aprisionamento, o Brasil está indo à contramão dos outros três países que possuem elevadas populações carcerárias, enquanto China, Estados Unidos da América (EUA) e Rússia estão diminuindo suas taxas o Brasil está em crescimento.

De 1990 à 2014, a população carcerária teve um aumento de 575%, sendo três quartos das vagas para indivíduos do sexo masculino e o restante dividi-se em vagas femininas e mistas. As faixas etárias que possuem maior número absoluto de indivíduos são de 18 à 29 anos, sendo que a maioria da população carcerária é autodeclarada negra e mais da metade da população possui apenas o ensino fundamental incompleto.

Segundo o DEPEN (2014), os crimes tentados ou consumados mais comuns são tráfico de drogas e roubo com 27% e 21%, respectivamente, onde a maioria do sexo masculino cometerá ou tentará praticar tráfico (25%) e roubo (21%), já 63% dos indivíduos do sexo feminino cometerá ou tentará praticar o crime de tráfico.

Desse modo, o principal objetivo destes locais que é a ressocialização do indivíduo à sociedade, fica negligenciado. Vários fatores além destes apresentados corroboram para que o indivíduo egressante não consiga uma efetiva ressocialização, fazendo com que este tornem-se reincidente.

No Brasil a reincidência é tida como um agravante de pena de privação de liberdade, impede o indivíduo de cumprir pena de reclusão no regime semiaberto, impede o pagamento de multa em vez de prisão para crimes dolosos, aumenta o período de reclusão em diversas modalidades de penas entre outros agravantes (BRASIL, 1984)

Estudos sobre reincidência no Brasil são raros, visto que são inúmeras as divergências

que ocorrem na construção de metodologias, materiais e dados sobre o assunto, assim, surgem diversos valores quando são apresentados pela mídia sobre a real taxa de reincidentes no país.

As divergências ocorrem de diversas maneiras, sendo um exemplo a definição de reincidente, que segundo BRASIL (1940) um indivíduo é considerado reincidente se este for condenado judicialmente por algum crime no prazo de cinco anos, após a extinção da pena anterior. Porém diversas outras definições são encontradas na literatura, onde temos a reincidência penitenciária, onde o egresso retorna a penitenciária após uma pena ou por medida de segurança, reincidência criminal, onde há mais de uma condenação independente do prazo legal e a reincidência genérica quando há mais de um ato criminal, independente da condenação ou autuação (CHIQUEZI, 2009).

Devido a esta ampla definição, diversos recortes são feitos nos poucos estudos disponíveis para leitura, estudos em sua maioria com mais de décadas de conclusão, onde a realidade penal era completamente diferente da atual.

O relatório final da Comissão Parlamentar de Inquérito (CPI) de 2008, divulgou que a taxa de reincidentes era em torno de 70% à 80% dependendo da unidade federativa em que se encontrava o detento, contudo a CPI não produziu uma pesquisa para verificar a veracidade de tais dados (DEPUTADOS, 2009).

O Instituto de Pesquisa Econômica Aplicada (IPEA) divulgou um relatório analisando dados sobre a reincidência legal de algumas unidades federativas do Brasil: os dados mostram que em torno de 24,5% dos indivíduos são reincidentes, que o maior percentil de indivíduos reincidentes são da faixa etária de 18 à 24 anos do sexo masculino. Quanto a cor, a maioria dos reincidentes são da cor branca, em torno de 53,7%. Sobre a escolaridade e ocupação a maioria dos reincidentes possuem o ensino fundamental incompleto, chegando a 58,5% dos reincidentes e 92,5% possuem ocupação. Sobre o tipo penal imputado sobre os indivíduos a maioria dos reincidentes cometeram crimes contra o patrimônio, ultrapassando os 50% (IPEA, 2015).

1.1 SISTEMA PENAL PARANAENSE

No mês de fevereiro de 2016 o estado do Paraná possuía uma população carcerária de 28.261, o que com isto, ocupava a quinta posição dos estados que mais possuem presos em números absolutos. De 2011 à 2014 o número de indivíduos privados de liberdade diminuiu em torno de $7,9\%$. Ainda assim a taxa de superlotação é grande, existindo o dobro de indivíduos por vagas alocados nos Setor de Carceragem Temporária (SECAT) e 4% de indivíduos a mais

nas penitenciárias. No Paraná, 18.042 indivíduos do sexo masculino e 1.047 indivíduos do sexo feminino estavam privados de liberdade alocados em Penitenciárias e 8.455 indivíduos do sexo masculino e 688 indivíduos do sexo feminino privados de liberdade alocados no SECAT. A taxa de ressocialização do sistema paranaense é de aproximadamente 42%, de indivíduos que trabalham é de 21% e que estudam é de 27%. Estes dados são fornecidos através do Sistema de Transparência na Gestão Carcerária (SIGEP-PR) sendo atualizados diariamente.

Os SECAT por lei, deveriam receber somente indivíduos capturados em flagrante e estes permanecerem somente o tempo necessário para as atribuições legais, após este período, ficariam à disposição da Justiça para serem soltos ou transferidos aos Centro de Detenção Provisória (CDP) , algo que não ocorre na maioria dos SECAT paranaense. Os SECAT abrigam cerca de 32% de toda população carcerária do Paraná, são locais que afrontam a dignidade das pessoas que ali estão, sendo em sua maioria, locais sem nenhum tipo de estrutura para manter em cárcere os indivíduos, violando a Lei de Execução Penal.

1.2 CARACTERIZAÇÃO DO SECAT DE PRIMEIRO DE MAIO

Primeiro de Maio é um município paranaense pertencente à região metropolitana de Londrina, com aproximadamente 11 mil habitantes, onde suas principais atividades econômicas são agricultura e turismo. Nitidamente uma cidade pequena e pouco desenvolvida, possui apenas uma Delegacia de Polícia Civil, localizada no centro da cidade, ao lado do Fórum Municipal.

O SECAT está localizado no interior da Delegacia de Polícia (DP) , possuindo capacidade máxima para seis indivíduos, dividindo-se em três celas, sendo estas subdivididas em 01 cela para indivíduos do sexo masculino, 01 cela para indivíduos do sexo feminino e 01 cela para indivíduos que cometeram crimes contra a dignidade sexual, devido ao fato dos outros detentos não aceitarem estes indivíduos na mesma cela, garantindo sua integridade física.

A estrutura física é bem precária, não possui nenhum tipo de sala especial, consultório, salas de atendimento ou refeitório. A refeição dos detentos é servida a partir de uma empresa terceirizada pelo Estado que fornece duas refeições diárias (marmitas) para cada detento. Não é desenvolvido nenhum tipo de trabalho ou ação educativa pelos detentos por falta de estrutura física e pessoal. Os detentos possuem o direito do banho de sol e as pastorais religiosas comparecem periodicamente para um apoio espiritual.

Ao término do periódo de observação, o SECAT possuía 18 detentos divididos em 02 indivíduos do sexo feminino, 02 indivíduos que cometeram crimes contra a dignidade sexual e 14 indivíduos do sexo masculino.

A única competência que o SECAT de Primeiro de Maio consegue cumprir é da custódia dos detentos que ali estão.

1.3 CONTEXTUALIZAÇÃO

A situação carcerária é tema de grandes debates no cenário político e social. Conhecer alguns aspectos deste cenário proporciona novos horizontes, estudos de diversas áreas estão sendo realizados com este intuito: ampliar o debate e entender o que esta relacionado ao tipo de perfil dos indivíduos e se há correlações com alguma covariável em comum.

Este estudo, buscou utilizar-se de um recorte adequado para o estudo, onde é considerado "reincidente" o indivíduo que por algum motivo teve mais que uma passagem pela DP da comarca de Primeiro de Maio, buscando utilizar técnicas de análise de sobrevivência para acrescentar informações dos indivíduos que passam pela DP e esboçar um possível perfil do indivíduo reincidente da comarca de Primeiro de Maio.

Um dos objetivos deste trabalho será traçar o perfil do reincidente da comarca de Primeiro de Maio/PR, através de covariáveis que foram fornecidas pelo Sistema de Registros Policial da Polícia Civil do Paraná e pelo livro de detentos do SECAT. As covariáveis serão sexo, profissão, tipo de crime cometido, tipo de benefício adquirido ao sair do SECAT, cor da pele e faixa etária.

Outro objetivo, é utilização da distribuição de Burr XII em dados penais, sendo uma alternativa à distribuição de Weibull, comumente utilizado na literatura em dados deste tipo. Serão realizadas comparações entre as duas distribuições para verificar a usabilidade das mesmas em dados deste tipo.

Para a conclusão dos objetivos acima, serão estudados diversos modelos de análise de sobrevivência, métodos paramétricos e não-paramétricos, inferência frequentista (via estimação de máxima verossimilhança dos parâmetros), inferência Bayesiana, análise de regressão linear e a utilização de softwares como R e SAS.

O trabalho está organizado em quatro capítulos. No primeiro capítulo é apresentado uma contextualização do tema, o problema proposto, os objetivos do trabalho, a contribuição para a comunidade e a estrutura do trabalho. No segundo capítulo é desenvolvido toda a metodologia do trabalho, apresentando toda a parte teórica que serviu de base para o progresso do trabalho, apresentando os estimadores não paramétricos e paramétricos, diversas distribuições, modelos de longa duração e inferências clássicas e Bayesiana, juntamente com o estudo da regressão dos parâmetros. No terceiro capítulo são apresentados os resultados obtidos através das

inferências realizadas com os dados dos ex-detentos da comarca de Primeiro de Maio - PR. No quarto e último capítulo é realizada a análise dos resultados obtidos no capítulo anterior e são descritas algumas perspectivas futuras em relação ao desenvolvimento do trabalho. O trabalho contém ainda apêndices que contêm, no apêndice A e B são apresentadas algumas ferramentas utilizadas na análise dos dados, no apêndice C apresenta-se a divisão das categorias de crimes feita pelo Código Penal Brasileira e no apêndice D são apontados os códigos do software `SAS` e por fim no apêndice E são apresentados os gráficos de convergência dos parâmetros que foram obtidos através de simulações do método de Monte Carlo via Cadeia de Markov.

2 METODOLOGIA

A análise de sobrevivência consiste em um conjunto de técnicas estatísticas utilizadas em diversas áreas, tais como: engenharia, realizando testes para estudar a duração de certos produtos, vida útil de certos componentes, entre outros exemplos que podem ser verificados em Meeker e Escobar (1998); criminal, estudando o tempo de reincidência de indivíduos privados de liberdade e tempo de ocorrência de crimes, Araujo (2004) apresentou um estudo utilizando dados da Penitenciária Estadual de Maringá; demográfica, estudando tempos de nascimentos, mortes, casamentos e outros eventos de interesse; médica, realizando estudos em doenças, novos medicamentos, novos tratamentos, dentro da área médica podemos citar trabalhos como de Lambert et al. (2007) que aplica a estatística em amostras de populações que possuem câncer, Coelho-Barros (2014) expõe um trabalho semelhante ao de Lambert et al. (2007), tendo como objetivo estudar diversos modelos, inclusive Weibull e Burr XII, para análise de sobrevivência em casos univariado e bivariados, isto é, dois tempos de sobrevivência para cada unidade amostral; Martinez et al. (2013) em um estudo também na área da saúde, apresenta uma análise com a distribuição de Weibull modificada, utilizando inferência Bayesiana , utilizando-se de um banco de dados obtido do Hospital do Câncer de Barretos.

A Análise de Sobrevivência, usualmente, é utilizada na análise do tempo até a ocorrência de eventos de interesse.

Caracteriza-se este tempo como tempo de falha, em que esta "falha" não implica necessariamente em falhas no sentido de defeito ou erro, podendo ser o tempo até a cura de uma doença, morte de um paciente, reincidência de um indivíduo ao crime ou ainda inadimplência de um serviço. Dessa forma, a análise de sobrevivência pode responder algumas questões do tipo: " Qual é a proporção da população que sobrevive a um dado tempo? Daqueles que sobrevivem, qual é a taxa de óbitos? Quais as características ou tratamentos que aumentam ou não a probabilidade de sobrevivência?".

O aspecto mais relevante na análise de sobrevivência é a presença de observações censuradas, isto é, quando o evento de interesse não é observado para uma certa unidade amos-

tral. As causas desta censura são as mais diversas possíveis, podendo ser por exemplo uma desistência durante o estudo por parte do indivíduo ou ainda uma observação onde houve a "falha" por outras causas e não devido ao evento de interesse. Com isto sabemos que todo conhecimento sobre o tempo de falha do evento em estudo é superior ao período observado. Caso o pesquisador desconsiderasse as informações contidas nas censuras, as técnicas da estatística usuais conseguiriam responder as perguntas existentes, porém, de forma viesada.

Quando este tempo de "falha" coincide com o sentido literal de falha de algum equipamento na área da engenharia e afins, a análise de sobrevivência é chamada de análise de confiabilidade, na qual os estudos são voltados para o tempo até que certo equipamento fique inapto ao uso ou apresente uma primeira falha. O grande objetivo em estudos deste tipo é para uma melhor estimação do tempo de vida do produto, que é uma informação crucial para o fabricante, como por exemplo o tempo de falha até que uma lâmpada de um projetor venha a queimar.

Este "tempo" também não precisa estar necessariamente ligado ao seu sentido literal, podendo ser outra unidade de medida, como por exemplo o caso em que se mede a quilometragem rodada de um pneu até o aparecimento de algum defeito, ou ainda o número de ciclos de uma linha de produção até o registro de falha na montagem.

O diferencial das técnicas utilizadas pela análise de sobrevivência é a não exclusão de observação incompletas, ou seja, na análise de sobrevivência existe observações censuradas. A não utilização destas informações censurada levam a perda de dados e análise incorreta das informações presentes.

A censura ocorre por diversos motivos, sendo os mais comuns o término do experimento sem a ocorrência do evento de interesse. Em estudos na área médica a desistência de pacientes durante o estudo, pacientes que mudaram de cidades, são alguns exemplos.

As censuras são classificadas em diversos tipos, sendo a mais comum e a presente neste estudo, censura do tipo I e à direita. Neste tipo de censura um tempo de observação é fixado e a observação do evento de interesse é uma variável aleatória, isto é, quando o tempo de ocorrência do evento de interesse é maior que o tempo de observação do estudo. Em outras palavras:

$$Y_i = min(T_i, t_c) = \begin{cases} T_i, & \text{se} \quad T_i \leq t_c \\ t_c, & \text{se} \quad t_c < T_i \end{cases}, (T \in \mathbb{R}; T \geq 0) \tag{1}$$

em que Y_i são as observações, T_i é o tempo de cada observação, t_c é o tempo pré estabelecido do estudo e $\in$ significa "pertence".

Com a variável T sempre positiva e usualmente contínua ($T \geq 0$), define-se algumas funções de grande importância para a análise de sobrevivência sendo elas: Função Densidade de Probabilidade, Função de Sobrevivência e Função de Risco.

Sendo $T \geq 0$ a função densidade de probabilidade é a função que descreve o tempo de sobrevivência de um indivíduo como o limite da probabilidade de um indivíduo morrer em um intervalo de tempo $[t, t+\Delta t]$, ou seja:

$$f(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t+\Delta t)}{\Delta t} \tag{2}$$

em que a função $f(t)$ é sempre positiva para todo $t > 0$.

Para representar a função de distribuição acumulada denota-se:

$$F(t) = P(T \leq t) = \int_0^t f(x)dx \tag{3}$$

A função de sobrevivência é a probabilidade de cada indivíduo sobreviver até o tempo t e é dada por:

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x)dx \tag{4}$$

Observe que $S(t)$ é monótona decrescente então $S(0) = 1$ e $S(\infty) = \lim_{t\to\infty} S(t) = 0$ e que a função de distribuição acumulada é definida como a probabilidade de um indivíduo não sobreviver até o tempo t, ou seja, $F(t) = 1 - S(t)$.

A outra função de grande importância é a função de risco $h(t)$ que é a função que descreve os riscos de falha durante o tempo $[t, t+\Delta t]$, em outras palavras, a função de risco apresenta que os riscos durante certo tempo t variam, e é descrita como:

$$h(t) = \lim_{\Delta t \to \infty} \frac{P(t \leq T < t+\Delta t | T \geq t)}{\Delta t} \tag{5}$$

tal que,

$$h(t) = \frac{f(t)}{S(t)} \tag{6}$$

em que $f(t)$ é a função densidade de probabilidade, que é obtida através da derivada da função de sobrevivência em relação a t.

A função de risco $h(t)$ é estritamente positiva, podendo ser monótona crescente ou não crescente (decrescente), ou não monótona em forma de banheira, ou seja, no tempo inicial o risco é maior e com o passar do tempo o risco decresce até chegar a um valor mínimo constante e após este valor a função se torna crescente.

Conhecendo estas três funções pode-se apresentar algumas relações existentes entre elas, com isto, conhecendo apenas uma dentre elas, consegue-se obter qualquer outra função através das relações abaixo:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d\log S(t)}{dt}; \tag{7}$$

$$f(t) = -\frac{d}{dt}S(t); \tag{8}$$

$$H(t) = \int_0^t h(t)dt = -\log S(t); \tag{9}$$

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t H(t)dt\right\} \tag{10}$$

em que $f(t)$ é a função densidade de probabilidade, $S(t)$ é a função de sobrevivência, $h(t)$ é a função de risco, $H(t)$, P é a probabilidade, F é a função de distribuição acumulada e $H(t)$ é a função acumulada de risco.

2.1 ESTIMADOR PRODUTO LIMITE DE KAPLAN-MEIER

Para realizar as estimações dos parâmetros das funções apresentadas anteriormente são consideradas algumas técnicas de estimação, sendo técnicas paramétricas e não paramétricas.

O estimador de Kaplan e Meier (1958a), ou estimador produto limite, é um estimador não paramétrico para a função de sobrevivência $S(t)$. Este estimador é uma adaptação da função empírica de sobrevivência, onde, na ausência de censura é definida como:

$$\hat{S}(t) = \frac{\#(t_i > t)}{n} \tag{12}$$

em que n é o número total de observações no estudo e # significa número de observações.

A função $\hat{S}(t)$ é uma função escada nos tempos observados, onde o tamanho desta escada é de $\frac{1}{n}$, caso exista empates nestes tempos t o tamanho da escada é multiplicado pelo número de empates.

O estimador é construído a partir de probabilidades condicionais, em que dado uma informação anterior, sabe-se a próxima, ou seja, para um indivíduo estar vivo em um tempo $t = 5$ este indivíduo deve ter sobrevivido a um tempo $t = 1$, onde os tempos $t = 1$ e $t = 5$ são as respectivas censuras ordenadas. Então, considerando que:

- As censuras são todas ordenadas, ou seja, $t_1 < t_2 < ... < t_k$;
- d_j é o número de falhas no tempo t_j, onde $j = 1, ..., k$;

- n_j é o número de indivíduos vivo imediatamente antes do tempo t_j.

tem-se o estimador de Kaplan-Meier dado por:

$$\hat{S}(t) = \prod_{i:t_j<t} \left(\frac{n_i - d_i}{n_i} \right) \tag{13}$$

sendo (13) o estimador de máxima verossimilhança da função $S(t)$, provado no artigo original de Kaplan e Meier (1958a). Como já mencionado acima, este estimador é muito usado devido suas propriedades.

A partir deste estimador é possível encontrar as outras funções e realizar as inferências necessárias para o prosseguimento de estudos. Outros autores a partir do estimador de Kaplan-Meier colaboraram com o desenvolvimento de técnicas não-paramétricas. Pode-se citar o estimador de Nelson-Aalen, proposto por Nelson (1972) e retomado por Aalen (1978) onde foram provadas suas propriedades assintóticas. Este estimador faz uso da função de risco acumulada, sendo expresso por:

$$\hat{\Lambda}(t) = \sum_{j:t_j<t} \left(\frac{d_j}{n_j} \right), \tag{14}$$

em que d_j e n_j são definidos como o estimador de Kaplan e Meier (1958a). Assim, um estimador para a função de sobrevivência, pode ser obtido através de Nelson-Aalen, substituindo (14) em (10), ou seja:

$$\hat{S}(t) = \exp\left\{ -\hat{\Lambda}(t) \right\}. \tag{15}$$

2.2 ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA

Para a estimação de parâmetros na estatística frequentista, existem diversas ferramentas para realizar tais estimações, sendo um dos mais conhecidos o método dos mínimos quadrados, geralmente utilizado em análise de regressão linear, porém tais métodos não abrangem a informação de censura, sendo inviáveis na prática quando trata-se informações que a possuem. Logo o método de estimação via EMV (Estimação de Máxima Verossimilhança) surge como uma opção de aplicabilidade em grandes amostras.

A ideia da EMV é a partir dos resultados observados, encontrar no meio de inúmeras distribuições (em relação aos parâmetros do modelo já escolhido) aquela que conforme os valores de seus parâmetros possa ter gerado tal amostra, um exemplo é se for escolhida uma distribuição com dois parâmetros, dentro das inúmeras observações dos valores dos parâmetros o EMV escolherá qual o melhor par que explique a amostra observada. Para isto é utilizado a

Função de Verossimilhança, denotada por:

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta) \tag{16}$$

em que θ é um vetor de parâmetros e $f(t_i)$ é a função densidade de probabilidade. Assim, o objetivo é encontrar qual o resultado que maximize tal função.

Para banco de dados que possuem observações censuradas, a função (16) é modificada, incluindo as informações de censura. Os dados envolvendo Análise de Sobrevivência são divididos em duas partes: os tempos censurados e os tempos não censurados, sendo o comportamento dos tempos exatamente observados representados pela função densidade de probabilidade $f(t)$ e os censurados representados pela função de sobrevivência $S(t)$, já que o tempo de falha é maior que as observações censuradas. Sendo as r primeiras observações de um banco de dados não censuradas e as $n-r$ observações censuradas, é possível escrever a função (16) como:

$$L(\theta) = \prod_{i=1}^{r} f(t_i;\theta) \prod_{i=r+1}^{n} S(t_i;\theta) \tag{17}$$

Considerando um indicador de censura, denotado por δ, é possível reescrever (17) como:

$$L(\theta) = \prod_{i=1}^{r} [f(t_i;\theta)]^{\delta_i} [S(t_i;\theta)]^{1-\delta} \tag{18}$$

em que δ_i é um parâmetro binário, assumindo valor 0 para as observações censuradas e 1 para as observações não censuradas.

Para encontrar os EMV dos parâmetros dos modelos deve-se maximizar a função (18) em relação ao vetor de parâmetros θ, maximizar (18) ou sua função logarítmica que retornam os mesmos resultados. Na prática é maximizado a função log-verossimilhança. Neste trabalho, a título de notação, o termo log refere-se ao logaritmo neperiano. Logo, a equação será dada abaixo:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0 \tag{19}$$

Geralmente resulta em um sistema de equações que não possuem soluções triviais. Assim utiliza-se de métodos numéricos iterativos para a resolução de tais sistemas. O método de Newton-Raphson é um dos algoritmos mais utilizados na resolução deste tipo de problema.

O algoritmo de Newton-Raphson baseia-se na expansão de $U(\hat{\theta})$ em torno de um ponto inicial $\theta_{(0)}$ através de séries de Taylor, onde tem-se na primeira iteração,

$$U(\hat{\theta}_{(1)}) = U(\theta_{(0)}) + U'(\theta_{(0)})(\hat{\theta}_{(1)} - \theta_{(0)}). \tag{20}$$

onde $U'(\theta_{(0)}) = \frac{\partial^2 \log L(\theta)}{\partial^2 \theta}\Big|_{\theta=\hat{\theta}_{(0)}}$, pode-se repetir este processo n vezes, até o erro ser o menor esperado. Um critério de convergência pode ser dado quando,

$$\frac{|\hat{\theta}_{(n)}|}{|\hat{\theta}_{(n+1)}|} < \varepsilon, \tag{21}$$

em que ε é tão pequeno quanto quanto desejável (ARENALES; DAREZZO, 2008).

2.2.1 INTERVALOS DE CONFIANÇA

Ao trabalhar com estimação via EMV, uma importante propriedade é que este estimador é assintótico, isto é, dentro de certas condições, a distribuição do vetor estimado $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ é Normal multivariada de média θ e matriz de variância-covariância $Var(\hat{\theta})$, de modo que:

$$\hat{\theta} \sim N_K\left(\theta, Var(\hat{\theta})\right), \tag{22}$$

onde k é a dimensão do vetor $\hat{\theta}$.

Uma segunda propriedade muito útil é a seguinte:

$$Var(\hat{\theta}) \approx -\left[E(F(\theta))\right]^{-1}, \tag{23}$$

isto é, a matriz de variância-covariância é aproximadamente (o $\approx$ neste caso tem o significado de "aproximado") o negativo da inversa esperança da matriz de Fisher (F), que é obtida através da derivada segunda do logaritmo do valor da função de verossimilhança de θ. Em algumas situações em que fica inviável o cálculo da esperança, utiliza-se somente $-\left[F(\theta)\right]^{-1}$, onde esta matriz é um estimador consistente de (23).

Os elementos da diagonal principal de (23) são as variâncias dos estimadores e os outros elementos são as covariâncias entre eles.

Na construção de intervalos de confiança deve-se estimar o erro-padrão, isto é, $[Var(\hat{\theta})]^{1/2}$.

Nos casos em que θ é somente um escalar, um intervalo de confiança com $(1-\alpha)100\%$ de confiança para θ é dado por:

$$\hat{\theta} \pm Z_{\frac{\alpha}{2}} \sqrt{\hat{Var}(\hat{\theta})}, \tag{24}$$

onde $Z_{\frac{\alpha}{2}}$ é um valor tabelado. Caso θ seja multivariado, os intervalos de confiança são construídos separadamente para cada parâmetro de θ.

2.2.2 TESTES DE HIPÓTESES

Quando é feita uma conjectura a respeito de um parâmetro θ desconhecido com distribuição F_θ, chamamos de hipótese estatística. Em estatística clássica, assume-se duas hipóteses, a H_0, ou Hipótese Nula, onde é a hipótese que assume-se como verdadeira na construção do teste, é a teoria, efeito, causa ou relação que busca-se testar e H_A, ou Hipótese Alternativa, assim é:

H_0: $\theta \in \Theta_0$ é a hipótese nula, em que $\Phi_0 \subset \Theta$, e

H_A: $\theta \in \Theta_0^C$ é a hipótese alternativa, em que Θ_0^C é o complemento de Θ_0,

em que Θ é um espaço paramétrico qualquer.

Outro conceito fundamental ao realizar teste de hipóteses são os possíveis erros que podem ocorrer, sendo:

Erro tipo I: a probabilidade de rejeitar H_0, quando H_0 é verdadeira, isto é, P(rejeitar $H_0|H_0$ é verdadeira)

Erro tipo II: a probabilidade de não rejeita H_0, quando H_A é verdadeira, isto é, P(não rejeitar $H_0|H_A$ é verdadeira).

2.3 MODELOS PARAMÉTRICOS

A seguir são apresentados alguns modelos paraméricos mais utilizados na área de Análise de Sobrevivência.

2.3.1 DISTRIBUIÇÃO EXPONENCIAL

Uma das distribuições mais simples, no sentido matemático, é a distribuição exponencial, já que ela possui um único um parâmetro e sua função de risco $\lambda(t)$ é caracterizada por uma função constante.

A função de densidade de probabilidade para a variável aleatória tempo de falha T é dada por:

$$f(t) = \frac{1}{\alpha}\exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, \tag{25}$$

em que $t \geq 0$ e o parâmetro $\alpha > 0$ é o tempo médio de vida, e ainda, α segue a mesma unidade do tempo de falha.

Com isto, pode-se apresentar as funções de sobrevivência e risco da distribuição exponencial, encontradas a partir das relações apresentadas no início deste capítulo, dadas respectivamente por:

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)\right\} \tag{26}$$

$$h(t) = \frac{1}{\alpha} \tag{27}$$

Observar que a função de risco (27) da distribuição exponencial é constante. Logo, observações novas ou velhas que ainda não falharam possuem o mesmo risco de falha.

2.3.2 DISTRIBUIÇÃO DE WEIBULL

Uma das distribuições usualmente utilizadas na área de análise de sobrevivência é a distribuição de Weibull, o fato que faz desta distribuição ser usualmente utilizada é que sua função de risco pode assumir diversas formas, sendo todas monótonas, isto é, crescente, decrescente ou constante. Sua função de densidade de probabilidade para uma variável aleatória T é dada por:

$$f(t) = \frac{\beta}{\mu^{\beta}} t^{\beta-1} \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right],\ t \geq 0, \tag{29}$$

em que $\beta > 0$ e $\mu > 0$ são os parâmetros de forma e escala, respectivamente. O parâmetro μ tem a mesma unidade do tempo de falha e o parâmetro β não possui unidade.

Utilizando das equivalências do início do capítulo chega-se as funções de sobrevivência e risco, respectivamente, apresentadas abaixo:

$$S(t) = \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right], \tag{30}$$

$$h(t) = \frac{\beta}{\mu^{\beta}} t^{\beta-1} \tag{31}$$

Observar que se $\beta = 1$ a expressão (29) se torna a função densidade de probabilidade da distribuição exponencial, fazendo com que a distribuição exponencial seja um caso particular da distribuição de Weibull.

Em alguns casos ao analisar dados de tempo de vida é mais prático utilizar o logaritmo, devido a suas propriedades e a não deformidade das análises, quando utiliza-se o logaritmo na distribuição de Weibull, chega-se a um caso que é chamado de distribuição do valor extremo ou de Gumbel. Assim, se a variável aleatória T tem distribuição de Weibull com a $f(t)$ dada por

(29), então uma variável $Y = \log(T)$, possui a seguinte função de densidade de probabilidade:

$$f(y) = \frac{1}{\sigma} \exp\left\{ \left(\frac{y-\alpha}{\sigma} \right) - \exp\left\{ \frac{y-\alpha}{\sigma} \right\} \right\}, \tag{33}$$

em que os parâmetros de locação e escala, σ e μ, respectivamente, apresentam as seguintes relações de igualdade com a distribuição de Weibull: $\beta = \frac{1}{\sigma}$ e $\mu = \exp\{\alpha\}$. A distribuição de Gumbel aparece durante a modelagem das informações, informações adicionais sobre esta distribuição estará na seção 2.6.3.

2.3.3 DISTRIBUIÇÃO LOG-NORMAL

Assim como a distribuição de Gumbel e Weibull possuem similaridades, a distribuição log-normal e a distribuição normal possuem similaridades, sendo que as relações existentes da distribuição normal são válidas para a distribuição log-normal, desde que se utilize o logaritmo do parâmetro desejado, ou seja, ainda continua-se com média μ e desvio-padrão σ, porém considerando os logaritmos dos resultados.

A função de densidade de probabilidade da distribuição log-normal, sendo T uma variável aleatória, é dada por:

$$f(t) = \frac{1}{\sqrt{2\pi} t \sigma} \exp\left\{ -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, \; t > 0 \tag{34}$$

As funções de sobrevivência e de falha não apresentam forma explícita, sendo representadas por:

$$S(t) = \Phi\left(\frac{-\log(t) + \mu}{\sigma} \right) \tag{35}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{36}$$

$$\tag{37}$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão (média 0 e variância 1).

2.3.4 DISTRIBUIÇÃO LOG-LOGÍSTICA

Outra distribuição muito utilizada quando analisa-se o tempo de falha de algo, é a distribuição log-logística, sendo geralmente utilizada como alternativa a distribuição de Wei-

bull. Para uma variável aleatória T, tem-se como função densidade de probabilidade:

$$f(t) = \frac{\gamma}{\alpha^{\gamma}} t^{\gamma-1} \left(1 + \left(\frac{t}{\alpha}\right)^{\gamma}\right)^{-2}, \; t > 0, \tag{39}$$

em que $\alpha > 0$ e $\gamma > 0$ são os parâmetros de forma e escala, respectivamente.

As funções de sobrevivência e de risco, são dadas respectivamente por:

$$S(t) = \frac{1}{1 + (\frac{t}{\alpha})^{\gamma}} \tag{40}$$

$$h(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^{\gamma}]} \tag{41}$$

Para dados de tempo, usualmente trabalha-se com a forma logarítmica, conforme ocorre em distribuições já apresentadas, logo, utilizando o mesmo parâmetro T já apresentado acima, tem-se o seu logaritmo como $Y = \log(T)$ e a função de densidade de probabilidade, função de sobrevivência e função de risco são apresentadas, respectivamente por:

$$f(y) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \left(1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right)^{-2}; \tag{43}$$

$$S(y) = \frac{1}{1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}}; \tag{44}$$

$$h(y) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \left(1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right)^{-1}. \tag{45}$$

2.3.5 DISTRIBUIÇÃO BURR XII

Uma distribuição mais flexível em relação as distribuições apresentadas até então é proposta por Burr (1942). Esta distribuição consegue assumir diversos casos particulares, englobando funções como Weibull, gamma, exponencial, log-normal entre outras, fazendo com que seja muito flexível no ponto de vista da modelagem, assumindo diversas formas.

A função densidade de probabilidade com três parâmetros, a função de sobrevivência

e a função de risco, são dadas, respectivamente, por:

$$f(t) = \frac{\alpha}{\mu^{\alpha}} t^{\alpha-1} \left[1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}\right]^{-\left(1+\frac{1}{\lambda}\right)}; \quad (47)$$

$$S(t) = \left[1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}\right]^{-\frac{1}{\lambda}}; \quad (48)$$

$$h(t) = \frac{\alpha\left(\frac{1}{\mu}\right)^{\alpha} t^{\alpha-1}}{1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}}. \quad (49)$$

em que $\mu > 0$ é o parâmetro de locação; $\alpha > 0$ e $\lambda > 0$ são parâmetros de forma. E que para $\alpha \to 0^{+}$ tem-se a distribuição Weibull como caso particular. A função de risco para a distribuição Burr XII é decrescente se $\alpha \leq 1$ e unimodal com moda em $t = \frac{(\alpha-1)^{1/\alpha}}{\mu^{-1}\lambda^{1/\alpha}}$ quando $\alpha > 1$.

2.4 MODELOS DE MISTURA DE LONGA DURAÇÃO

Em uma análise preliminar do banco de dados, verifica-se que uma porção considerada de indivíduos naõ enfrenta o evento de interesse, assim, a utilização de modelos de mistura de longa duração é uma alternativa para uma inferência mais precisa.

Em análise de sobrevivência, um modelo de mistura de longa duração, também conhecido como modelo de fração de cura, assume que a população em estudo é uma mistura de indivíduos suscetíveis a um evento de interesse, e indivíduos não suscetíveis, em que nunca é observado o evento de interesse. Esses indivíduos não estão em risco com respeito ao evento de interesse e são considerados imunes, não suscetíveis ou curados (MALLER; ZHOU, 1996), dependendo do contexto do estudo que geram os dados.

Diferentes metodologias, paramétricas e não paramétricas, podem ser consideradas para modelar a proporção de imunes. Nesse sentido, vários autores podem ser citados, como por exemplo, Boag (1949), Berkson e Gage (1952), Haybittle (1965), Farewell (1982, 1986), Meeker (1987), Dunsmuir et al. (1989), Gamel et al. (1990), Ghitany e Maller (1992), Taylor (1995), Copas e Heydari (1997), Ng e McLachlan (1998), Angelis et al. (1999), Peng e Dear (2000), Sy e Taylor (2000) e Yu et al. (2004). Wienke et al. (2006) propõem um modelo de fração de cura para dados de sobrevivência bivariados.

De acordo com Maller e Zhou (1996), em um modelo de fração de cura assume-se que uma certa fração p de indivíduos na população é curada ou nunca experimenta o evento

de interesse (são imunes). Logo $1-p$ é a fração de indivíduos não curados. A função de sobrevivência, nesse caso, pode ser escrita considerando a seguinte mistura (BERKSON; GAGE, 1952),

$$S(t) = p + (1-p)S_0(t), \tag{51}$$

em que $p \in (0,1)$ é o parâmetro de mistura (proporção de imunes) e $S_0(t)$ é a função de sobrevivência basal para a população de indivíduos não curados (indivíduos suscetíveis). Considerando uma amostra aleatória de tempos de vida (t_i, δ_i), $i = 1, \ldots, n$, a contribuição do *i-ésimo* indivíduo para a função de verossimilhança é dada por,

$$L_i = [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}, \tag{52}$$

em que δ_i é a variável indicadora de censura, ou seja, $\delta_i = 1$ quando o tempo de sobrevivência é exatamente observado e $\delta_i = 0$ quando o tempo é censurado (não observado) para o *i-ésimo* indivíduo.

A partir da função de sobrevivência definida em (51), é possível obter a função densidade de probabilidade, utilizando o resultado $f(t_i) = -\frac{d}{dt}S(t_i)$, dada por,

$$f(t_i) = (1-p) f_0(t_i), \tag{53}$$

em que $f_0(t_i)$ é a função densidade de probabilidade para os indivíduos suscetíveis. Substituindo a função densidade (53) e a função de sobrevivência (51) na função de verossimilhança (52) obtêm-se a seguinte função de verossimilhança para o modelo de mistura de longa duração,

$$L_i = [(1-p) f_0(t_i)]^{\delta_i} [p + (1-p) S_0(t_i)]^{1-\delta_i}. \tag{54}$$

Portanto a função log-verossimilhança considerando todas as observações é dada por,

$$l = r \log(1-p) + \sum_{i=1}^{n} \delta_i \log f_0(t_i) + \sum_{i=1}^{n} (1-\delta_i) \log[p + (1-p) S_0(t_i)], \tag{55}$$

em que, $r = \sum_{i=1}^{n} \delta_i$ é o número de observações não censuradas.

Usualmente, assume-se que a função de sobrevivência $S_0(t)$, em (51), é a função de sobrevivência das distribuições exponencial ou Weibull. Peng et al. (1998) considerou a função de sobrevivência da distribuição Fisher-Snedecor generalizada. A distribuição Fisher-Snedecor generalizada é um supermodelo que inclui os modelos mais usuais de sobrevivência como casos particulares, como por exemplo, as distribuições exponencial, Weibull e log-normal. Yamaguchi (1992) considera a distribuição log-gama generalizada para modelos de fração de cura no contexto de regressão com tempos de falha acelerados. A distribuição de Gompertz é conside-

rada em Gieser et al. (1998), enquanto que as distribuições Weibull exponenciada e exponencial exponenciada são consideradas, respectivamente, por Cancho e Bolfarine (2001) e Kannan et al. (2010). Um modelo de fração de cura utilizando a distribuição Conway-Maxwell Poisson é proposto por Rodrigues et al. (2009) como alternativa ao modelo discutido por Yin e Ibrahim (2005). Shao e Zhou (2004) propõem um modelo de mistura de longa duração considerando a distribuição Burr XII.

2.4.1 DISTRIBUIÇÃO BURR XII NA PRESENÇA DE FRAÇÃO DE CURA

Considere (48), o modelo Burr XII na presença de fração de cura tem função densidade de probabilidade, função distribuição e função de sobrevivência definidas, respectivamente, por,

$$f(t \mid \theta) = (1-p)\frac{\alpha}{\mu^{\alpha}}t^{\alpha-1}\left[1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}\right]^{-\left(1+\frac{1}{\lambda}\right)}; \tag{56}$$

$$F(t \mid \theta) = (1-p)\left\{1-\left[1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}\right]^{-\frac{1}{\lambda}}\right\}; \tag{57}$$

$$S(t \mid \theta) = p+(1-p)\left[1+\lambda\left(\frac{t}{\mu}\right)^{\alpha}\right]^{-\frac{1}{\lambda}}, \tag{58}$$

em que $\theta = (\mu, \alpha, \lambda, p)$, μ é o parâmetro de escala, α e λ são os parâmetros de forma e p é a proporção de indivíduos imunes ou não suscetíveis. Supor os dados na forma (t_i, δ_i), $i = 1, \ldots, n$, em que $\delta_i = 1$ se t_i não é censurado e $\delta_i = 0$ caso contrário e que $f(t_i)$ é dado por (56). Assumindo censuras à direita para os tempos de sobrevivência, a função de verossimilhança é dada por,

$$L(\theta \mid \mathbf{t}, \delta) = L_1(\theta \mid \mathbf{t}, \delta) \times L_2(\theta \mid \mathbf{t}, \delta). \tag{60}$$

Dessa forma, as funções log-verossimilhança, $l_j(\theta \mid \mathbf{t}, \delta) = \log\left[L_j(\theta \mid \mathbf{t}, \delta)\right]$, $j = 1, 2$, são dadas, respectivamente, por,

$$l_1(\theta \mid \mathbf{t}, \delta) = r\log(1-p) + r\log(\alpha) - r\alpha\log(\mu) + (\alpha-1)\tilde{t} - \left(1+\frac{1}{\lambda}\right)\sum_{i=1}^{n}\delta_i\log(A_i); \tag{61}$$

$$l_2(\theta \mid \mathbf{t}, \delta) = \sum_{i=1}^{n}(1-\delta_i)\log\left\{p+(1-p)A_i^{-\frac{1}{\lambda}}\right\}, \tag{62}$$

em que $r = \sum_{i=1}^{n}\delta_i$, $\tilde{t} = \sum_{i=1}^{n}\delta_i\log(t_i)$, $A_i = 1+B_i$ e $B_i = \lambda\left(\frac{t_i}{\mu}\right)^{\alpha}$.

Sejam os tempos de sobrevivência observados, (t_i, δ_i), $i = 1, \ldots, n$, definindo $l(\theta \mid \mathbf{t}, \delta) = \log L(\theta \mid \mathbf{t}, \delta)$; então, os estimadores de máxima verossimilhança para $\theta = (\mu, \alpha, \lambda, p)$, deno-

tados por $\widehat{\theta} = \left(\widehat{\mu}, \widehat{\alpha}, \widehat{\lambda}, \widehat{p}\right)$, são obtidos resolvendo, por algum método numérico, o seguinte sistema de equações,

$$\begin{cases} \frac{\partial}{\partial\mu} l(\theta \mid \mathbf{t}, \delta) = -\frac{r\alpha}{\mu} + \frac{\alpha\left(1+\frac{1}{\lambda}\right)}{\mu} \sum_{i=1}^{n} \frac{\delta_i B_i}{A_i^{\alpha}} + \frac{(1-p)\alpha}{\lambda\mu} \sum_{i=1}^{n} \frac{A_i^{-\left(1+\frac{1}{\lambda}\right)} B_i}{p+(1-p)A_i^{-\frac{1}{\lambda}}} = 0 \\ \frac{\partial}{\partial\alpha} l(\theta \mid \mathbf{t}, \delta) = \frac{r}{\alpha} - r\log(\mu) + \tilde{t} - \left(1+\frac{1}{\lambda}\right) \sum_{i=1}^{n} \frac{\delta_i B_i \log\left(\frac{t_i}{\mu}\right)}{A_i} - \frac{(1-p)}{\lambda} \sum_{i=1}^{n} \frac{A_i^{-\left(1+\frac{1}{\lambda}\right)} B_i \log\left(\frac{t_i}{\mu}\right)}{p+(1-p)A_i^{-\frac{1}{\lambda}}} = 0 \\ \frac{\partial}{\partial\lambda} l(\theta \mid \mathbf{t}, \delta) = \frac{1}{\lambda^2} \sum_{i=1}^{n} \delta_i \log(A_i) - \frac{\lambda+1}{\lambda^2} \sum_{i=1}^{n} \frac{\delta_i B_i}{A_i} + \frac{(1-p)}{\lambda^2} \sum_{i=1}^{n} \frac{A_i^{-\frac{1}{\lambda}} \left[\log(A_i) - B_i A_i^{-1}\right]}{p+(1-p)A_i^{-\frac{1}{\lambda}}} = 0 \\ \frac{\partial}{\partial p} l(\theta \mid \mathbf{t}, \delta) = -\frac{r}{1-p} + \sum_{i=1}^{n} \frac{1-A_i^{-\frac{1}{\lambda}}}{p+(1-p)A_i^{-\frac{1}{\lambda}}} = 0 \end{cases} . \quad (64)$$

Intervalos de confiança e testes de hipóteses de interesse podem ser obtidos usando métodos assintóticos, como a normalidade assintótica dos estimadores de máxima verossimilhança (EMV) ou usando testes de razão de verossimilhança (LAWLESS, 1982).

2.4.2 DISTRIBUIÇÃO DE WEIBULL NA PRESENÇA DE FRAÇÃO DE CURA

Para facilitar os cálculos é considerado a distribuição de Weibull dada em (29) com a parametrização $\lambda = \frac{1}{\mu^{\beta}}$. Assumindo o modelo de mistura (51), a função log-verossimilhança para β, λ e p (ver (55)), é dada por,

$$l(\theta \mid \mathbf{t}, \delta) = r\log(1-p) + r\log(\beta) + r\log(\lambda) + (\beta-1)v - \lambda A_1(\beta) + A_2(\beta, \lambda, p), \quad (65)$$

em que $\theta = (\beta, \lambda, p)$, $r = \sum_{i=1}^{n} \delta_i$, $v = \sum_{i=1}^{n} \delta_i \log(t_i)$, $A_1(\beta) = \sum_{i=1}^{n} \delta_i t_i^{\beta}$ e $A_2(\beta, \lambda, p) = \sum_{i=1}^{n} (1-\delta_i)$ $\log\left[p + (1-p)\exp\left(-\lambda t_i^{\beta}\right)\right]$.

As primeiras derivadas de $l(\theta \mid \mathbf{t}, \delta)$ em relação a β, λ e p, são dadas, respectivamente, por,

$$\begin{aligned} \frac{\partial}{\partial\beta} l(\theta \mid \mathbf{t}, \delta) &= \frac{r}{\beta} + v - \lambda \frac{\partial}{\partial\beta} A_1(\beta) + \frac{\partial}{\partial\beta} A_2(\beta, \lambda, p), \\ \frac{\partial}{\partial\lambda} l(\theta \mid \mathbf{t}, \delta) &= \frac{r}{\lambda} - A_1(\beta) + \frac{\partial}{\partial\lambda} A_2(\beta, \lambda, p), \\ \frac{\partial}{\partial p} l(\theta \mid \mathbf{t}, \delta) &= -\frac{r}{1-p} + \frac{\partial}{\partial p} A_2(\beta, \lambda, p), \end{aligned} \quad (66)$$

em que,

$$\begin{aligned}
&\frac{\partial}{\partial\beta}A_1(\beta) = \sum_{i=1}^{n} \delta_i t_i^{\beta}\log(t_i);\\
&\frac{\partial}{\partial\beta}A_2(\beta,\lambda,p) = -\lambda(1-p)\sum_{i=1}^{n}\frac{(1-\delta_i)t_i^{\beta}\exp\left(-\lambda t_i^{\beta}\right)\log(t_i)}{p+(1-p)\exp\left(-\lambda t_i^{\beta}\right)};\\
&\frac{\partial}{\partial\lambda}A_2(\beta,\lambda,p) = -(1-p)\sum_{i=1}^{n}\frac{(1-\delta_i)t_i^{\beta}\exp\left(-\lambda t_i^{\beta}\right)}{p+(1-p)\exp\left(-\lambda t_i^{\beta}\right)};\\
&\frac{\partial}{\partial p}A_2(\beta,\lambda,p) = \sum_{i=1}^{n}\frac{(1-\delta_i)\left[1-\exp\left(-\lambda t_i^{\beta}\right)\right]}{p+(1-p)\exp\left(-\lambda t_i^{\beta}\right)}.
\end{aligned} \tag{68}$$

Igualando a zero as equações obtidas em (66) e resolvendo o sistema resultante por método numérico, tem-se os estimadores de máxima verossimilhança para β, λ e p, denotados por $\widehat{\beta}$, $\widehat{\lambda}$ e $\widehat{p}$. Os $100\times(1-\psi)\,\%$ intervalos de confiança para β, λ e p podem ser obtidos a partir da normalidade assintótica dos estimadores de máxima verossimilhança.

2.5 MÉTODOS BAYESIANOS

A utilização de métodos Bayesianos é uma alternativa na análise de dados. Estes métodos são baseados no Teorema de Bayes.

Seja Ω um espaço amostral em que $A_1, A_2, \ldots, A_n$ sejam eventos mutuamente exclusivos, que formem uma partição do espaço Ω e que a intersecção de dois eventos distintos seja vazia. Logo, a probabilidade da união de todos estes eventos é igual a 1. Assim, para qualquer outro evento $B \subset \Omega$ tem-se o Teorema de Bayes dado por:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{k} P(B|A_j)P(A_j)} \tag{69}$$

para todo $i = 1, ..., k$ (ACHCAR et al., 2012). Assim, antes do conhecimento de qualquer informação sobre o evento A_i, atribui-se uma probabilidade a priori para A_i, que é dada por $P(A_i)$. Essa probabilidade é atualizada a partir da ocorrência do evento B, chamada de probabilidade condicional, que é dada pelo Teorema de Bayes (69), isto é, $P(A_i|B)$ (a probabilidade de A_i dado B).

Pode-se escrever o Teorema de Bayes como:

$$P(\theta|x) \propto L(\theta;x)P(\theta),$$

em outras palavras,

distribuição a posteriori $\propto$ *verossimilhança* $\times$ *distribuição a priori*, onde $\propto$ significa

"proporcional" e $\times$ é o simbolo de multiplicação.

2.5.1 DISTRIBUIÇÕES A PRIORI

Para utilizar de informações a priori é necessário a especificação de distribuições de probabilidade para os parâmetros de interesse. Estas distribuições devem representar o conhecimento que se tem sobre os parâmetros antes da realização do estudo. Os parâmetros da distribuição a priori, chamados de *hiperparâmetros*, são definidos de forma à contemplar as informações a priori, caso existam.

Fundamentalmente, uma distribuição a priori deve respeitar os limites do espaço paramétrico, conduzir a uma posteriori integrável ou própria e refletir o conhecimento do especialista, quando disponível.

As distribuições a priori para um parâmetro qualquer podem ser obtidas através de diversas formas, sendo as mais usuais:

- Distribuições a priori definidas no domínio de variação do parâmetro de interesse, por exemplo, ao escolher uma distribuição normal para um parâmetro que possui uma variação em todo intervalo real;
- Distribuições a priori elaboradas através de informações fornecidas por especialistas, por exemplo, um químico fornecer informações sobre determinada droga para um estudo de fármacos;
- Distribuições a priori construídas através de métodos de elicitações, por exemplo, métodos estruturais através de histogramas, onde através de uma partição do espaço amostral paramétrico, um especialista irá determinar as probabilidades para intervalo, de modo que após esta etapa, construa-se um histograma e escolhe-se uma família de distribuições que melhor modele este histograma. Outro exemplo seria, o método preditivo de elicitação, onde o especialista poderá fornecer informações nas observações e não no parâmetro;
- Distribuições a priori baseadas em métodos Bayesianos empíricos em dados ou experimentos prévios;
- Distribuições a priori não informativas, quando se possui total ignorância sobre os parâmetros de interesse.

As distribuições a priori não informativas são utilizadas por diversos motivos, sendo os mais comuns, a ignorância sobre os parâmetros de interesse, permitir comparações com a

inferência clássica e comparar resultados obtidos através de uma priori subjetiva.

Uma primeira ideia que se pode ter em relação a distribuições a priori não informativas é que todos os possíveis valores que o parâmetro em questão pode assumir são igualmente prováveis, isto é, utilizando uma distribuição a priori uniforme [1], assim fazendo $P(\theta) \propto k$, (proporcional a uma constante) onde θ varia em um subconjunto da reta, significando que nenhum valor particular possui preferência (BAYES, 1763). A utilização deste método sem um conhecimento prévio pode ocasionar erros, como no caso se a variação de θ for ilimitada, levando a um caso em que a distribuição a priori é imprópria (não integrável).

2.5.2 INFERÊNCIA BAYESIANA

Métodos Bayesianos podem ser considerados na análise dos dados de sobrevivência com fração de cura. A análise Bayesiana é baseada em métodos de simulação de MCMC (Monte Carlo via Cadeia de Markov), para gerar amostras da distribuição a posteriori de interesse. Dessa forma, uma análise Bayesiana para o modelo de mistura de longa duração considerando as distribuições Burr XII e Weibull é realizada.

Amostras da distribuição conjunta a posteriori de interesse são simuladas utilizando métodos MCMC, como o popular algoritmo de Gibbs (GELFAND; SMITH, 1990; CASELLA; GEORGE, 1992) e o algoritmo Metropolis-Hastings (CHIB; GREENBERG, 1995) (ver. Apêndice A).

Serão considerados valores para os hiperparâmetros nos quais não seja fornecida nenhuma informação a respeito do evento em estudo. Assim, os valores escolhidos para os hiperparâmetros são valores que ocasionam em uma elevada variância.

2.5.2.1 ANÁLISE BAYESIANA PARA A DISTRIBUIÇÃO BURR XII

Para a análise Bayesiana considerando a distribuição Burr XII na presença de fração de cura, assume-se a distribuição a priori não informativa uniforme $U(0;1)$ definida no intervalo $(0,1)$ para a probabilidade de cura p e para o parâmetro de forma λ. E assume-se a distribuição a priori não informativa uniforme[1] $U(0;1000)$ definida no intervalo $(0,1000)$ para os parâmetros de locação μ e forma α. Assume-se independência a priori entre p, μ, α e λ.

Considerando a distribuição Burr XII sem a presença de fração de cura, assume-se distribuição a priori não informativa gama [2] $Gama(0,001;0,001)$ para os parâmetros de locação

[1] $f(x) = \frac{1}{b-a}$, onde $-\infty < a < b < \infty$.

[2] $f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ onde Γ é a função Gama dada por $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ onde $\alpha, \beta > 0$.

μ e forma α, e distribuição a priori não informativa gama [2] $Gama(1;1)$ para o parâmetro de forma λ. Assume-se independência a priori entre μ, α e λ.

2.5.2.2 ANÁLISE BAYESIANA PARA A DISTRIBUIÇÃO WEIBULL

Para a análise Bayesiana considerando a distribuição Weibull na presença de fração de cura, assume-se distribuição a priori não informativa uniforme[1] $U(0;1)$ definida no intervalo $(0,1)$ para a probabilidade de cura p, distribuição a priori não informativa gama[2] $Gama(0,001;0,001)$ para o parâmetro de escala μ e distribuição a priori não informativa gama[2] $Gama(0,01;0,01)$ para o parâmetro de forma β.Assume-se, independência a priori entre os parâmetros β, μ e p.

Quando é considerada a distribuição Weibull sem a presença de fração de cura, assume-se distribuição a priori não informativa uniforme[1] $U(0;20000)$ definida no intervalo $(0,20000)$ para o parâmetro de escala μ e distribuição a priori não informativa uniforme[1] $U(0;10000)$ definida no intervalo $(0,10000)$ para o parâmetro de forma β. Assume-se, também, independência a priori entre os parâmetros β e μ.

É importante observar que para os modelos introduzidos aqui, são utilizados alguns programas computacionais introduzidos na literatura, como por exemplo, a procedure `MCMC` (SAS, 2010a) do `software SAS`, que só requer a introdução da distribuição para os dados e as distribuições a priori para os parâmetros do modelo proposto. Assim, não são introduzidas as distribuições condicionais a posteriori necessárias para a geração de amostras da posteriori conjunta de interesse usando o amostrador de Gibbs ou o algoritmo de Metropolis-Hastings.

2.6 MODELOS DE REGRESSÃO

Ao trabalhar com inferência de dados, frequentemente são incorporadas informações que agregam valores aos dados em estudo, as covariáveis.

As informações das covariáveis podem ser expressas através de um vetor

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{ip}),$$

sobre o indivíduo i, onde os dados são formados de n observações na forma (t_i, δ_i, x_i).

A forma mais eficiente de verificar o efeito de tais covariáveis é utilizar um modelo de regressão apropriado para os dados (COLOSIMO; GIOLO, 2006).

Ao utilizar dados em análise de sobrevivência, são dispostas duas classes de modelos de regressão, os modelos paramétricos (modelos de tempo de vida acelerados ou modelos de

regressão locação-escala) e modelos semiparamétricos (ou modelos de regressão de Cox). A classe paramétrica é mais eficiente, mas não tão flexível quando comparada com a de Cox.

O modelo de regressão mais conhecido é o de regressão linear, neste modelo busca-se associar uma covariável a resposta através de um modelo linear, e o gráfico da covariável em relação a resposta, deverá apresentar evidências de uma relação linear, ou seja, a nuvem de pontos que é formada, deverá ser próxima a uma reta.

2.6.1 REGRESSÃO PARAMÉTRICA

Este modelo é representado por:

$$Y = \beta_0 + \beta_1 x + \varepsilon, \tag{70}$$

onde Y é a resposta, x é a covariável em questão, β_0 e β_1 são os parâmetros a ser estimados e ε é o erro aleatório, que neste caso, segue distribuição normal.

Ao analisar dados de análise de sobrevivência o modelo (70) geralmente não é utilizado, devido a assimetria dos dados em direção a maiores tempo de sobrevivência, diante desse problema, é proposto uma transformação de variável e a utilização do modelo determinístico:

$$\exp\{\beta_0 + \beta_1 x\}, \tag{71}$$

com a distribuição log-normal para o erro aleatório.

2.6.2 MODELO DE REGRESSÃO EXPONENCIAL

O modelo de regressão mais simples de utilização é o modelo exponencial, onde só é trabalhada uma única covariável. Ao combinar a equação determinística (71) e uma distribuição exponencial com a média unitária ($f(\varepsilon) = \exp\{\varepsilon\}$) para o erro padrão, obtem-se o modelo de regressão exponencial abaixo:

$$T = \exp\{\beta_0 + \beta_1 x\}\varepsilon, \tag{72}$$

onde este modelo é uma função de ligação logarítmica e a resposta com distribuição exponencial, onde ε possui uma distribuição assimétrica.

O modelo (72) pode ser linearizado, quando é considerado seu logaritmo, assim:

$$Y = \log(T) = \beta_0 + \beta_1 x + \nu, \tag{73}$$

onde ν é $\log(\varepsilon)$.

A diferença entre (73) e o modelo linear (70) é a distribuição que ν segue, onde em (73) é a distribuição do valor extremo padrão ($f(\nu) = \exp\{\nu - \exp\{\nu\}\}$), distribuição que é muito útil à análise de sobrevivência, já que caracteriza de forma satisfatória diversos dados.

A função de sobrevivência para Y condicional a x é dada por:

$$S(y|x) = \exp\left\{-\exp\{y - (\beta_0 + \beta_1 x)\}\right\}, \tag{74}$$

e para T condicional a x:

$$S(t|x) = \exp\left\{-\left(\frac{t}{\exp\{\beta_0 + \beta_1 x\}}\right)\right\}. \tag{75}$$

A estimação dos parâmetros $\theta = (\beta_0, \beta_1)$ é via EMV, diferentemente do método linear, onde é utilizado o método dos mínimos quadrados. No caso de dados que não possuem normalidade e presença de censuras, tal método é inviável.

A função de verossimilhança para o modelo (73) é dada por:

$$L(\theta) = \prod_{i=1}^{n} [f(y_i|x_i)]^{\delta_i} [S(y_i|x_i)]^{1-\delta}, \tag{76}$$

onde $y_i = \log(t_i)$.

Para modelos na forma de (72) a função de verossimilhança é a seguinte:

$$L(\theta) = \prod_{i=1}^{n} [f(t_i|x_i)]^{\delta_i} [S(t_i|x_i)]^{1-\delta} \tag{77}$$

O método de obtenção dos valores estimados de θ é o mesmo aplicado na estimação dos valores para o modelo geral, isto é, toma-se o logaritmo da função de verossimilhança, resolve o sistema resultante das derivadas parciais dos parâmetros de θ numericamente e assim, determina-se os valores de θ.

2.6.3 MODELO DE REGRESSÃO WEIBULL

Uma generalização do modelo de regressão exponencial é acrescentar um novo parâmetro de escala, isto é, em modelos lineares assumir que para os erros, a distribuição é normal com variância σ^2, no local da distribuição normal padronizada.

A forma generalizada com p covariáveis é dada por:

$$Y = \log\{T\} = x'\beta + \sigma\nu, \tag{78}$$

onde $x' = (1, x_1, \ldots, x_p)$ é o vetor de covariáveis, $\beta = (\beta_0, \ldots, \beta_p)$ e σ são os parâmetros desconhecidos e ν segue distribuição de valor extremo padrão, com densidade $f(\nu) = \exp\{\nu - e^{\nu}\}$.

Dessa forma, T tem uma distribuição Weibull, e ainda, $\log\{T\}$ segue distribuição de valor extremo com parâmetro de escala e locação dependendo das covariáveis.

A função de sobrevivência para Y condicional a x é apresentada por:

$$S(y|x) = \exp\left\{-\exp\left\{\frac{y - x'\beta}{\sigma}\right\}\right\}, \tag{79}$$

e a função de sobrevivência para T condicional a x é dada por:

$$S(t|x) = \exp\left\{-\left(\frac{t}{\exp\{x'\beta\}}\right)^{\frac{1}{\sigma}}\right\}. \tag{80}$$

A estimação dos parâmetros é via EMV novamente e sua função de verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n}\left(\frac{1}{\sigma}\exp\left\{\frac{y - x\beta}{\sigma} - \exp\left\{\frac{y - x\beta}{\sigma}\right\}\right\}\right)^{\delta_i}\left(\exp\left\{-\exp\left\{\frac{y - x\beta}{\sigma}\right\}\right\}\right)^{1-\delta_i} \tag{81}$$

o método de estimação dos parâmetros é análogo ao já explicitado acima.

O modelo (78) é conhecido também como **modelo de tempo de vida acelerado**, já que ao analisar as covariáveis envolvidas, estas podem acelerar ou retardar o tempo de vida.

As distribuições pertencentes a esta família possuem função densidade de probabilidade na forma de:

$$f(y, \beta, \sigma) = \frac{1}{\sigma} g\left(\frac{y - \beta}{\sigma}\right), -\infty < y < \infty, \tag{82}$$

e função de sobrevivência na forma de $G\left(\frac{y-\beta}{\sigma}\right)$ (PARANAIBA, 2012).

Quando a distribuição Burr XII é utilizada na forma de (78), esta passa a ser chamada de log-Burr XII, possuindo a função de sobrevivência de y dado t como:

$$S(y|t) = \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-k}, \tag{83}$$

e de y dado x por:

$$S(y|x) = \left[1 + \exp\left(\frac{y - x\beta}{\sigma}\right)\right]^{-k}. \tag{84}$$

2.6.4 INFERÊNCIA PARA MODELOS DE REGRESSÃO PARAMÉTRICOS

Os parâmetros são quantidades desconhecidas do modelo de regressão, desse modo, através de uma amostra aleatória, busca-se estimar tais parâmetros. A seguir são apresentadas algumas técnicas de estimação para tais parâmetros desconhecidos do modelo de regressão.

2.6.4.1 MÉTODO DE MÁXIMA VEROSSIMILHANÇA

A estimação utilizando o método de máxima verossimilhança é análogo ao explicitado acima, valendo as propriedades para estimação de intervalos de confiança e testes de hipótese, quando a amostra é grande o suficiente e possui certas condições de regularidade.

Quando é analisada a interpretação dos coeficientes estimados, uma proposta é através de tempos medianos (HOSMER; LEMESHOW, 1999). Assim, quando uma variável é binária e considerando a razão dos tempos medianos com $x = 1$ no numerador, se $\hat{\beta}$ é negativo (positivo) implica que indivíduos com $x = 1$ apresentam tempo mediano de sobrevivência reduzido (aumentado) em $[e^{\hat{\beta}} \times 100\%]$ relativamente aos indivíduos no outro grupo binário ($x = 0$), fixando as demais covariáveis, tal análise pode ser estendida a casos de covariáveis contínuas ou categóricas.

2.6.4.2 ANÁLISE BAYESIANA

O método Bayesiano, apresentado na seção 2.5, também é aplicado na estimação dos parâmetros dos modelos de regressão.

Utilizando das mesmas técnicas de estimação, resolução das distribuições via métodos de simulação de Monte Carlo via Cadeia de Markov e análise de convergência, são obtidas as distribuições a posteriori de interesse para os parâmetros desconhecidos.

2.6.4.3 ESTATÍSTICAS AIC E DIC

Critérios de informação também são utilizados para a escolha do melhor modelo sobre os parâmetros em estudo. Para maiores informações, ver apêndice B.

2.6.5 VARIÁVEIS QUALITATIVAS

A incorporação de dados qualitativos em estudos de regressão é realizada através da utilização de variáveis *dummy*, no qual são consideradas variáveis binárias que indicam a

presença ou não das informações qualitativas do estudo no modelo.

A utilização de tais variáveis torna o modelo de regressão muito mais flexível, incorporando os valores 0 (no caso de ausência de um atributo) e 1 (no caso de presença de um atributo).

Na prática se a variável qualitativa possui K categorias, são incluídas no modelo $K-1$ variáveis *dummies*, de forma que contenha informações sobre todas as categorias pertencentes à variável (MISSIO; JACOBI, 2007).

3 RESULTADOS

Para aplicar a metodologia proposta e verificar o desempenho dos modelos propostos é considerado um conjunto de dados composto por todos os indivíduos detidos por algum motivo no SECAT da comarca de Primeiro de Maio entre dezembro de 2009 até dezembro de 2015 (final do período de observação). Os dados de entrada e saída, tipo de delito cometido, motivo da prisão e tipo de benefício que o fez ser solto foram obtidos através do livro de presos do SECAT, que é preenchido manualmente. As informações adicionais como sexo, profissão, cor da cútis, escolaridade, data de nascimento e se o indivíduo possui ou não tatuagens no corpo foram obtidas através do Sistema de Registro Policial da Polícia Civil do Paraná. A variável dependente de interesse é o tempo em dias da soltura até o detento reincidir ao crime. Os crimes estão divididos em categorias, segundo o código penal Brasileiro BRASIL (1940) e podem ser observados no Apêndice C. Os benefícios adquiridos ao sair do SECAT de Primeiro de Maio são:

- Alvará de Soltura com ressalvas;
- Fiança;
- Transferências para outras unidades carcerárias;
- Regime aberto;
- Prisão domiciliar.

O conjunto de dados coletado é composto por 356 indivíduos que foram detidos no SECAT da comarca de Primeiro de Maio entre dezembro de 2009 à dezembro de 2015, além do tempo até a reincidência ao crime (em dias) e observou-se o seguinte comportamento, descrito nas Tabelas 1 a 8 .

A idade média do primeiro delito é de 32 anos, com um desvio padrão de 11 anos, sendo o indivíduo com a menor idade possuía 15 anos e o indivíduo com a maior idade quando foi preso a primeira vez na comarca de Primeiro de Maio é de 78 anos.

Tabela 1: Sexo dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Sexo	Frequência Relativa (%)	Frequência Absoluta
Masculino	85,98%	306
Feminino	13,46%	48
Não informado	0,56%	02

Tabela 2: Vínculo empregatício dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Profissão	Frequência Relativa (%)	Frequência Absoluta
Desempregado	4,50%	16
Empregados	26,40%	94
Não informado	69,10%	246

Tabela 3: Cor da cútis dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Cútis	Frequência Relativa (%)	Frequência Absoluta
Autodeclarada Branca	59,61%	214
Autodeclarada Parda	27,47%	97
Autodeclarada Negra	9%	34
Não informado	3,92%	11

Tabela 4: Escolaridade dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Escolaridade	Frequência Relativa (%)	Frequência Absoluta
Primeiro grau completo	12,92%	46
Segundo grau completo	8,14%	29
Terceiro grau completo	1,68%	06
Não alfabetizado	35,11%	125
Não informado	42,15%	150

Tabela 5: Tatuagem dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Possui tatuagens pelo corpo	Frequência Relativa (%)	Frequência Absoluta
Sim	13,19%	46
Não	86,81%	310

Tabela 6: Faixa etária da apreensão dos indivíduos privados de liberdade no SECAT de Primeiro de Maio - PR.

Faixa Etária	Frequência Relativa (%)
Menos de 18 anos	2%
Entre 18 e 24 anos de idade	28,60%
Entre 25 e 29 anos de idade	18,50%
Entre 30 e 34 anos de idade	14,60%
Entre 35 e 45 anos de idade	22,30%
Entre 46 e 60 anos de idade	11,90%
Entre 61 e 70 anos de idade	2%
Acima de 71 anos de idade	0,01%

Tabela 7: Benefício adquirido pelos indivíduos ao sair do SECAT de Primeiro de Maio - PR.

Benefício Adquirido	Frequência Relativa (%)	Frequência Absoluta
Alvará de soltura com ressalvas	60,95%	217
Pagamento de fiança	18,82%	67
Transferência	18,53%	66
Outros benefícios	1,70%	06

Tabela 8: Tipo de crime cometido para entrar no SECAT de Primeiro de Maio - PR.

Tipo de crime	Frequência Relativa (%)	Frequência Absoluta
Crime contra a pessoa	28,65%	102
Crime contra o patrimônio	20,50%	73
Legislação específica (Drogas)	17,41%	62
Crimes de trânsito	9,55%	34
Legislação específica (Desarmamento)	5,33%	19
Crime contra a dignidade sexual	3,93%	14
Crimes ambientais	3,93%	14
Legislação específica (ECA)	2,80%	10
Crime contra a administração pública	2,24%	08
Outras combinações de crimes	5,66%	20

Para a covariável "Tipo de Crime Cometido" foi admitido o crime de maior pena, baseando-se no código penal brasileiro, visto que a maioria dos detentos infringiram mais de um tipo de crime e ainda foi considerado os dados dos reincidentes somente da primeira passagem pelo SECAT da comarca de Primeiro de Maio - PR, já que alguns foram presos mais de uma vez.

O objetivo da análise desses dados é estudar o comportamento da variável tempo até a reincidência ao crime utilizando a distribuição de Weibull e Burr XII. Existem evidências que sugerem que uma grande proporção de detentos não voltam a reincidir ao crime após cumprir a pena, portanto um modelo de mistura de longa duração será proposto.

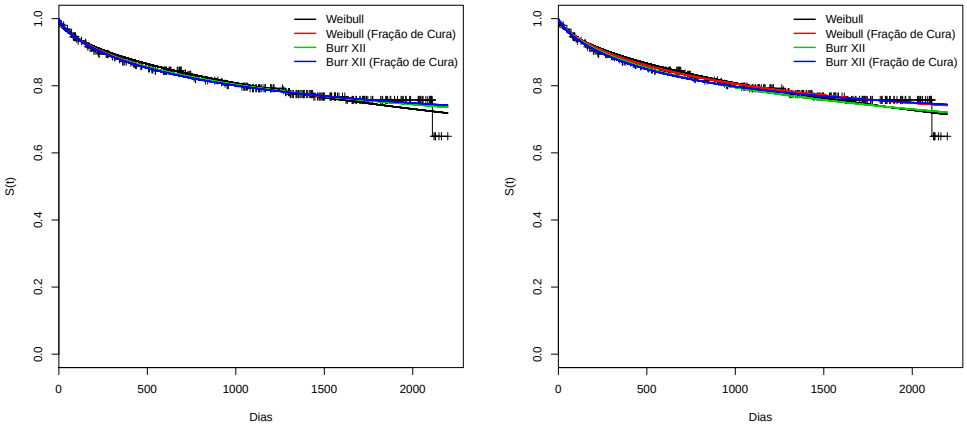
Considerando o objetivo de ajustar o tempo de reincidência ao crime aos modelos Weibull e Burr XII na presença ou não de fração de cura, os parâmetros de interesse foram estimados através das metodologias clássica (Estimadores de Máxima Verossimilhança) e Bayesiana.

Os estimadores de máxima verossimilhança foram obtidos utilizando o procedimento (procedure) `NLMIXED` do software `SAS` (SAS, 2010b), pelo algoritmo de Newton-Raphson. Para obter as estimativas Bayesianas foi utilizado o método MCMC disponível no software `SAS 9.3` na procedure `MCMC` (SAS, 2010a). Uma única cadeia é utilizada para ambos os modelos considerando 250.000 simulações para cada parâmetro com um *burn-in* de tamanho 15.000 para eliminar os possíveis efeitos dos valores iniciais da simulação. Os valores simulados foram selecionados de 150 em 150, para se ter amostras aproximadamente não correlacionadas,

no que resulta em uma amostra final de tamanho 2.000. Assume-se distribuições a priori não informativas para cada parâmetro dos modelos. Diagnósticos usuais de convergência observados na literatura estão avaliados na procedure `MCMC` do software `SAS`; nesse caso a indicação de convergência para todos os parâmetros foi observada.

Seja T a variável aleatória que representa o tempo até a reincidência do detento ao crime, as estimativas de máxima verossimilhança para os parâmetros dos modelos Weibull e Burr XII na presença ou não de fração de cura são apresentadas na Tabela 9 e as inferências considerando a análise Bayesiana são apresentadas na Tabela 10. Nas Tabelas 9 e 10 também é possível observar o AIC (*Akaike Information Criterion*) e as estimativas de Monte Carlo para o DIC (*Deviance Information Criterion*), utilizados como critério de discriminação de modelos. Menores valores de AIC e DIC indicam melhores modelos. Os valores dos intervalos de confiança e credibilidade, juntamente com as amplitudes estão disponíveis nas tabelas.

A Figura 1 apresenta as curvas da função de sobrevivência estimadas pelo método de máxima verossimilhança e Bayesiano, quando considerado os modelos de Weibull e Burr XII na presença ou não de fração de cura, juntamente com as curvas de Kaplan-Meier que são obtidas através da função de sobrevivência empírica dos dados, estimadas não parametricamente (KAPLAN; MEIER, 1958b).



(a) EMV (b) Bayesiano

Figura 1: Ajuste dos modelos para o tempo até a reincidência do detento ao crime, considerando as distribuições Burr XII e Weibull, com e sem fração de cura.

Observado o ajuste dos dados aos modelos de sobrevivência (ver Figura 1), conclui-se que todos os modelos se ajustam bem aos tempos de reincidência, porém, os modelos de mistura aparentemente apresentam melhor ajuste. É possível perceber, também, que o modelo Burr XII

Tabela 9: Estimadores de Máxima Verossimilhança.

Distribuição	Parâmetro	Estimativa	Erro Padrão	I.C. (95%)	Amplitude	AIC
Weibull	μ	15997	5931.23	$(4331.97; 27662)$	23330.03	1283.3
	β	0.5577	0.06193	$(0.4359; 0.6794)$	0.2435	
Weibull (fração de cura)	μ	1222.15	1028.68	$(-800.75; 3245.05)$	4045.8	1283.7
	β	0.6701	0.1044	$(0.4648; 0.8754)$	0.4106	
	p	0.6606	0.1133	$(0.4377; 0.8834)$	0.4457	
Burr XII	μ	3805.83	3641.98	$(-3356.66; 10968)$	14324.66	1283.4
	α	0.7312	0.1611	$(0.4144; 1.0479)$	0.6335	
	λ	4.5583	3.7153	$(-2.7492; 11.8658)$	14.615	
Burr XII (fração de cura)	μ	983.10	622.48	$(-240.94; 2207.15)$	2448.09	1285.7
	α	0.7233	0.1633	$(0.4022; 1.0445)$	0.6423	
	λ	0.4046	2.1019	$(-3.7278; 4.5370)$	8.2648	
	p	0.6513	0.2153	$(0.2279; 1.0747)$	0.8468	

Tabela 10: Médias a posteriori

Distribuição	Parâmetro	Média	Desvio Padrão	Percentis (2.5 e 97.5)	Amplitude	DIC
Weibull	μ	14566.7	3163.5	(8461.6;19712.7)	11251.1	1282.523
	β	0.5768	0.0502	(0.4897;0.6868)	0.1971	
Weibull (fração de cura)	μ	1206.1	620.5	(490.7;2792.1)	2301.4	1282.566
	β	0.6881	0.0832	(0.5416;0.8634)	0.3218	
	p	0.6678	0.0678	(0.5037;0.7722)	0.2685	
Burr XII	μ	3436.8	1469.2	(1345.9;7066.6)	5720.7	1283.186
	α	0.7661	0.0891	(0.6125;0.9647)	0.3522	
	λ	4.2508	1.7401	(1.3130;8.0274)	6.7144	
Burr XII (fração de cura)	μ	766.7	158.4	(444.1;989.1)	545	1282.615
	α	0.7704	0.0890	(0.6146;0.9522)	0.3376	
	λ	0.5124	0.2888	(0.0279;0.9781)	0.9502	
	p	0.6707	0.0474	(0.5705;0.7574)	0.1869	

sem fração de cura se aproxima de ambos os modelos com fração de cura, mostrando sua maior flexibilidade em relação ao modelo Weibull.

Considerando o modelo Burr XII, sem fração de cura, na presença das covariáveis sexo e idade do primeiro delito afetando o parâmetro de locação μ, tem-se o seguinte modelo de regressão,

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right), \tag{85}$$

em que, x_1 é uma variável binária que assume valor 1 se o indivíduo é do sexo masculino e valor 0 se do sexo feminino e x_2 é a variável que representa a idade em anos completos em que o indivíduo cometeu seu primeiro delito; os parâmetros β_1 e β_2 medem, respectivamente, o efeito das covariáveis sexo e idade do primeiro delito no tempo de reincidência ao crime. A escolha pelo modelo Burr XII sem a presença de fração de cura é devido a sua melhor flexibilidade quando comparado com os outros modelos e por ser um modelo mais simples se comparado com o modelo Burr XII na presença de fração de cura.

A escolha por utilizar apenas duas covariáveis (Sexo e Idade do Primeiro Delito) é devido aos valores dos erros padrões que foram obtidos na incorporação das demais covariáveis ao modelo, os valores dos erros padrões para os parâmetros que medem os efeitos das outras covariáveis estavam inflacionados, quando comparados com as estimativas dos parâmetros. Isto ocorreu devido à falta de dados sem censura (exatamente observados) para os indivíduos com as informações das covariáveis, levando a estimativas inflacionadas dos erros padrões. Veja a tabela 20, que apresenta as estimativas dos parâmetros com seus erros padrões e intervalos de confiança, e as tabelas 11, 12, 13, 14, 15, 16, 17, 18 e 19, apresentam as frequências absolutas das variáveis *dummy* e censuras de cada valor, onde o valor 0 na censura significa que o indivíduo é censurado e o valor 1 significa que a observação foi completa.

Tabela 11: Frequência Absoluta da variável *dummy* "Emprego" *vs* Censura

Emprego	Censura		Total
	0	1	
Desempregado	11	5	16
Empregado	81	13	94
Total	92	18	110
Dados não informados: 246			

Tabela 12: Frequência Absoluta da variável *dummy* "Cor da Pele" *vs* Censura

Cor da Pele	Censura		Total
	0	1	
Branco	199	49	248
Pardo	80	17	97
Total	279	66	345
Dados não informados: 11			

Tabela 13: Frequência Absoluta da variável *dummy* "Cor da Pele" *vs* Censura

Cor da Pele	Censura		Total
	0	1	
Branco	252	59	311
Negro	27	7	34
Total	279	66	345
Dados não informados: 11			

Tabela 14: Frequência Absoluta da variável *dummy* "Escolaridade" *vs* Censura

Escolaridade	Censura		Total
	0	1	
Sem Escolaridade	129	31	160
1^o grau completo	35	11	46
Total	164	42	206
Dados não informados: 150			

Tabela 15: Frequência Absoluta da variável *dummy* "Escolaridade" *vs* Censura

Escolaridade	Censura		Total
	0	1	
Sem Escolaridade	136	41	177
2^o grau completo	28	1	29
Total	164	42	206
Dados não informados: 150			

Tabela 16: Frequência Absoluta da variável *dummy* "Escolaridade" *vs* Censura

Escolaridade	Censura		Total
	0	1	
Sem Escolaridade	158	42	200
3^o grau completo	6	0	6
Total	164	42	206
Dados não informados: 150			

Tabela 17: Frequência Absoluta da variável *dummy* "Benefício ao sair do SECAT" *vs* Censura

Benefício ao sair do SECAT	Censura		Total
	0	1	
Outro Motivo	221	67	288
Fiança	66	1	67
Total	287	68	355
Dados não informados: 1			

Tabela 18: Frequência Absoluta da variável *dummy* "Benefício ao sair do SECAT" *vs* Censura

Benefício ao sair do SECAT	Censura		Total
	0	1	
Outro Motivo	239	50	289
Transferência	48	18	66
Total	287	68	355
Dados não informados: 1			

Tabela 19: Frequência Absoluta da variável *dummy* "Benefício ao sair do SECAT" *vs* Censura

Benefício ao sair do SECAT	Censura		Total
	0	1	
Outro Motivo	283	67	350
Outros Benefícios	4	1	5
Total	287	68	355
Dados não informados: 1			

Tabela 20: Estimativa dos parâmetros

Variável	Parâmetro	Estimativa	Erro Padrão	Intervalo de Confiança 95%
Emprego	β_1	206.44	124.85	$(-40.9878; 453.87)$
Cor da Pele	β_1	0.002648	0.5658	$(-1.1104; 1.1157)$
	β_2	-0.5479	0.8498	$(-2.2193; 1.1235)$
Escolaridade	β_1	-0.1830	0.8706	$(-1.8994; 1.5335)$
	β_2	3.7932	1.9806	$(-0.1116; 7.6981)$
	β_3	19.4628	292.54	$(-557.30; 596.22)$
Benefício ao sair do SECAT	β_1	135.30	315.16	$(-484.51; 755.11)$
	β_2	93.8203	3297.36	$(-6390.99; 6578.63)$
	β_3	8.1868	8493.49	$(-16696; 16712)$

Os valores das estimativas da Tabela 20 foram obtidos através do modelo bruto, isto é, com o acréscimo de uma covariável por vez, obtendo a estimativa e a resposta de cada variável sozinha no modelo 85.

Nas Tabelas 21 e 22 tem-se, respectivamente, os estimadores de Máxima Verossimilhança e os resultados inferenciais considerando a análise Bayesiana para os parâmetros do modelo de regressão proposto em (85).

Para a análise Bayesiana no modelo de regressão são consideradas a distribuição a priori não informativa normal $N(0; 100)$ com média 0 e variância 100 para o parâmetro β_0, distribuição a priori não informativa normal $N(0; 10)$ com média 0 e variância 10 para os parâmetros β_1 e β_2, e distribuição a priori não informativa gama $Gama(1; 1)$ para os parâmetros de forma λ e α.

Tabela 21: Estimadores de Máxima Verossimilhança para os parâmetros do modelo de regressão.

Parâmetro	Estimativa	Erro Padrão	Intervalo de Confiança 95%
β_0	8.1246	1.5514	$(5.0735; 11.1757)$
β_1	-2.2136	0.9041	$(-3.9917; -0.4355)$
β_2	0.0903	0.0258	$(0.0395; 0.1411)$
α	0.6571	0.1392	$(0.3834; 0.9308)$
λ	1.8223	2.5565	$(-3.2055; 6.8502)$

Tabela 22: Médias a posteriori para os parâmetros do modelo de regressão.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade 95%
β_0	8.4237	1.1515	$(6.2167; 10.7335)$
β_1	-2.2106	0.8963	$(-3.8946; -0.4923)$
β_2	0.0936	0.0262	$(0.0435; 0.1451)$
α	0.6092	0.0803	$(0.4597; 0.7697)$
λ	1.1108	0.9947	$(0.000021; 3.0162)$

A partir dos resultados das Tabelas 21 e 22, é possível concluir que as covariáveis sexo e idade do primeiro crime afetam o tempo de reincidência ao crime, visto que o valor zero não está incluído nos intervalos de confiança e credibilidade para os parâmetros β_1 e β_2.

4 ANÁLISE DOS RESULTADOS E PERSPECTIVAS FUTURAS

Observado o ajuste dos dados aos modelos de sobrevivência (ver Figura 1), conclui-se que todos os modelos se ajustam bem aos tempos de reincidência, porém, os modelos de mistura aparentemente apresentam melhor ajuste. É possível perceber, também, que o modelo Burr XII sem fração de cura se aproxima de ambos os modelos com fração de cura, mostrando sua maior flexibilidade em relação ao modelo Weibull.

O uso de métodos usuais para MCMC como a procedure `MCMC` do software `SAS`, permite uma boa simplificação na obtenção das inferências para os modelos propostos. No Apêndice D tem-se os programas computacionais utilizados para a obtenção das estimativas dos parâmetros para os modelos propostos.

Usualmente na análise de dados de sobrevivência tem-se a presença de fração de cura, quando uma certa proporção de indivíduos não experimentam o evento de interesse. Para a análise desse tipo de dados, pode-se utilizar diferentes formulações paramétricas, como, por exemplo, os modelos de mistura. Essas formulações usualmente assumem uma distribuição paramétrica, como por exemplo, Weibull, log-normal ou exponencial para os indivíduos susceptíveis. Como visto na Seção 3 o uso da distribuição Burr XII pode ser de grande interesse prático, pois esse modelo apresenta uma grande flexibilidade no ajuste aos dados se comparado com outras distribuições mais usuais. Além disso, a distribuição Burr XII ainda foi pouco explorada na análise de dados de sobrevivência, principalmente na presença de fração de cura.

A partir dos resultados dados nas Tabelas 9, 10, 21 e 22 observa-se, considerando a presença ou não de fração de cura, que as estimativas pontuais para os parâmetros dos modelos propostos são muito similares, mas os erros-padrão são bem menores considerando a metodologia Bayesiana esta comparação pode ser feita pois as distribuições a priori utilizadas são não informativas. Isto implica em estimativas bem mais precisas, é importante salientar que os resultados clássicos são obtidos utilizando métodos assintóticos nem sempre bem precisos e dependentes do tamanho amostral e a proporção de dados censurados. Pela Tabela 21 é possível verificar que o intervalo de confiança para o parâmetro λ se inicia em -3.2055, o que não

ocorre com a estimativa Bayesiana observada na Tabela 22, isto é um problema recorrente em estimativas via métodos frequentista, visto que $\lambda > 0$. Observa-se, também, que os critérios DIC e AIC para todos os modelos propostos são muito próximos.

Por fim, pode-se concluir ao analisar os resultados obtidos pelas estimativas no modelo de regressão que indivíduos do sexo masculino tendem a reincidir com um tempo menor se comparado com os indivíduos do sexo feminino, devido ao valor negativo do parâmetro β_1. E ainda que, quanto mais velho for o indivíduo, maior o tempo de reincidência ao crime, devido ao valor positivo da estimativa do parâmetro β_2, estas estimativas são possíveis devido a incorporação das covariáveis no parâmetro de locação μ do modelo de Burr XII.

Como propostas para a continuação da pesquisa, será utilizado um ponto de mudança para a função de sobrevivência, devido a mudança de direção da curva de sobrevivência estimada via Kaplan-Meier para $t > 2000$ (ver Figura 1) e a incorporação das covariáveis não incluídas nesta etapa do trabalho.

4.1 PUBLICAÇÕES

A realização deste trabalho gerou algumas publicações e uma submissão de artigo, sendo:

1. *Aplicação das Distribuições Burr XII e Weibull em Dados de Reincidência ao Crime*, XXII Simpósio Nacional de Probabilidade e Estatística, 24 a 29 de Julho de 2016, Porto Alegre, RS (Publicação de Resumo nos Anais do Evento)

2. *Emprego das Distribuições de Weibull e Burr XII em Dados Penais*, V Semana Acadêmica de Matemática da UTFPR, 26 a 30 de Setembro de 2016, Cornélio Procópio, PR (Comunicação Científica)

3. *Análise de Sobrevivência Aplicada à Dados de Reincidência ao Crime de Detentos do SECAT da Comarca de Primeiro de Maio, PR* Revista Brasileira de Biometria - UNESP (Submissão de Artigo)

REFERÊNCIAS

AALEN. Nonparametric inference for a family of counting processes. **Annals of Statistics**, 1978.

ACHCAR, J. A. et al. **Uma Introducao aos Metodos Bayesianos**. Sao Paulo: Universidade de Sao Paulo, 2012.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **Proceedings of the 2 nd International Symposium on Information Theory.**, p. 176–723, 1973.

ANGELIS, R. D. et al. Mixture models for cancer survival analysis: application to population-based data with covariates. **Statistics in Medicine**, v. 18, n. 4, p. 441–454, 1999.

ARAUJO, A. M. M. D. Dissertação, **Aplicação de Modelos de Mistura de Longa Duração em Dados de Reincidência ao Crime**. 2004.

ARENALES, S.; DAREZZO, A. **Cálculo numérico: aprendizagem com apoio de software**. ed.1. São Paulo: Thomson Learning, 2008.

BAYES, T. An essay towards solving in the doctrine of chances. **Philosophical Transactions of the Royal Society London.**, 1763.

BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, v. 47, p. 501–515, 1952.

BOAG, J. Maximum likelihood estimation of the proportion of patients cured by Cancer therapy. **Journal of the Royal Statistical Society, B**, v. 11, p. 15–53, 1949.

BRASIL. **Código Penal Brasileiro**. Brasil: BRASIL, 1940.

BRASIL. **Lei de Execução Penal**. Brasil: BRASIL, 1984.

BRAZIL. **Constituição Política do Império do Brazil**. Brasil: Legislativo, 1824.

BURR, I. W. Cumulative frequency functions. **Annals of Mathematical Statistics**, v. 13, p. 215–232, 1942. ISSN 0003-4851.

CAMPOS, G. de; SOUSA, R. R. de. **O trabalho prisional como eixo de reintegração social: a experiência do projeto liberdade com dignidade pela ótica dos presos**. jul./dez. 2013.

CANCHO, V. G.; BOLFARINE, H. Modeling the presence of immunes by using the exponentiated-Weibull model. **Journal of Applied Statistics**, v. 28, n. 6, p. 659–671, 2001. ISSN 0266-4763.

CASELLA, G.; GEORGE, E. I. Explaining the Gibbs sampler. **Amer. Statist.**, v. 46, n. 3, p. 167–174, 1992. ISSN 0003-1305. Disponível em: <http://dx.doi.org/10.2307/2685208>.

CHIB, S.; GREENBERG, E. Undestanding the metropolis-hastings algorithm. **The American Statistician**, v. 49, n. 4, p. 327–335, 1995.

CHIQUEZI, A. **Reincidencia criminal e sua atuacao como circunstancia agravante**. Sao Paulo, 2009. PUC.

COELHO-BARROS, E. A. **Modelagem em Análise de Sobrevivência Para Dados Médicos Bivariados Utilizando Funções Cópulas e Fração de Cura**. Tese (Doutorado) — Universidade de São Paulo, 2014.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. Curitiba: ABE - Projeto Fischer, 2006.

COPAS, J. B.; HEYDARI, F. Estimating the risk of reoffending by using exponential mixture models. **Journal of the Royal Statistical Society, A**, v. 160, n. 2, p. 237–252, 1997.

DEPEN. **Levantamento Nacional de Informações Penitenciárias INFOPEN**. Brasil: DEPEN, 2014.

DEPUTADOS, C. dos. **CPI SISTEMA CARCERÁRIO**. Brasilia: Biblioteca Digital da Câmera dos Deputados, 2009.

DUNSMUIR, W. et al. Modeling the transitions between employement states for young australians. **Australian Journal of Statistics**, v. 31, n. A, p. 165–196, 1989.

EHLERS, R. S. **Bayesian statistics**. 03 2014. University of Sao Paulo.

FAREWELL, V. T. The use of mixture models for the analysis of survival data with long-term survivors. **Biometrics**, v. 38, p. 1041–1046, 1982.

FAREWELL, V. T. Mixture models in survival analysis: Are they worth the risk? **The Canadian Journal of Statistics**, v. 14, n. 3, p. 257–262, 1986. ISSN 0319-5724.

GAMEL, J. W.; MCLEAN, I. W.; ROSENBERG, S. H. Proportion cured and mean log survival time as functions of tumor size. **Statistics in Medicine**, v. 9, p. 999–1006, 1990.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, n. 410, p. 398–409, 1990. ISSN 0162-1459.

GELMAN, A.; RUBIN, D. B. Inference from interative simulation using multiple sequences. **Statistical Science**, v. 7, p. 457–511, 1992.

GHITANY, M. E.; MALLER, R. A. Asymptotic results for exponential mixture models with long term survivors. **Statistics**, v. 23, p. 321–336, 1992.

GIESER, P. W. et al. Modelling cure rates using the gompertz model with covariate information. **Statistics in Medicine**, n. 17, p. 831–839, 1998.

HAYBITTLE, J. L. A two parameter model for the survival curve of treated cancer patients. **Journal of the American Statistical Association**, v. 53, p. 16–26, 1965.

HOSMER, D. W.; LEMESHOW, S. **Applied Survival Analysis: Regression Modeling of Time to Event Data**. 1st. ed. New York, NY, USA: John Wiley & Sons, Inc., 1999. ISBN 0471154105.

IPEA, I. de P. E. A. **Reincidência Criminal no Brasil**. Rio de Janeiro: IPEA, 2015.

KANNAN, N. et al. The generalized exponential cure rate model with covariates. **Journal of Applied Statistics**, v. 37, n. 9-10, p. 1625–1636, 2010. ISSN 0266-4763.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, vol. 53, n. no. 282, p. pp. 457–481, Jun. 1958.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, p. 457–481, 1958. ISSN 0162-1459.

LAMBERT, P. C. et al. Estimating and modeling the cure fraction in population-based cancer survival analysis. **Biostatistics**, v. 8, n. n. 3, p. p. 576–594, out. 2007.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: John Wiley and Sons, 1982. xi+580 p. ISBN 0-471-08544-8.

MALLER, R. A.; ZHOU, X. **Survival analysis with long-term survivors**. Chichester: John Wiley & Sons Ltd., 1996. xviii+278 p. (Wiley Series in Probability and Statistics: Applied Probability and Statistics). ISBN 0-471-96201-5.

MARQUES, G. R. A relação de trabalho no regime fechado de execução de pena privativa de liberdade. 2013.

MARTINEZ, E. Z. et al. Mixture and non-mixture cure fraction models based on the generalized modified weibull distribution with an application to gastric cancer data. **Computer methods and programs in biomedicine**, p. p. 343–355, Jul. 2013.

MEEKER, W. Q. Limited failure population life tests: Application to integrated circuit reliability. **Technometrics**, v. 29, n. 1, p. 51–65, 1987.

MEEKER, W. Q.; ESCOBAR, L. A. Statistical methods for reliability data using sas software. 1998.

MISSIO, F.; JACOBI, L. F. Variáveis dummy: especificações de modelos com parâmetros variáveis. **Ciência e Natura**, v. 29, n. 1, p. 111, 2007.

NELSON, W. Theory and applications of hazard plotting for censored failure data. **Technometrics**, 1972.

NG, S. K.; MCLACHLAN, G. J. On modifications to the long-term survival mixture model in the presence of competing risks. **Journal of Statistical Computation and Simulation**, v. 61, p. 77–96, 1998. ISSN 0094-9655.

PARANAIBA, P. F. **Caracterizacao e extensoes da distribuicao Burr XII: propriedades e aplicacoes**. Tese (Doutorado) — Universidade de Sao Paulo, Piracicaba, 2012.

PENG, Y.; DEAR, K. B. G. A nonparametric mixture model for cure rate estimation. **Biometrics**, v. 56, p. 237–243, 2000.

PENG, Y.; DEAR, K. B. G.; DENHAM, J. W. A generalized F mixture model for cure rate estimation. **Statistics in Medicine**, v. 17, n. 8, p. 813–830, 1998.

RODRIGUES, J. et al. COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. **J. Statist. Plann. Inference**, v. 139, n. 10, p. 3605–3611, 2009. ISSN 0378-3758.

SAS. **The MCMC Procedure, SAS/STAT User's Guide, Version 9.22**. Cary, NC: SAS Institute Inc.: SAS, 2010. 4102–4326 p.

SAS. **The NLMIXED Procedure, SAS/STAT User's Guide, Version 9.22**. Cary, NC: SAS Institute Inc.: SAS, 2010. 4967–5062 p.

SHAO, Q.; ZHOU, X. A new parametric model for survival data with long-term survivors. **Stat Med**, v. 23, n. 22, p. 3525–43, 2004.

SILVA, R. M. R. da; SELLOS-KNOERR, V. C. de. **O Trabalho como Instrumento da Promoção da Dignidade do Preso**. 2015.

SOUZA, R. M. de. **Modelagem estatistica em estudos de bioequivalencia sob o enfoque Bayesiano**. Tese (Doutorado) — Universidade de Sao Paulo, Ribeirao Preto, 2015.

SY, J. P.; TAYLOR, J. M. G. Estimation in a Cox proportional hazards cure model. **Biometrics**, v. 56, p. 227–236, 2000.

TAYLOR, J. M. G. Semiparametric estimation in failure time mixture models. **Biometrics**, v. 51, p. 899–907, 1995.

WIENKE, A.; LOCATELLI, I.; YASHIN, A. I. The modelling of a cure fraction in bivariate time-to-event data. **Austrian Journal of Statistics**, v. 35, n. 1, p. 67–76, 2006.

YAMAGUCHI, K. Accelerated failure-time regression model with a regression model for the survining fraction: an application to the analysis of âpermanent employmentâ in japan. **Journal of the American Statistical Association**, v. 87, p. 284–292, 1992.

YIN, G.; IBRAHIM, J. G. Cure rate models: a unified approach. **The Canadian Journal of Statistics**, v. 33, n. 4, p. 559–570, 2005. ISSN 0319-5724.

YU, B.; TIWARI, R. C.; CRONIN, K. Z. Cure fraction estimation from the mixture cure models for grouped survival times. **Statistics in Medicine**, v. 23, p. 1733–1747, 2004.

APÊNDICE A – MÉTODOS DE MONTE CARLO VIA CADEIA DE MARKOV

Ao trabalhar com o método de Monte Carlo via Cadeia de Markov (MCMC) para simular as distribuições condicionais ou a posteriori de cada parâmetro, utiliza-se comumente o Algoritmo de Gibbs, quando são conhecidas as funções de distribuição da posteriori conjunta e caso contrário utiliza-se do algoritmo de Metropolis-Hastings.

A ideia é obter uma amostra da distribuição a posteriori e calcular estimativas amostrais de características dessa distribuição utilizando os métodos iterativos, baseados em Cadeias de Markov. Uma questão que deve ser levada em consideração quando se trabalha com Cadeias de Markov é que as iterações iniciais influenciam os valores da Cadeia de Markov, devido a este fato, na prática os valores iniciais são descartados, já que após um certo número de iterações, espera-se que a cadeia convirja para a distribuição de equilíbrio. Esses valores iniciais possuem o nome de *Amostra de Aquecimento*.

A.1 O ALGORITMO DE METROPOLIS-HASTINGS

Quando não são conhecidas as distribuições condicionais, são utilizados métodos de amostragem por importância ou o algoritmo de Metropolis-Hastings para gerar amostras das distribuições a posteriori, dentro de uma dada probabilidade de aceite, esse mecanismo é o que garante a convergência do algoritmo para uma distribuição de equilíbrio.

Se a cadeia está no estado θ e um valor θ' seja gerado através de uma distribuição proposta $q(.|\theta)$, o valor de θ' será aceito com probabilidade:

$$\alpha(\theta,\theta') = min\left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)}\right). \tag{86}$$

em que $\pi(\theta)$ é a distribuição de interesse.

Note que não é necessário o conhecimento da distribuição π por completo e ainda, uma cadeia pode permanecer no mesmo estado por diversas iterações.

Em termos gerais, o algoritmo de Metropolis-Hastings pode ser descrito como:

1.Inicialize o contador de iterações com $t = 0$ e especifique um valor para θ^0;

2.Gere um novo valor θ' da distribuição $q(.|\theta)$;

3.Calcule a probabilidade de aceitação de $\alpha(\theta, \theta')$ e gere $u \sim U(0,1)$;

4.Se $u \leq \alpha$ aceite o valor e faça $\theta^{t+1} = \theta'$, caso contrário, faça $\theta^{t+1} = \theta$;

5.Incremente o contador t para $t+1$ e volte para o passo 2.

Veja que se um valor é rejeitado, o valor atual é considerado na próxima iteração, um "salto" na direção "ascendente" é sempre aceito, já um "salto" na direção "descendente" só é aceito dentro de uma certa probabilidade.

O desempenho do algoritmo pode ser influenciado pela escolha da locação e da escala da distribuição geradora, assim, o pesquisador deverá possuir certa experiência na escolha destes valores para os parâmetros de escala e forma (CHIB; GREENBERG, 1995) (ACHCAR et al., 2012).

A.2 O ALGORITMO DE GIBBS

Na utilização do algoritmo de Gibbs a cadeia sempre irá mover-se à um novo valor, isto é, não existe o mecanismo de aceitação ou rejeição dos valores.

O funcionamento do algoritmo de Gibbs é o seguinte: simula-se quantidades aleatórias de distribuições condicionais completas $\pi(\theta_i|y, \theta_{(i)})$ de modo que produza uma cadeia de Markov. Assim, supor um conjunto aleatório de valores iniciais $\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}$ para o vetor de parâmetros θ. O algoritmo é escrito como:

(i)Gerar $\theta_1^{(1)}$ de $\pi(\theta_1|y, \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)})$;

(ii)Gerar $\theta_2^{(1)}$ de $\pi(\theta_2|y, \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)})$;

(iii)Gerar $\theta_3^{(1)}$ de $\pi(\theta_3|y, \theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_k^{(0)})$;

.

.

.

(k)Gerar $\theta_k^{(1)}$ de $\pi(\theta_k|y, \theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{k-1}^{(1)})$

De modo que substituindo os valores iniciais pelas próximas iterações até um valor suficientemente grande de iterações, até que ocorra a convergência, gerando uma amostra de $\pi(\theta)$. Observe que o algoritmo de Gibbs é um caso particular do algoritmo de Metropolis-Hastings, em que os elementos de θ são atualizados um de cada vez (ou em blocos), tomando a distribuição condicional completa como proposta e a probabilidade de aceitação igual a 1 (EHLERS, 2014).

A convergência dos algoritmos pode ser mostrada graficamente, podendo ser construídos histogramas, em que histogramas similares de θ_j e θ indicam convergência do algoritmo, análise de séries temporais das amostras selecionadas e através de análise de variância, proposta por Gelman e Rubin (1992).

A verificação de convergência (ou não convergência) de tais métodos é de responsabilidade do analista, sendo de grande dificuldade na prática em determinados casos.

APÊNDICE B – CRITÉRIOS DE INFORMAÇÕES

Ao estimar informações sobre quaisquer dados, ocorre uma perca de informações em relação ao real valor dos dados. Para mensurar essa perca utiliza-se de critérios de informações, de modo que quanto menor for esse valor, melhor o modelo se ajusta aos valores reais. O AIC *(Akaike Information Criterion)* e DIC *(Deviance Information Criterion)* serão apresentados abaixo.

B.1 AKAIKE INFORMATION CRITERION (AIC)

O critério desenvolvido por Akaike (1973) assume a existência de um modelo "real" que descreve os dados que é desconhecido e busca escolher dentro de inúmeros modelos, aquele que minimiza a diferença entre o modelo dito "real" e o escolhido, e é representado por:

$$AIC = -2\log(L) + 2k, \tag{87}$$

onde L é a função de verossimilhança e k são os parâmetros estimados do modelo.

O critério de Akaike é baseado na "Teoria da Informação", isto é, sempre deve ser comparado com modelos conhecidos e aplicados ao mesmo estudo, de forma que os melhores modelos apresentarão menores valores de AIC. Grandes amostras geram valores mais consistentes, distribuições com poucos parâmetros podem ocasionar em um valor alto de AIC.

O critério de Akaike é de grande utilização, já que não é necessário a utilização de modelos encaixados, isto é, a mesma família de distribuições para realizar a comparação dos valores de AIC.

Para uma melhor utilização do AIC, deve-se observar se o número da amostra é suficientemente superior ao número de parâmetros das distribuições e se as distribuições que serão comparadas não possuem uma elevada amplitude de parâmetros, quando comparadas. Caso ocorra algum evento assim, é recomendável a utilização do AIC_c, que é o Critério de Akaike Corrigido.

B.2 DEVIANCE INFORMATION CRITERION (DIC)

Quando são avaliados modelos Bayesianos, um dos critérios de informação utilizados na seleção do melhor modelo é o *Deviance Information Criterion* ou DIC, utilizado quando as distribuições a posteriori são obtidas através de simulações MCMC.

O critério pode ser escrito como:

$$DIC = \hat{D} + 2p_D, \tag{88}$$

em que $\hat{D}$ é o valor da *deviance* estimado a posteriori para os parâmetros de interesse e p_D são os números dos parâmetros do modelo em análise, que é dado por $p_D = \tilde{D} - \hat{D}$ em que $\tilde{D}$ é a média a posteriori da deviance (SOUZA, 2015).

De forma análoga ao AIC, quando compara-se valores de DIC, os melhores valores são os menores.

APÊNDICE C – DIVISÃO DAS CATEGORIAS DE CRIMES

•Crimes contra a pessoa:

Homicídio Simples (art. 121, caput): Matar alguém.

Aborto (art. 124): Provocar aborto em si mesma ou consentir que outrem lho provoque.

Lesão Corporal (art. 129): Ofender a integridade corporal ou a saúde de outrem.

Violência Doméstica (art. 129 §9): Se a lesão for praticada contra ascendente, descendente, irmão, cônjuge ou companheiro, ou com quem conviva ou tenha convivido, ou, ainda, prevalecendo-se o agente das relações domésticas, de coabitação ou de hospitalidade.

Sequestro e cárcere privado (art. 148): Privar alguém de sua liberdade, mediante sequestro ou cárcere privado.

Outros crimes entre os artigos 122 e 154-A

•Crimes contra o patrimônio:

Furto (art. 155): Subtrair, para si ou para outrem, coisa alheia móvel.

Roubo (art. 157): Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência a pessoa, ou depois de havê-la, por qualquer meio, reduzido à impossibilidade de resistência.

Latrocínio (art. 157 §3º): Se da violência resulta lesão corporal grave, a pena é de reclusão, de cinco a quinze anos, além da multa; se resulta morte, a reclusão é de vinte a trinta anos, sem prejuízo da multa.

Extorsão (art. 158): Constranger alguém, mediante violência ou grave ameaça, e com o intuito de obter para si ou para outrem indevida vantagem econômica, a fazer, tolerar que se faça ou deixar de fazer alguma coisa.

Apropriação indébita (art. 168): Apropriar-se de coisa alheia móvel, de que tem a posse ou a detenção.

Estelionato (art. 171): Obter, para si ou para outrem, vantagem ilícita, em prejuízo alheio, induzindo ou mantendo alguém em erro, mediante artifício, ardil, ou qualquer outro meio fraudulento.

Receptação (art. 180): Adquirir, receber, transportar, conduzir ou ocultar, em proveito próprio ou alheio, coisa que sabe ser produto de crime, ou influir para que terceiro, de boa-fé, a adquira, receba ou oculte.

Outros crimes entre os artigos 156 e 179

•Crimes contra a dignidade sexual:

Estupro (art. 213): Constranger alguém, mediante violência ou grave ameaça, a ter conjunção carnal ou a praticar ou permitir que com ele se pratique outro ato libidinoso.

Atentado violento ao pudor (art. 214): Constranger alguém, mediante violência ou grave ameaça, a praticar ou permitir que com ele se pratique ato libidinoso diverso da conjunção carnal.

Estupro de vulnerável (art. 217-A): Ter conjunção carnal ou praticar outro ato libidinoso com menor de 14 (catorze) anos.

Corrupção de menores (art. 218): Induzir alguém menor de 14 (catorze) anos a satisfazer a lascívia de outrem.

Artigos 215, 216-A, 218-A, 218-B,, 227, 228, 229, 230

•Crimes contra a paz pública:

Formação de quadrilha ou bando (art. 288): Associarem-se 3 (três) ou mais pessoas, para o fim específico de cometer crimes.

•Crimes contra a fé pública:

Uso de documento falso (art. 304): Fazer uso de qualquer dos papéis falsificados ou alterados.

•Crimes contra a Administração Pública:

Peculato (art. 312 e 313): Apropriar-se o funcionário público de dinheiro, valor ou qualquer outro bem móvel, público ou particular, de que tem a posse em razão do cargo, ou desviá-lo, em proveito próprio ou alheio apropriar-se de dinheiro ou qualquer utilidade que, no exercício do cargo, recebeu por erro de outrem.

Corrupção passiva (art. 317): Solicitar ou receber, para si ou para outrem, direta ou indiretamente, ainda que fora da função ou antes de assumi-la, mas em razão dela, vantagem indevida, ou aceitar promessa de tal vantagem.

•Crimes praticados por particular contra a Administração Pública:

Corrupção ativa (art. 333): Oferecer ou prometer vantagem indevida a funcionário público, para determiná-lo a praticar, omitir ou retardar ato de ofício.

•Legislação específica:

•Grupo: Drogas:

Tráfico de Drogas (art. 33 da Lei 11343/06): Importar, exportar, remeter, preparar, produzir, fabricar, adquirir, vender, expor à venda, oferecer, ter em depósito, transportar, trazer consigo, guardar, prescrever, ministrar, entregar a consumo ou fornecer drogas, ainda que gratuitamente, sem autorização ou em desacordo com determinação legal ou regulamentar.

Associação para o tráfico (art. 35 da Lei 11343/06): Associarem-se duas ou mais pessoas para o fim de praticar, reiteradamente ou não, qualquer dos crimes previstos nos art. 33.

•Grupo: Estatuto do Desarmamento (Lei 10826/03):

Porte ilegal de arma de fogo de uso permitido (art. 14): Portar, deter, adquirir, fornecer, receber, ter em depósito, transportar, ceder, ainda que gratuitamente, emprestar, remeter, empregar, manter sob guarda ou ocultar arma de fogo, acessório ou munição, de uso permitido, sem autorização e em desacordo com determinação legal ou regulamentar.

Disparo de arma de fogo (art. 15): Disparar arma de fogo ou acionar munição em lugar habitado ou em suas adjacências, em via pública ou em direção a ela, desde que essa conduta não tenha como finalidade a prática de outro crime.

•Grupo: Crimes de trânsito (Lei 9503/97):

Homicídio culposo na condução de veículo automotor (art. 302): Praticar homicídio culposo na direção de veículo automotor.

Outros artigos do 303 a 312

•Grupo: Outras legislações específicas:

Estatuto da criança e do adolescente "ECA" (Lei 8069/90)

Crimes contra o Meio Ambiente (Lei 9605/98)

Crimes de Tortura (Lei 9455/97)

APÊNDICE D – PROGRAMAS DO SOFTWARE SAS

Esse apêndice apresenta os programas computacionais utilizados na resolução dos problemas propostos.

Listing D.1: Distribuição Weibull (*Procedure NLMIXED* e *Procedure MCMC*).

```
/* Modelo Weibull */
proc nlmixed data=dados tech=nra;
        parms mu=15997 beta=0.5577;
        bounds mu>0, beta>0;

            t = TempoReincidencia;
        delta = Censura;

    lf = log(beta)-beta*log(mu)+(beta-1)*log(t)-(t/mu)**beta;
        ls = -(t/mu)**beta;
        l      = delta*lf+(1-delta)*ls;

        model t~general(l);
run;
proc mcmc data=dados nmc=200000 seed=34512 dic nbi=15000 THIN=150
diag=all;
        parms mu=3000 beta=0.4;
        prior mu:       ~       uniform(0,20000);
        prior beta:     ~       uniform(0,10000);

            t = TempoReincidencia;
        delta = Censura;

    lf = log(beta)-beta*log(mu)+(beta-1)*log(t)-(t/mu)**beta;
```

```
25          ls = -(t/mu)**beta;
26          l        = delta*lf+(1-delta)*ls;
27
28          model t~general(l);
29  run;
30
31  /* Modelo Weibull (mixture)*/
32  proc nlmixed data=dados tech=nra;
33          parms mu=1222.15 beta=0.6701 p=0.6606;
34          bounds mu>0, beta>0;
35
36              t = TempoReincidencia;
37          delta = Censura;
38
39  S0 = exp(-(t/mu)**beta);
40  lf = log(1-p)+log(beta)-beta*log(mu)+(beta-1)*log(t)-(t/mu)**beta;
41  ls = log(p+(1-p)*S0);
42  l  = delta*lf+(1-delta)*ls;
43
44          model t~general(l);
45  run;
46  proc mcmc data=dados nmc=200000 seed=34512 dic nbi=15000 THIN=100
47  diag=all;
48          parms mu=1200 beta=0.6701 p=0.6606;
49          prior mu:       ~       gamma(0.001,iscale=0.001);
50          prior beta:     ~       gamma(0.01,iscale=0.01);
51          prior p:        ~       uniform(0,1);
52
53
54              t = TempoReincidencia;
55          delta = Censura;
56
57  S0 = exp(-(t/mu)**beta);
58  lf = log(1-p)+log(beta)-beta*log(mu)+(beta-1)*log(t)-(t/mu)**beta;
59  ls = log(p+(1-p)*S0);
```

```
60 l  = delta*lf+(1-delta)*ls;
61
62          model t~general(l);
63 run;
```

Listing D.2: Distribuição Burr XII (*Procedure NLMIXED* e *Procedure MCMC*).

```
1  /* Modelo Burr */
2  proc nlmixed data=dados tech=nra;
3           parms mu=3805.83 beta=0.7312 lambda=4.5583;
4           bounds mu>0, beta>0, lambda>0;
5
6                t = TempoReincidencia;
7           delta = Censura;
8
9  lf = log(beta)-beta*log(mu)+(beta-1)*log(t)-(1+1/lambda)
10 *log(1+lambda*(t/mu)**beta);
11 ls = -(1/lambda)*log(1+lambda*(t/mu)**beta);
12 l         = delta*lf+(1-delta)*ls;
13
14          model t~general(l);
15 run;
16 proc mcmc data=dados nmc=250000 seed=34512 dic nbi=15000 THIN=150
17 diag=all;
18          parms mu=3805.83 beta=0.7312 lambda=4.5583;
19          prior mu:          ~   gamma(0.001,iscale=0.001);
20          prior beta:        ~   gamma(0.001,iscale=0.001);
21          prior lambda:   ~      gamma(1,iscale=1);
22
23               t = TempoReincidencia;
24          delta = Censura;
25
26 lf = log(beta)-beta*log(mu)+(beta-1)*log(t)-(1+1/lambda)
27 *log(1+lambda*(t/mu)**beta);
28 ls = -(1/lambda)*log(1+lambda*(t/mu)**beta);
29 l         = delta*lf+(1-delta)*ls;
30
```

```
31          model t~general(l);
32  run;
33  /* Modelo Burr (mixture)*/
34  proc nlmixed data=dados;
35          parms mu=983.1 beta=0.7330 lambda=0.6446 p=0.6390;
36          bounds mu>0, beta>0, lambda>0, 0<p<1;
37
38              t = TempoReincidencia;
39          delta = Censura;
40
41  S0 = (1+lambda*(t/mu)**beta)**(-1/lambda);
42  lf = log(1-p)+log(beta)-beta*log(mu)+(beta-1)*log(t)-(1+1/lambda)
43  *log(1+lambda*(t/mu)**beta);
44  ls = log(p+(1-p)*S0);
45  l  = delta*lf+(1-delta)*ls;
46
47          model t~general(l);
48  run;
49  proc mcmc data=dados nmc=200000 seed=34512 dic nbi=15000 THIN=150
50  diag=all;
51          parms mu=500 beta=0.7 lambda=0.4 p=0.6;
52          prior mu:                ~       uniform(0,1000);
53          prior beta:              ~       uniform(0,1000);
54          prior lambda:            ~       uniform(0,1);
55          prior p:                 ~       uniform(0,1);
56
57
58          t = TempoReincidencia;
59          delta = Censura;
60
61  S0 = (1+lambda*(t/mu)**beta)**(-1/lambda);
62  lf = log(1-p)+log(beta)-beta*log(mu)+(beta-1)*log(t)-(1+1/lambda)*
63  log(1+lambda*(t/mu)**beta);
64  ls = log(p+(1-p)*S0);
65  l  = delta*lf+(1-delta)*ls;
```

```
66
67 model t~general(l);
68 run;
```

Listing D.3: Distribuição Burr XII (modelo de Regressão) (*Procedure NLMIXED* e *Procedure MCMC*).

```
1  proc nlmixed data=dados tech=nra;
2  parms beta0=8.1245 beta1=-2.2136 beta2=0.09028 alpha=0.6571
3  lambda=1.8225;
4  bounds alpha>0, lambda>0;
5
6                t = TempoReincidencia;
7           delta = Censura;
8
9       mu = exp(beta0+beta1*x1+beta2*IdadePrimeiroDelito);
10
11 lf = log(alpha)-alpha*log(mu)+(alpha-1)*log(t)-(1+1/lambda)
12 *log(1+lambda*(t/mu)**alpha);
13 ls = -(1/lambda)*log(1+lambda*(t/mu)**alpha);
14 l         = delta*lf+(1-delta)*ls;
15
16          model t~general(l);
17 run;
18 proc mcmc data=dados nmc=200000 seed=34512 dic nbi=10000
19 THIN=100 diag=all STATS(PERCENTAGE=(2.5 97.5));
20 parms beta0=8 beta1=0 beta2=0 alpha=0.6571 lambda=1.82;
21          prior beta0:  ~ normal(0,var=100);
22          prior beta1:  ~ normal(0,var=10);
23          prior beta2:  ~ normal(0,var=10);
24          prior alpha:  ~ gamma(1,iscale=1);
25          prior lambda: ~ gamma(1,iscale=1);
26
27               t = TempoReincidencia;
28          delta = Censura;
29
30      mu = exp(beta0+beta1*x1+beta2*IdadePrimeiroDelito);
```

```
31
32 lf = log(alpha)-alpha*log(mu)+(alpha-1)*log(t)-(1+1/lambda)*
33 log(1+lambda*(t/mu)**alpha);
34 ls = -(1/lambda)*log(1+lambda*(t/mu)**alpha);
35 l      = delta*lf+(1-delta)*ls;
36
37       model t~general(l);
38 run;
```

APÊNDICE E – GRÁFICOS DE CONVERGÊNCIA DAS SIMULAÇÕES MCMC

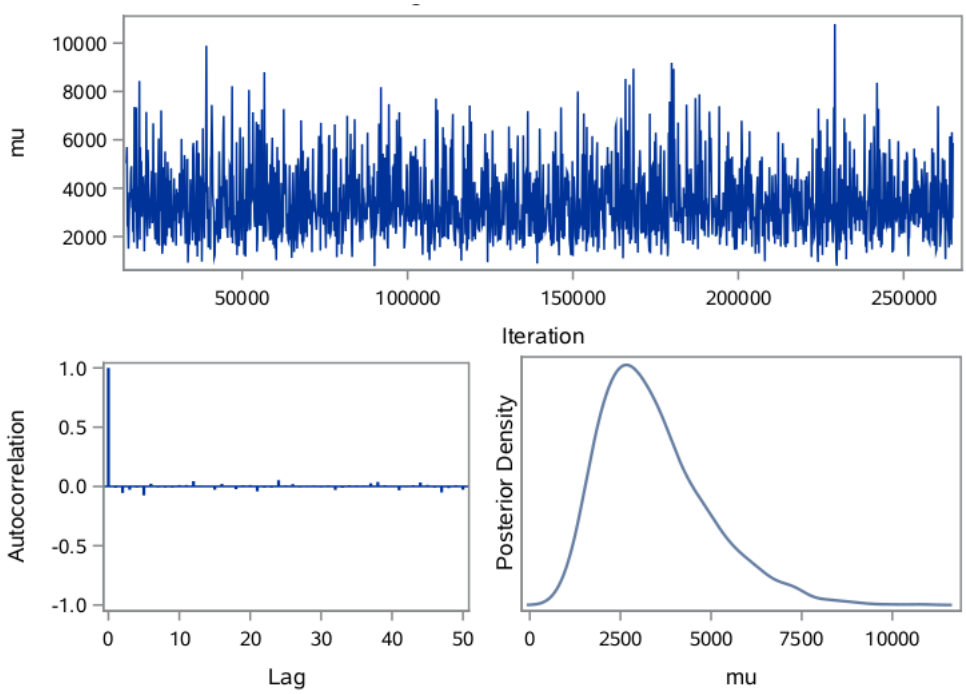


Figura 2: Convergência do parâmetro μ do modelo de Burr XII sem fração de cura

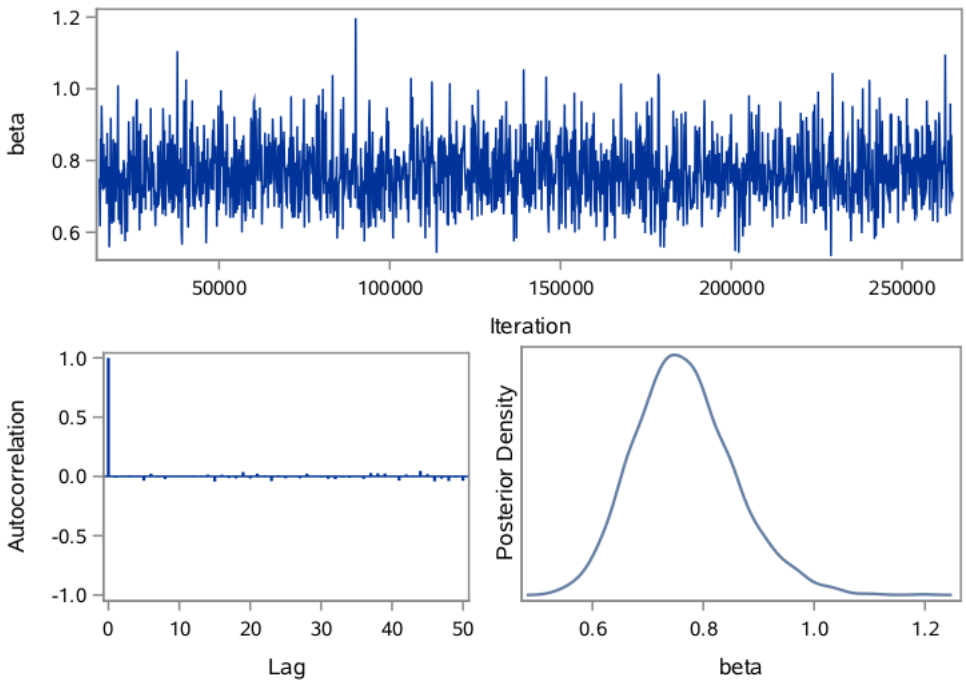


Figura 3: Convergência do parâmetro β do modelo de Burr XII sem fração de cura

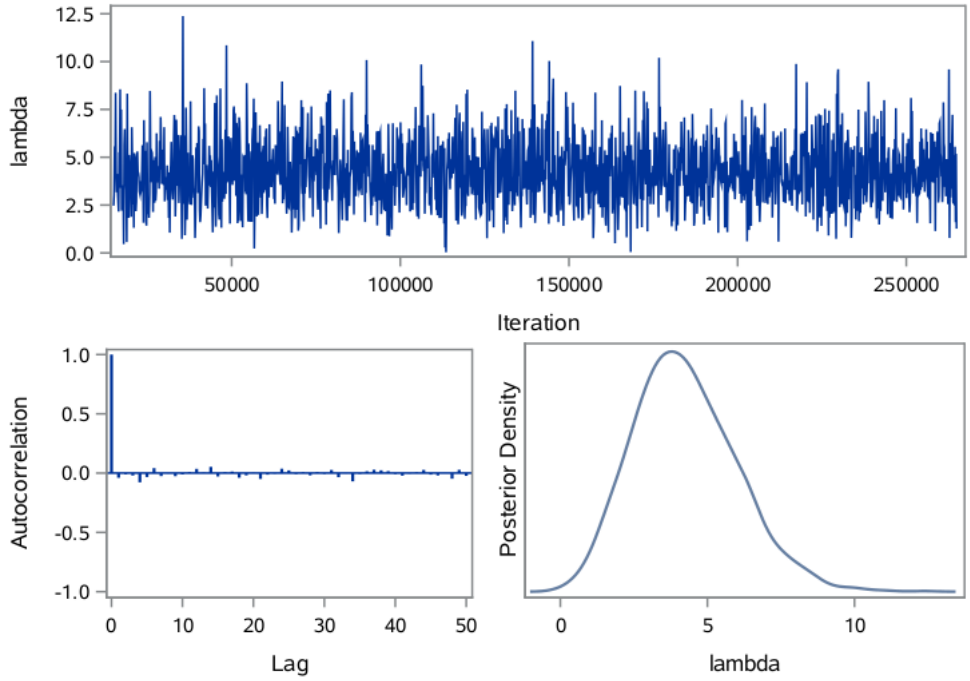


Figura 4: Convergência do parâmetro λ do modelo de Burr XII sem fração de cura

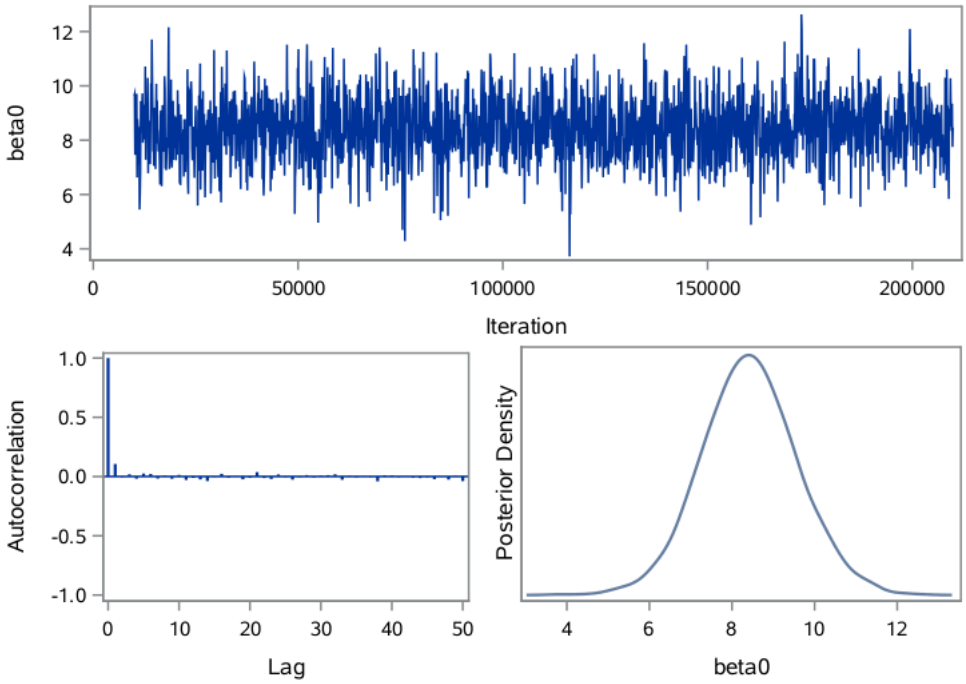


Figura 5: Convergência do parâmetro β_0 do modelo de regressão

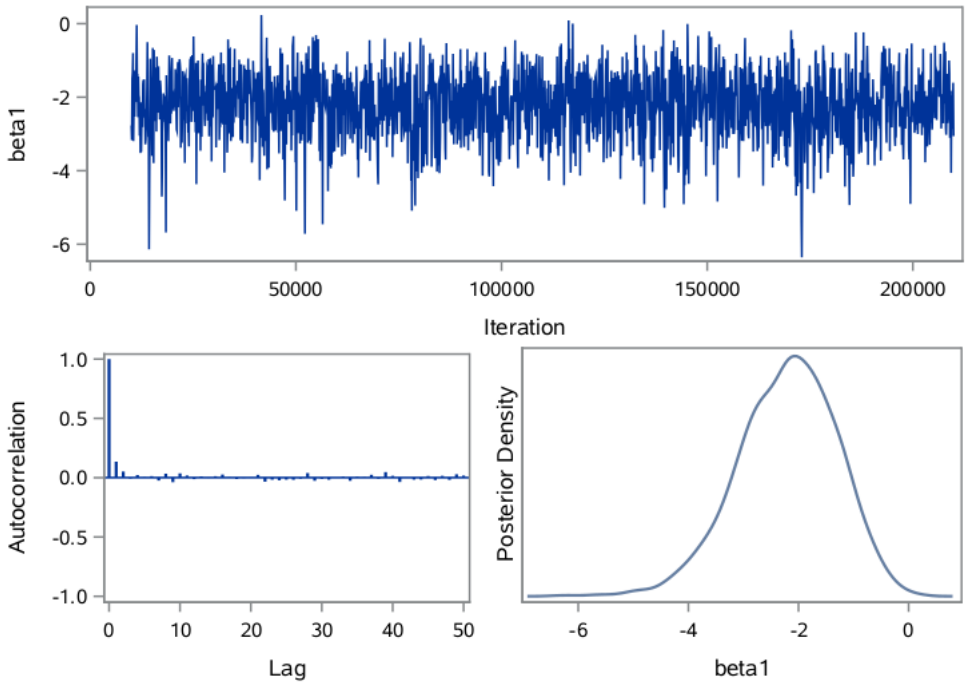


Figura 6: Convergência do parâmetro β_1 do modelo de regressão

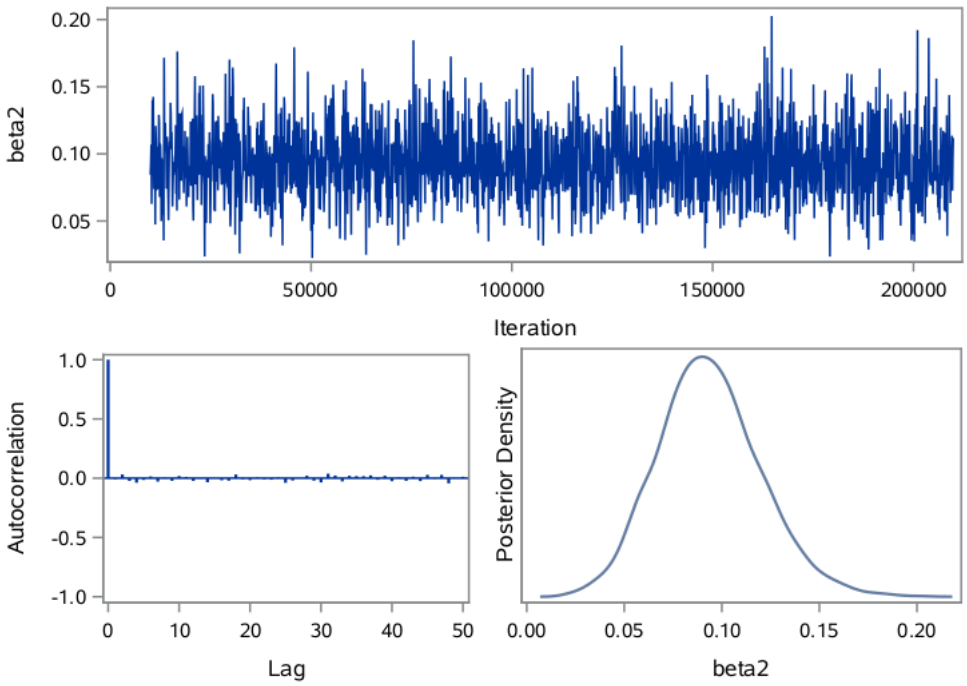


Figura 7: Convergência do parâmetro β_2 do modelo de regressão

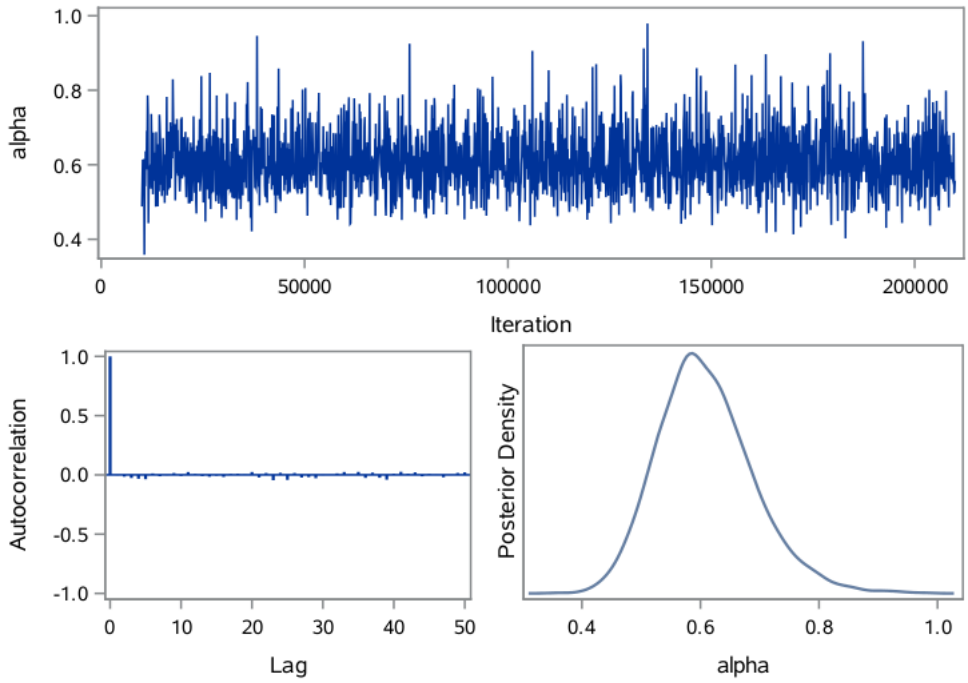


Figura 8: Convergência do parâmetro α do modelo de regressão

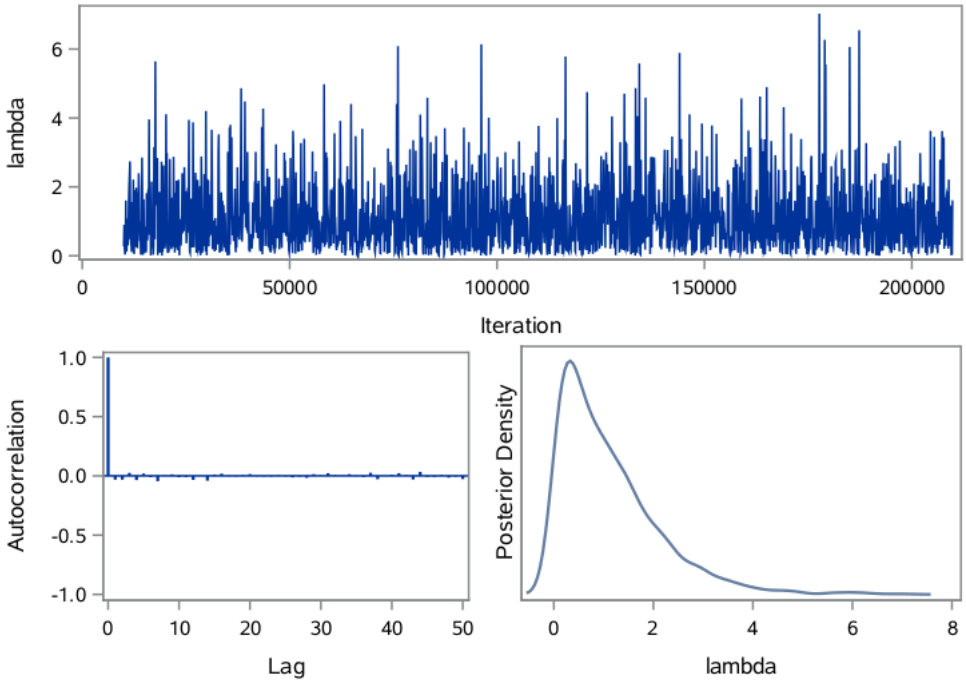


Figura 9: Convergência do parâmetro λ do modelo de regressão