



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Welton Chagas

**ANÁLISE DE SOBREVIVÊNCIA APLICADA A DADOS PENAIS DO
MUNICÍPIO DE PRIMEIRO DE MAIO**

**CURITIBA
2019**



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Welton Chagas

ANÁLISE DE SOBREVIVÊNCIA APLICADA A DADOS PENAIIS DO MUNICÍPIO DE PRIMEIRO DE MAIO

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório B do Curso de Estatística do Setor de Ciências Exatas da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Nivea da Silva Matuda Machado

Coorientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA
2019**

AGRADECIMENTOS

Gostaria de agradecer primeiramente a Deus, que permitiu que tudo isso acontecesse ao longo da minha vida e que, não somente nestes anos como universitário, mas em todos os momentos, tem sido o maior mestre que alguém pode conhecer.

Agradeço à minha avó Iraci e à minha mãe Jucimara, que me apoiaram incondicionalmente durante todo esse processo, me incentivando nas horas difíceis de desânimo e cansaço.

Gostaria de estender meu agradecimento aos professores do Departamento de Estatística da UFPR. Mestres dedicados e compreensivos que me passaram ensinamentos que vou levar para a vida e me permitiram ser eu mesmo sempre.

Às minhas orientadoras, Professora Dra. Suely Ruiz Giolo, que me acompanhou integralmente na primeira etapa do trabalho e comprou a ideia do projeto, assim como na segunda parte, em que pude contar com a orientação da Professora Dra. Nivea da Silva Matuda Machado, que também abraçou a ideia e prestou a maior assistência a esse projeto.

Agradeço também ao acadêmico Matheus Henrique Pimenta Zanon e ao Dr. Emílio Augusto Coelho Barros, que gentilmente forneceram os dados para estudos e foram sempre solícitos para sanar dúvidas que iam surgindo sobre os dados.

De forma geral, sem citar nomes, aos meus colegas que ingressaram no curso em 2013 ou que foram incorporados no decorrer do curso. Foram inúmeras as vezes que nos reunimos em horários mais diversos possíveis para estudarmos. Tenho a certeza que cada um, da sua maneira, faz parte dessa conquista.

*"Mesmo desacreditado e ignorado por todos,
não posso desistir,
pois para mim,
vencer é nunca desistir".
(Albert Einstein)*

RESUMO

O sistema penitenciário no Brasil apresenta uma situação caótica, com os presídios lotados e presos condenados no mesmo ambiente daqueles que ainda não foram a julgamento. Devido à falta de dados disponíveis nessa área, existem poucas publicações que abordam esse assunto, o que reforça a importância deste trabalho, que apresenta um estudo sobre a reincidência ao crime de 350 indivíduos que passaram pelo SECAT do Município de Primeiro de Maio, entre dezembro de 2009 e dezembro de 2015. Para a análise dos dados desse estudo, foram utilizadas técnicas, no contexto de análise de sobrevivência, com o objetivo de estudar o tempo até a reincidência ao crime desses indivíduos, após terem tido uma passagem pelo SECAT, levando-se em consideração algumas variáveis, tais como: sexo, idade e motivo da prisão. Inicialmente, foram utilizados o estimador de Kaplan-Meier e o teste *logrank* para proceder a uma análise descritiva das variáveis. Na sequência, foram ajustados os seguintes modelos de regressão: modelo de Cox, modelo de Cox com fragilidade Normal, e modelo de Cox com fragilidade Gama. Ao final das análises, o modelo selecionado foi o modelo de Cox. Com base neste modelo, foi constatado que os homens apresentaram risco maior de reincidência ao crime do que as mulheres. Risco maior de reincidência ao crime também foi observado quando o motivo de saída foi transferência, seguido dos motivos alvará de soltura e fiança/outros, nesta ordem. Os indivíduos mais jovens também apresentaram risco maior de reincidência ao crime.

Palavras-chave: Reincidência criminal. Modelo de Cox. Razão de riscos. SECAT.

Sumário

AGRADECIMENTOS.....	iii
RESUMO	v
1 INTRODUÇÃO.....	7
2 OBJETIVOS	10
2.1 Objetivo Geral.....	10
2.2 Objetivos Específicos	10
3 MATERIAL E MÉTODOS	11
3.1 Material.....	11
3.2 Métodos.....	13
4 RESULTADOS E DISCUSSÃO	21
5 CONSIDERAÇÕES FINAIS	39
REFERÊNCIAS	41
APÊNDICES.....	43

1 INTRODUÇÃO

Segundo o Sistema Online Geopresídios do Conselho Nacional de Justiça (CNJ), o Brasil ocupa atualmente a 4ª posição entre as penitenciárias com as maiores populações. Em 2017, o Sistema Brasileiro encontrava-se com 654.836 presos, com capacidade projetada para os manter de 403.345 vagas, mostrando assim um déficit de 251.491 vagas.

Mapeamento realizado pelo Fórum Brasileiro de Segurança Pública em 2017, com o apoio do Instituto de Pesquisa Econômica Aplicada (IPEA), indicou que a população carcerária no Brasil cresceu 83 vezes em setenta anos. Esses dados mostram a necessidade de se estudar e entender mais sobre a situação carcerária, a ressocialização em relação aos prisioneiros, a reincidência criminal e seus fatores determinantes, bem como sobre a eficácia de dispositivos alternativos como meios de contornar esta crise no sistema prisional brasileiro.

A situação do sistema penitenciário no Brasil pode ser considerada caótica, com presídios lotados e presos condenados no mesmo ambiente daqueles que ainda não foram a julgamento. Dados apontam que cerca de 35 mil presos vivem amontoados em 25 mil vagas no Rio Grande do Sul, considerado um dos piores do País. No Nordeste, o cenário também é alarmante, sendo considerada a região com maior taxa de ocupação do País, com quase dois presos por vaga, de acordo com o último levantamento realizado em janeiro de 2017 pelo Conselho Nacional do Ministério Público. Outro exemplo de superlotação, segundo uma pesquisa realizada pelo Conselho Nacional do Ministério Público (CNMP), é o Ceará, com 20 mil presos para 12 mil vagas.

O artigo 5º, inciso III da Constituição Federal Brasileira, diz que "ninguém será submetido à tortura nem a tratamento desumano ou degradante". Além disso, é atribuído ao Estado o dever de assegurar os direitos e garantias dos detentos, bem como a organização dos serviços destinados à execução da pena, sendo o principal objetivo da pena a regeneração dos condenados, os readaptando à vida social. O artigo 10 da Lei de Execução Penal (LEP) cita que "a assistência ao preso e ao internado, como dever do Estado, objetiva prevenir o crime e orientar o retorno à convivência em sociedade, estendendo-se esta ao egresso". Entre as

atenções básicas, a LEP prevê que devem ser prestadas aos presos, assistência à saúde, assistência psicológica, educacional, jurídica, religiosa, social e material.

Entretanto, os estudos mostram que isso não acontece e o próprio sistema acaba sendo violador desses direitos. Ou seja, embora a ressocialização dos detentos seja a principal finalidade do sistema penitenciário, a prisão deixou de ser uma medida de ressocialização para ser um ambiente de torturas, tratamento desumano e que despreza direitos, pelo que apontam os dados apresentados pelo Conselho Nacional de Justiça (CNJ). Com a precariedade do sistema prisional brasileiro, os presos acabam sofrendo dupla penalidade: primeiro sendo privados da liberdade em função da pena e, segundo, com o desrespeito aos seus direitos fundamentais, devido às péssimas condições dos presídios, ficando expostos a doenças graves, falta de assistência médica, entre outros fatores.

Entre os especialistas, predomina a opinião sobre a incapacidade da prisão no que se refere à ressocialização do condenado. Os ataques mais severos advêm dos adeptos da criminologia crítica, que censuram a ressocialização por implicar a violação do livre-arbítrio e da autonomia do sujeito. A opinião, no entanto, de que a prisão não é capaz de ressocializar não se estende aos rumos que deveriam ser dados à prisão. Quanto a isso, Baratta (1976) aponta duas grandes posições: realista e idealista. Os adeptos da posição realista, partem da premissa que a prisão não é capaz de constituir um espaço de ressocialização e defendem que o máximo que ela pode fazer é neutralizar o delinquente. No extremo oposto, estão os que se inserem na posição idealista, que permanecem na defesa da prisão como espaço de prevenção especial positiva (ressocialização). Apesar de admitir seu fracasso para este fim, advogam que é preciso manter seu objetivo principal, visto que o seu abandono acabaria reforçando o caráter exclusivamente punitivo da pena, dando à prisão a única função de excluir da sociedade aqueles que são considerados delinquentes.

O Conselho Nacional de Justiça (CNJ) aponta também a questão da separação dos presos. Em 2015, apenas 10% das prisões separavam presos primários dos reincidentes, 22% por periculosidade. Apenas um terço das prisões separou os que são de grupos ou facções criminosas dos demais. O Coordenador do Conselho Permanente sobre os Direitos dos Presos, Everaldo Patriota, comenta: “como é que eu vou misturar uma pessoa que fez um furto em uma loja ou supermercado com um traficante, latrocinista, ou até mesmo um comandante

de quadrilha. Muitas pessoas vão para o presídio e querem sair de lá, querem trabalhar, querem cumprir sua pena em paz, mas não conseguem porque não há política para isso”.

A Doutora em Direito da Universidade de Brasília, Rosa Mendes, cita outro aspecto importante, que é a permanência de presos provisórios na cadeia. Segundo o Conselho Nacional de Justiça, de um total de 244 mil presos, quase 40% são provisórios, ou seja, pessoas que ainda não foram a julgamento. Mendes comenta que "são pessoas que muitas vezes acabam esquecidas durante muitos anos e que acabam até cumprindo as penas que deveriam ser aplicadas ou que poderiam ser aplicadas dentro dessa condição de provisórios. Então, isso é um alimento de um sistema carcerário da pior espécie, da pior forma, porque são pessoas que sequer a condenação definitiva tem". Mendes também enfatiza que é preciso melhorar as condições dos presídios, mas que a construção de novas unidades não resolverá o problema, pois é necessário repensar a política criminal para que não se utilize a pena de prisão de forma indiscriminada como é feito.

Tendo em vista as discussões que envolvem o sistema prisional brasileiro, é importante a realização de estudos que possam contribuir de algum modo com essas discussões. Nesse contexto, este trabalho apresenta os resultados de um estudo sobre reincidência criminal realizado em uma cidade de baixa população (em torno de 11.000 habitantes) do estado do Paraná.

Em linhas gerais, o trabalho apresenta a seguinte estrutura. Nos Capítulos 2 e 3, são descritos os principais objetivos do trabalho, assim como o conjunto de dados e os métodos estatísticos utilizados para a análise dos mesmos. No Capítulo 4, são apresentados os resultados das análises realizadas e, por fim, tem-se, no Capítulo 5, as principais conclusões do estudo.

2 OBJETIVOS

2.1 Objetivo Geral

O presente estudo teve como objetivo estudar o tempo até a reincidência ao crime de indivíduos do Município de Primeiro de Maio, localizado no Estado do Paraná, que por algum motivo já tinham passagem pelo Setor de Carceragem Temporária (SECAT) do Município, buscando a identificação de possíveis fatores que podem estar relacionados ao tempo até a reincidência desses indivíduos.

2.2 Objetivos Específicos

- a) Identificar os perfis dos detentos, a partir das informações disponíveis no banco de dados, com maior risco de reincidência ao SECAT do Município de Primeiro de Maio;
- b) Apresentar uma análise descritiva dos dados com a finalidade de obter uma visão geral das informações disponíveis no estudo, sem infringir o direito de preservar integralmente a identificação de cada presidiário;
- c) Ajustar modelos estatísticos aos dados descritos na Seção 3.1.2, tendo como foco a identificação de fatores (tais como, sexo, idade, tipo de crime, dentre outros) que podem estar relacionados ao tempo até a reincidência criminal dos indivíduos sob estudo;
- d) Comparar e discutir os resultados dos modelos ajustados, apresentando as conclusões mais relevantes.

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Contextualização dos Dados do SECAT

Os dados analisados foram cedidos por Zanon e Barros (2018), tendo sido coletados pelos autores junto ao Setor de Carceragem Temporária (SECAT) de Primeiro de Maio, município paranaense localizado na região metropolitana de Londrina, que tem aproximadamente 11 mil habitantes. Por ter uma população baixa, o município possui um único SECAT com capacidade máxima para seis indivíduos, divididos em três celas: uma delas para indivíduos do sexo masculino, outra para indivíduos do sexo feminino, e a última para indivíduos que cometem crimes contra a dignidade sexual, para garantir a integridade física dos mesmos, já que outros detentos podem não os aceitar na mesma cela.

Por lei, o SECAT deve receber apenas indivíduos capturados em flagrante, os quais deveriam permanecer ali apenas o tempo necessário para as atribuições legais e então, após esse período, ficariam à disposição da justiça para serem soltos ou transferidos aos Centros de Detenção Provisório (CDP). Contudo, os SECAT's abrigam cerca de 32% da população carcerária do Paraná, mesmo sendo lugares que afrontam a dignidade das pessoas que ali estão por não contar com uma estrutura para mantê-los em cárcere (ZANON; BARROS, 2018).

3.1.2 Conjunto de Dados

De acordo com Zanon e Barros (2018), o banco de dados é composto de 365 indivíduos detidos por algum motivo, entre dezembro de 2009 e dezembro de 2015, no SECAT do município de Primeiro de Maio, sendo todos não reincidentes. Considera-se reincidente aquele indivíduo que, por algum motivo, tem mais de uma passagem pelo SECAT. Para todos os indivíduos, informações referentes à data de entrada, data de saída, tipo de crime cometido, motivo da prisão e motivo da saída, foram obtidos através do livro de registros de presos do SECAT, que é preenchido manualmente. Informações adicionais sobre o "sexo do indivíduo" e a "idade em que ele cometeu o primeiro delito" foram obtidas através do Sistema de Registro Policial da Polícia Civil do Paraná.

Neste trabalho, a variável resposta de interesse consiste do tempo (em dias) contado desde a liberação do indivíduo até a sua reincidência ao crime. Para os que não apresentaram reincidência, foi registrado o tempo desde a liberação até 05/12/2015 (data considerada para a análise dos dados). Desse modo, tem-se, para cada indivíduo i ($i = 1, \dots, 365$), o par de informação (t_i, δ_i) , sendo t_i o tempo registrado e $\delta_i = 1$, se ocorreu a reincidência e $\delta_i = 0$, se não ocorreu.

Devido à falta de informações para 15 dos 365 indivíduos, decidiu-se pela exclusão dos mesmos do conjunto de dados, restando assim 350 indivíduos para as análises. O Quadro 1 apresenta as informações (variáveis) disponíveis para esses indivíduos, de acordo com suas respectivas categorias.

Quadro 1 - Informações na base de dados de indivíduos presos no SECAT do município de Primeiro de Maio entre dezembro de 2009 e dezembro de 2015

Variável	Descrição
Prontuário	Identificação do indivíduo
Sexo	Feminino Masculino
Idade	Idade no primeiro delito, em anos
Tatuagem	Possui Não possui
Tipo de crime	Crime contra a pessoa Crime contra o patrimônio Desarmamento Drogas Trânsito Outros motivos
Ano do primeiro delito	2009 ou 2010 2011 2012 2013 2014 2015
Cor	Branca Parda Negra Amarela
Motivo de saída	Alvará de soltura Transferência Pagamento de fiança ou outros
Tempo até a reincidência	Tempo, em dias, até cometer novo crime
Censura	Houve ou não houve reincidência

Fonte: O autor (2019).

Os dados descritos foram previamente analisados por Zanon e Barros (2018) com o objetivo de estudar a variável T = tempo até a reincidência criminal. Para tanto, os autores fizeram uso de modelos paramétricos (no contexto de análise de sobrevivência), tanto na ausência quanto na presença de fração de imunes. Para o ajuste dos modelos, foram consideradas duas distribuições de probabilidade, a de Weibull e a Burr XII. Também foram comparados os procedimentos de estimação clássico e Bayesiano. Segundo os autores, o modelo que apresentou melhor ajuste aos dados foi o modelo Burr XII sem fração de imunes.

Embora o objetivo deste trabalho seja essencialmente o mesmo daquele publicado por Zanon e Barros (2018), ressalta-se que para o estudo da variável T (tempo até a reincidência criminal) não foram considerados modelos no contexto paramétrico, mas sim modelos no contexto semi-paramétrico, que não necessitam que uma distribuição de probabilidade seja assumida para a variável T . Além disso, foram consideradas todas as variáveis descritas no Quadro 1 (os autores mencionados consideraram as variáveis sexo e idade no 1º delito).

3.1.3 Recursos Computacionais

O *software* R, versão 3.5.2 (R CORE TEAM, 2018), foi utilizado para realizar a análise exploratória dos dados, bem como ajustar os modelos a esses dados. Os principais pacotes utilizados foram: *survival*, *survivalROC*, *lattice*, *latticeExtra*, *RcolorBrewer* e *mass*. No Apêndice A, estão disponibilizados alguns códigos em R utilizados para a análise dos dados.

3.2 Métodos

3.2.1 Conceitos de Análise de Sobrevivência

Considere que T seja uma variável aleatória contínua e não-negativa representando a variável resposta tempo até a ocorrência de um evento de interesse. Além da variável resposta T , considere a presença de censuras, que é uma característica importante dos dados de sobrevivência, a qual ocorre quando não é possível observar o evento de interesse para alguns indivíduos sob estudo.

Existem diferentes tipos de censura. A censura à esquerda ocorre quando o tempo até o evento é menor do que o tempo registrado. Por outro lado, a censura é dita "intervalar" quando a informação disponível é somente a de que o evento de interesse ocorreu em um intervalo de tempo conhecido. Já a censura à direita, o tipo mais frequentemente observado, ocorre quando, por alguma razão, o tempo registrado é menor do que o tempo até o evento (COLOSIMO; GIOLO, 2006).

Dentre as funções utilizadas em análise de sobrevivência para representar a variável T , está a função de distribuição acumulada definida por

$$F(t) = P(T \leq t),$$

a qual fornece a probabilidade de um indivíduo sob risco apresentar o evento de interesse (desfecho) até o tempo t .

Por outro lado, a função de sobrevivência, dada por

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t),$$

fornece a probabilidade do indivíduo sobreviver ao tempo t . Esta função é monótona decrescente, tal que $S(t) = 1$ quando $t = 0$ e $S(t) = 0$ quando $t \rightarrow \infty$.

Outra função de interesse em análise de sobrevivência é a função taxa instantânea de falha, definida como o limite da probabilidade de ocorrência do evento para um indivíduo em um pequeno intervalo de tempo, dado que ele sobreviveu ao tempo t (LAWLESS, 1982), ou seja,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

Para estimar as funções apresentadas, foram propostos alguns métodos. Dentre eles, uma técnica não paramétrica bastante conhecida para estimar a função de sobrevivência é o produto-limite de Kaplan-Meier, apresentado a seguir.

3.2.2 Estimador de Kaplan-Meier

O estimador não paramétrico de Kaplan e Meier (1958) foi proposto para estimar a função de sobrevivência na presença de uma amostra com observações censuradas. Consiste de uma versão atualizada das tábuas de vida clássicas utilizadas em Ciências Atuariais desenvolvidas por John Graunt, em 1662, e pelo conhecido astrônomo Edmond Halley em 1693.

O estimador de Kaplan-Meier consiste de um produto, em que cada fator desse produto é obtido particionando-se o eixo do tempo em pequenos intervalos, determinados pela ocorrência de uma falha, e estimando-se a probabilidade de sobrevivida em cada intervalo condicionado à probabilidade de sobrevivência aos intervalos anteriores. O produto dessas estimativas resulta na probabilidade incondicional de sobrevivida e sobrevivida geral (KAPLAN; MEIER, 1958).

Considerando $t_1 < t_2 < \dots < t_k$ os k tempos distintos e ordenados de falha, d_j o número de falhas em t_j , $j = 1, \dots, k$, e n_j o número de indivíduos sob risco (isto é, que não apresentaram falha ou desfecho e não foram censurados até o tempo imediatamente anterior ao tempo t_j), o estimador de Kaplan-Meier (KM) é dado por

$$\hat{S}(t) = P(T > t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right).$$

O estimador de KM permite incorporar informações provenientes de censuras em seu procedimento de estimação desde que haja independência entre os tempos de falha e de censura. Além disso, apresenta as seguintes propriedades estatísticas: *i*) é estimador não viciado para amostras grandes; *ii*) é fracamente consistente; *iii*) converge assintoticamente para um processo gaussiano; e *iv*) é estimador de máxima verossimilhança de $S(t)$.

As estimativas obtidas via o estimador KM são usualmente representadas graficamente e mostram o comportamento da curva de sobrevivência. Pressupõe uma única causa de falha de interesse, que ocorre com probabilidade igual a 1 se o tempo de seguimento for suficientemente longo, e trata como censura os tempos observados para aqueles indivíduos que não apresentam o evento de interesse até o final do tempo de seguimento (KAPLAN; MEIER, 1958).

Neste trabalho, o estimador de Kaplan-Meier foi utilizado para avaliar a influência das variáveis (uma a uma), apresentadas no Quadro 1, sobre o tempo de reincidência ao crime. Para proceder a esta avaliação também por meio de um teste estatístico, as curvas de sobrevivência associadas às respectivas categorias de cada uma das variáveis mostradas no Quadro 1 foram comparadas entre si por meio do teste *logrank* (MANTEL, 1966).

Após a análise via o estimador de Kaplan-Meier, os dados foram analisados por meio de modelos de regressão, que permitem levar em consideração todas as variáveis simultaneamente. Os modelos considerados estão descritos a seguir.

3.2.3 Modelo de Regressão de Cox

O modelo de regressão de Cox (COX, 1972) é um dos mais populares para a análise de dados de sobrevivência na presença de um conjunto de covariáveis $\mathbf{x} = (x_1, \dots, x_p)$. Considerando uma amostra de tamanho n , o modelo assume que os tempos $t_i, i = 1, \dots, n$, são independentes e que a taxa de falha é dada por

$$\lambda(t | \mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}),$$

em que $\lambda_0(t)$ denota a função risco de base, não negativa e em que não é assumida nenhuma forma paramétrica; $\boldsymbol{\beta}$ é o vetor de parâmetros de regressão desconhecidos; e \mathbf{x} é o vetor observado de covariáveis.

Devido à presença do componente não paramétrico $\lambda_0(t)$, o método de máxima verossimilhança usual não é apropriado para estimar o vetor de parâmetros $\boldsymbol{\beta}$. Cox (1975) propôs, então, o método da verossimilhança parcial, que consiste em condicionar a função de verossimilhança nos tempos de ocorrência do evento, de modo a eliminar a função $\lambda_0(t)$. Nos casos em que ocorrem tempos empatados, aproximações para a função de verossimilhança parcial foram propostas por Breslow (1974) e Efron (1977).

Para estimar o componente não paramétrico $\lambda_0(t)$, Breslow (1972) propôs um estimador para o risco acumulado $\Lambda_0(t)$, dado por

$$\hat{\Lambda}_0(t) = \sum_{k: t_k < t} \frac{d_k}{\sum_{l \in R_k} \exp(\mathbf{x}'_l \hat{\boldsymbol{\beta}})},$$

com d_k o número de falhas em t_k e R_k o conjunto de indivíduos sob risco em t_k .

Uma das suposições do modelo de Cox é a proporcionalidade dos riscos, o que implica que a razão dos riscos de dois indivíduos quaisquer não depende do tempo. Diversos métodos gráficos, bem como testes, encontram-se disponíveis na literatura para verificar a suposição de proporcionalidade dos riscos. De modo similar, métodos gráficos também estão disponíveis para avaliar a qualidade de ajuste do modelo. Tais métodos se baseiam, em geral, na análise dos resíduos de Schoenfeld, de Cox-Snell, *martingal* e *deviance* (COLOSIMO; GIOLO, 2006).

3.2.3.1 Análise dos Resíduos do Modelo de Cox

Para avaliar a suposição de riscos proporcionais, Schoenfeld (1982) propôs os resíduos padronizados de Schoenfeld, denotados por s_{iq} , para $i = 1, \dots, d$, com d o número de falhas, e $q = 1, \dots, p$, com p o número de parâmetros. Com base nesses resíduos, Grambsch e Therneau (1994) sugeriram os gráficos $s_{iq} + \hat{\beta}_q$ versus t , $q = 1, \dots, p$, que devem apresentar inclinação nula para que a suposição de taxas proporcionais seja considerada válida.

A suposição de riscos proporcionais também pode ser avaliada por meio do coeficiente de correlação de Pearson (ρ) entre os resíduos padronizados de Schoenfeld s_{iq} e t , para $q = 1, \dots, p$. Valores de ρ próximos de zero indicam não haver evidências para a rejeição da suposição de taxas de falha proporcionais.

No que se refere à avaliação da qualidade geral de ajuste do modelo, são utilizados os resíduos de Cox-Snell (1968), definidos, para o modelo de Cox, por

$$\hat{e}_i = \hat{\Lambda}(t_i | \mathbf{x}_i) = \hat{\Lambda}_0(t_i) \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}}), \quad i = 1, \dots, n,$$

em que $\hat{\Lambda}_0(t_i)$ é a função risco de base acumulada obtida a partir do modelo ajustado. Se o modelo estiver bem ajustado, os resíduos \hat{e}_i 's devem ser considerados uma amostra censurada da distribuição exponencial padrão.

Para avaliar outros aspectos do modelo, como a presença de *outliers* ou pontos influentes, são utilizados os resíduos *martingal* e os resíduos *deviance*. Os resíduos *martingal* são definidos por

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

em que δ_i é a variável indicadora de falha e \hat{e}_i o i -ésimo resíduo de Cox-Snell, enquanto os resíduos *deviance*, considerados uma tentativa de tornar os resíduos *martingal* mais simétricos em torno de zero (facilitando a detecção de *outliers*), são definidos por

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

Se o modelo estiver bem ajustado, os resíduos citados devem apresentar comportamento aleatório em torno de zero (COLOSIMO; GIOLO, 2006).

3.2.3.2 Seleção de Covariáveis no Modelo de Cox

Para a seleção de quais covariáveis permanecem no modelo de Cox, foi utilizado o método de seleção *stepwise*, uma ferramenta utilizada nos estágios exploratórios da construção de um modelo para identificar um subconjunto útil de preditores. Neste método, o modelo inicia sem covariáveis, sendo incluída uma variável em cada um dos dois passos iniciais e, na sequência, inclui-se ou exclui-se uma variável a cada passo de acordo com o nível de significância estabelecido. O processo de seleção adiciona sistematicamente a variável mais significativa ou remove a variável menos significativa em cada passo.

3.2.4 Modelo de Cox com Fragilidade no Contexto Univariado

O modelo de fragilidade é caracterizado pela inclusão no modelo de um efeito aleatório ou variável aleatória, denominada "fragilidade" no contexto de análise de sobrevivência, que representa as informações que não puderam ou não foram observadas para o indivíduo i ($i = 1, \dots, n$), tais como informações ambientais ou genéticas ou outras não consideradas no planejamento do estudo. Para o indivíduo i , o modelo de fragilidade fica expresso da seguinte forma

$$\lambda_i(t) = z_i \lambda_0(t) \exp\{x_i' \boldsymbol{\beta}\},$$

em que $\boldsymbol{\beta}$ denota o vetor de parâmetros desconhecidos associados às covariáveis x_i , $\lambda_0(t)$ corresponde à função risco de base não especificada e z_1, \dots, z_n aos valores das fragilidades, assumidas serem uma amostra proveniente de variáveis aleatórias Z_i independentes e identicamente distribuídas com distribuição de probabilidade conhecida de média 1 e variância desconhecida.

A introdução do "termo de fragilidade" (efeito aleatório) na função taxa de falha é indexada para cada indivíduo i com o objetivo de se avaliar a presença de heterogeneidade entre eles devido à existência de variáveis que não foram observadas ou não são observáveis. O efeito aleatório incorporado à função taxa de falha como um fator multiplicativo é, em geral, não negativo, tendo em vista que $\lambda_i(t) > 0$ (COLOSIMO; GIOLO, 2006).

Neste trabalho, o modelo de fragilidade foi utilizado para análise dos dados considerando-se para as variáveis Z_i , $i = 1, \dots, n$, as distribuições Normal e Gama.

3.2.5 Modelo de Mistura com Fração de Imunes

Quando os dados apresentam um percentual elevado de não ocorrência do evento de interesse, o modelo de mistura com fração de imunes aparece como uma possibilidade a ser investigada para a análise dos dados. De acordo com Maller e Zhou (1996), para que esse tipo de modelo seja uma opção viável é necessário que: *i*) o tempo de seguimento dos indivíduos seja suficientemente longo, e *ii*) a curva de sobrevivência apresente uma estabilização a partir de um determinado tempo, indicando a presença de uma fração razoável de indivíduos que não irão experimentar o evento de interesse, mesmo se forem acompanhados por um período longo de tempo. A essa fração de indivíduos denomina-se "imunes ao evento" ou "sobreviventes de longa duração".

O modelo mencionado permite estimar, simultaneamente, se o evento de interesse tem probabilidade elevada de ocorrer (denominada incidência) e quando ele poderá ocorrer, denominada latência (CORBIÈRE; JOLY, 2007).

Para apresentar a formulação do modelo de mistura com fração de imunes, considere que na população de interesse exista duas subpopulações distintas: uma suscetível ao evento ($U = 1$) e outra não suscetível ao evento ($U = 0$). Ainda, seja T a variável aleatória não negativa representando o tempo até o evento de interesse e suponha que exista na população em estudo uma proporção $p_0 = 1 - \pi$ de indivíduos imunes ao evento e, conseqüentemente, uma proporção $q_0 = \pi$ suscetível ao evento. O modelo fica, então, dado por

$$\begin{aligned} S_p(t | \mathbf{x}, \mathbf{z}) &= P(U = 0 | \mathbf{z})P(T > t | U = 0, \mathbf{x}) + P(U = 1 | \mathbf{z})P(T > t | U = 1, \mathbf{x}) \\ &= (1 - \pi(\mathbf{z})) + \pi(\mathbf{z})S_U(t | \mathbf{x}), \end{aligned}$$

com \mathbf{z} o vetor de covariáveis associado à $\pi(\mathbf{z})$, que indica a probabilidade do indivíduo apresentar o evento, e \mathbf{x} o vetor de covariáveis associado à $S_U(t | \mathbf{x})$.

Neste trabalho, foi investigado a viabilidade desse modelo para a análise dos dados descritos previamente, tendo em vista o percentual elevado de não ocorrência de reincidência criminal (em torno de 81%).

3.2.6 Qualidade de Predição dos Modelos

Para avaliar a qualidade de predição dos modelos apresentados nas seções anteriores, bem como compará-los entre si, foi utilizado o critério de informação de Akaike (AIC), introduzido por Hirotugu Akaike em 1974. De modo geral, pode-se ajustar diferentes modelos e optar por aquele que produzir o menor valor de AIC. O AIC penaliza os modelos com covariáveis desnecessárias, sendo calculado por

$$AIC = -2 \log(L(\boldsymbol{\theta})) + 2p,$$

em que $\log(L(\boldsymbol{\theta}))$ corresponde ao logaritmo da função de verossimilhança do modelo com vetor de parâmetros $\boldsymbol{\theta}$ e p ao número de parâmetros do modelo.

Tendo como foco o modelo de Cox, Heagerty e Zheng (2005) propuseram um método alternativo ao R^2 para avaliar a qualidade de predição do modelo. Tal método, consiste em uma versão tempo-dependente da área sob a curva ROC, comumente utilizada em regressão logística. Para o modelo de Cox, a qualidade de predição é avaliada em tempos específicos por meio da área sob a curva ROC, em cada t , denotada por $AUC(t)$.

Para obtenção da $AUC(t)$, são consideradas, para o modelo de Cox, as seguintes definições de sensibilidade e especificidade tempo-dependentes:

$$\begin{cases} \text{sens}(c, t) = P(M_i > c \mid T_i = t) = P(M_i > c \mid \delta_i(t) = 1) \\ \text{esp}(c, t) = P(M_i \leq c \mid T_i > t) = P(M_i \leq c \mid \delta_i(t) = 0), \end{cases}$$

em que $M_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$, $i = 1, \dots, n$, corresponde ao marcador utilizado para previsão de falha no tempo t e $c \in \mathbb{R}$ denota o ponto de corte utilizado como critério para classificar a previsão como falha ou censura no tempo t .

Os indivíduos são classificados em cada tempo fixo t como falha ou censura com base no seu real status no tempo t , assim se o evento ocorreu para o indivíduo i , este assumirá o status de censura $\delta_i(t) = 0$ para todo tempo $t < T_i$ e o status de falha $\delta_i(t) = 1$ para $t = T_i$. Dentre os métodos propostos para estimar $\text{sens}(c, t)$ e $\text{esp}(c, t)$, há um baseado no teorema de Bayes e no estimador de Kaplan e Meier (1958), e outro baseado no estimador do vizinho mais próximo (AKRITAS, 1994), denotado por NNE.

4 RESULTADOS E DISCUSSÃO

4.1 Análise descritiva

Inicialmente, foi realizada uma análise descritiva dos dados apresentados na Seção 3.1.2 a fim de identificar a quantidade de indivíduos em cada categoria das variáveis mostradas no Quadro 1. Para tanto, pode-se observar na Tabela 1 as frequências absolutas e relativas (percentuais) em cada uma das categorias, bem como os percentuais de falhas e censuras. De acordo os dados mostrados na Tabela 1, observa-se que os maiores percentuais de presos estão na faixa etária de 24 a 40 anos, com 51%. Para a variável sexo, dos 350 indivíduos, 86% são do sexo masculino. Em relação ao crime cometido, dentre as seis categorias consideradas, a que apresentou maior frequência foi o crime contra a pessoa, com 101 indivíduos, correspondendo a 29%.

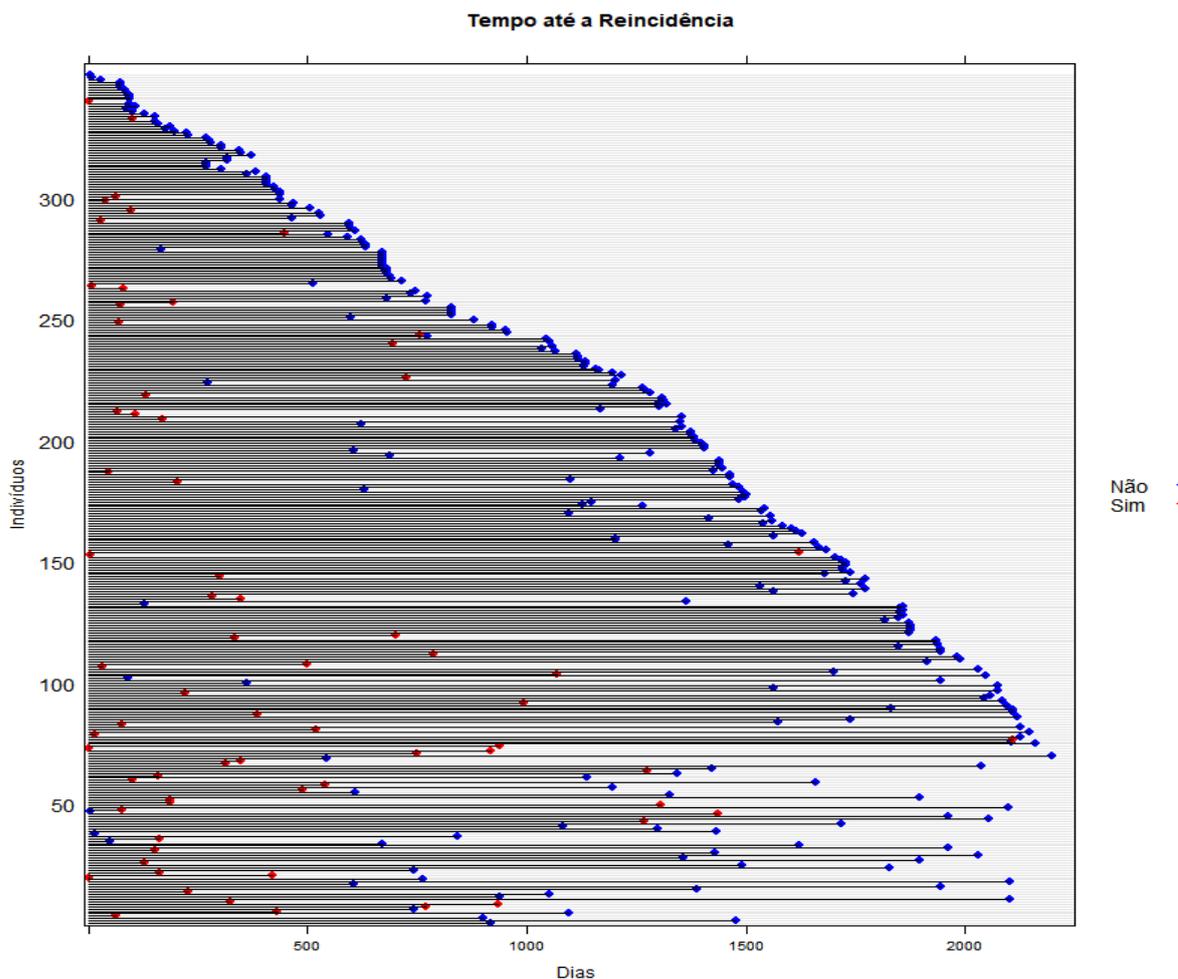
Quanto à variável cor, as maiores frequências correspondem à cor branca e parda, com 61% e 27% respectivamente. Quanto à existência de tatuagem no corpo, dos 350 indivíduos analisados, 305 (87%) deles não possuía. Os anos de delito com mais saídas registradas foram os anos de 2009 e 2010, com um total de 79 indivíduos liberados do SECAT. A variável motivo de saída, que conta com quatro categorias, mostrou que 214 indivíduos foram liberados com alvará de soltura, correspondendo a 61% dos 350 indivíduos avaliados.

O percentual de reincidência criminal também foi avaliado no estudo. Para os dados mostrados na Tabela 1, "falha" significa a reincidência do indivíduo, ou seja, o indivíduo foi liberado, mas por algum motivo voltou a ser preso durante o tempo de acompanhamento. Ao observar a variável idade, tem-se que dos 178 indivíduos na faixa etária de 24 a 40 anos, 139 são censuras (não reincidentes) e 39 são falhas (reincidentes), 75% e 25%, respectivamente. Para a variável sexo foram observados 64 homens e 4 mulheres como falhas, correspondendo a 22% e 9% dos avaliados. Para a variável tipo de crime, os crimes contra a pessoa e contra o patrimônio trazem os resultados mais significativos, dos 101 indivíduos que cometeram crime contra a pessoa, 24 (23%) voltaram a cometer crime, e dos 71 indivíduos que cometeram crime contra o patrimônio, um total de 18 (26%) também voltaram a cometer crime.

A categoria "2009 ou 2010", correspondente ao ano do primeiro delito, foi a que apresentou o maior percentual de reincidência criminal, com 27 indivíduos (34%). Dos indivíduos que se declararam de cor branca, 43 apresentaram reincidência (20%). Ainda, dos 304 que não possuíam tatuagem, 56 apresentaram reincidência, correspondendo a 19%. Quanto ao motivo de saída, a categoria "alvará de soltura" apresentou maior frequência; dos 214 indivíduos nesta categoria, observou-se 46 reincidentes (22% dos indivíduos).

Na Figura 1, tem-se a representação dos tempos desde a liberação até a reincidência ou não dos 350 indivíduos privados de liberdade no SECAT de Primeiro de Maio. Desta figura, é possível observar uma proporção elevada de não reincidentes (em torno de 81%), representados pela cor azul. Nota-se, também, que alguns indivíduos foram acompanhados por mais de 2000 dias.

Figura 1 – Tempos em dias até a reincidência ou não dos indivíduos privados de liberdade no SECAT do município de Primeiro de Maio



Fonte: O autor (2019).

Tabela 1 - Frequências absolutas e respectivos percentuais de indivíduos, falhas e censura, observados no SECAT de Primeiro de Maio

Variável	Categoria	Frequência absoluta	Frequência relativa (%)	Censura	Falha	Censura %	Falha%
Sexo	Masculino	302	86%	239	64	78%	22%
	Feminino	48	13%	44	4	91%	9%
Idade	Até 23 anos	92	27%	69	23	75%	25%
	Entre 24 a 40 anos	178	51%	139	39	78%	22%
	Acima de 40 anos	80	22%	75	5	93%	7%
Tipo de Crime	Crime contra a pessoa	101	29%	77	24	77%	23%
	Crime contra o patrimônio	70	21%	53	18	74%	26%
	Legislação específica (Drogas)	62	17%	51	11	82%	18%
	Crimes de trânsito	32	10%	29	3	91%	9%
	Legislação específica (Desarmamento)	19	5%	16	3	84%	16%
	Outros crimes	66	18%	57	9	86%	14%
Ano do primeiro delito	2009 ou 2010	79	23%	52	27	65%	34%
	2011	63	18%	50	13	79%	21%
	2012	64	19%	57	7	89%	11%
	2013	37	10%	28	9	75%	25%
	2014	62	17%	54	8	87%	13%
	2015	45	13%	42	3	93%	7%
Cor	Branca	214	61%	172	43	80%	20%
	Parda	97	27%	80	17	82%	18%
	Negra	34	10%	90	7	92%	8%
	Amarela	5	2%	4	1	80%	20%
Tatuagem	Possui	46	13%	34	12	74%	26%
	Não possui	304	87%	249	56	81%	19%
Motivo_da_saída	Alvará de soltura com ressalvas	214	61%	168	46	78%	22%
	Transferência	66	19%	48	18	72%	28%
	Pagamento de fiança ou outros	70	20%	67	3	95%	5%

Fonte: O autor (2019).

Na sequência, foi realizada uma análise univariada (uma a uma) das variáveis mostradas na Tabela 1. Para tanto, foi utilizado o estimador de Kaplan-Meier, que permite visualizar as curvas de sobrevida associadas às categorias de cada variável, auxiliando-nos a verificar a existência de associação delas com o tempo até a reincidência e, em consequência, na escolha das que são possíveis candidatas para o ajuste do modelo de Cox.

A Figura 2 apresenta as curvas de sobrevida obtidas pelo estimador de Kaplan-Meier para as variáveis: sexo, idade no 1º delito, tipo de crime cometido, cor, tatuagem, motivo de saída e ano de saída, sendo possível observar que as variáveis cor, tatuagem e ano do 1º delito possivelmente apresentam diferenças não significativas, uma vez o comportamento das curvas para as categorias dessas três variáveis não serem muito distintos.

Figura 2 - Curvas de sobrevida estimadas por Kaplan-Meier para as variáveis da Tabela 1
(continua)

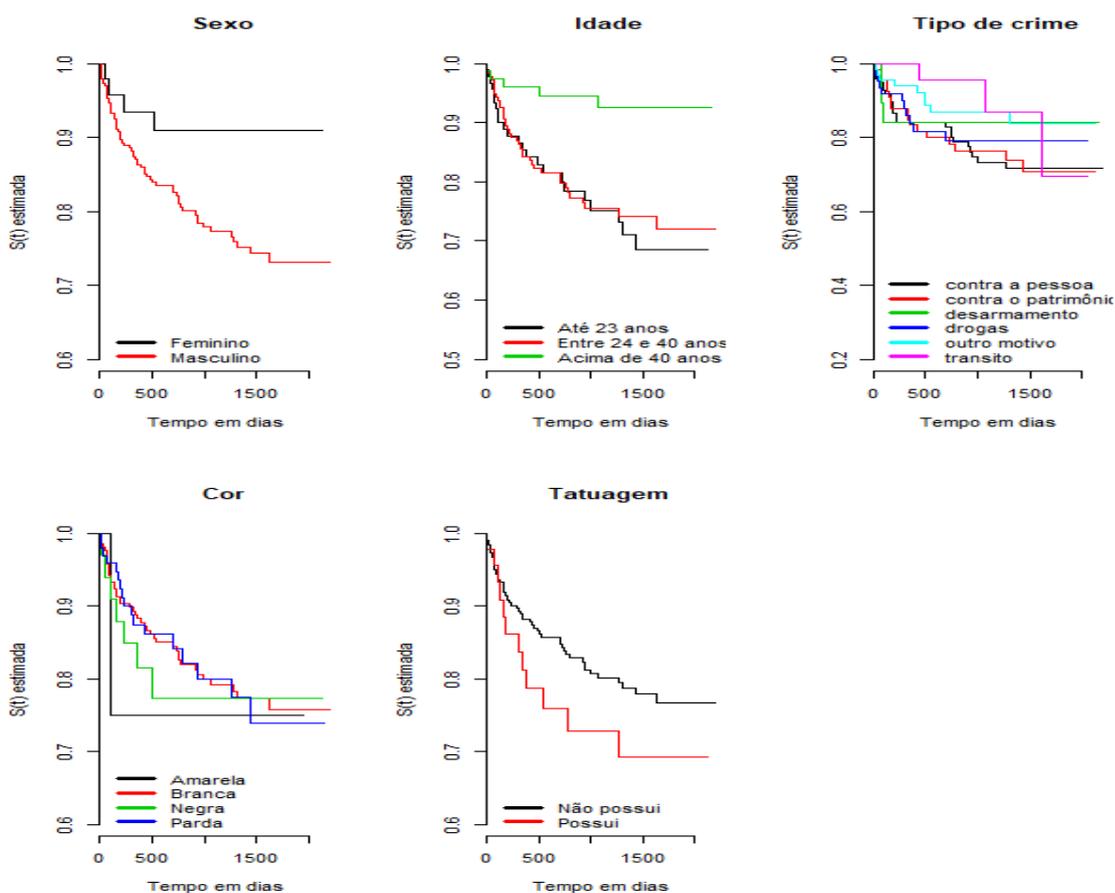
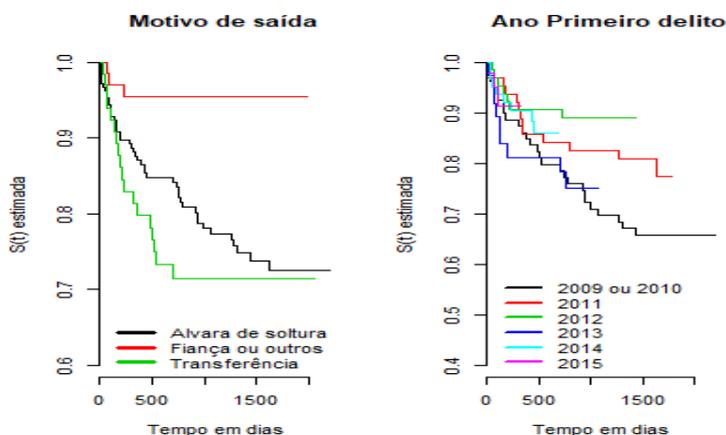


Figura 2 - Curvas de sobrevivência estimadas por Kaplan-Meier para as variáveis da Tabela 1
(conclusão)



Fonte: O autor (2019).

A seguir, foi utilizado, para cada variável, o teste *logrank* a fim de proceder a comparação entre suas respectivas curvas de sobrevivência. Na Tabela 2, estão dispostos os resultados observados, sendo possível notar que as variáveis tipo de crime, tatuagem e cor não apresentaram diferenças significativas, com valores *p* superiores a 0,10. Apesar de a variável tipo de crime apresentar um valor *p* alto, ela será mantida para o ajuste inicial do modelo de Cox, por se tratar de uma informação importante para o estudo. Dessa forma, as variáveis que são possíveis candidatas para o ajuste do modelo são: sexo, idade no 1º delito, tipo de crime, motivo de saída e ano do crime cometido.

Tabela 2 - Resultados do teste *logrank* para as variáveis disponíveis no conjunto de dados do SECAT do município de Primeiro de Maio

Variável	Estatística do teste	Valor p
Sexo	4,2	0,04
Idade	10,8	0,004
Tipo de crime	5,0	0,4
Cor	0,1	1,00
Tatuagem	1,5	0,2
Motivo de saída	9,8	0,008
Ano do primeiro delito	9,5	0,09

Fonte: O autor (2019).

4.2 Resultados do Ajuste do Modelo de Cox

Com base na análise exploratória dos dados, o modelo de Cox foi ajustado, inicialmente, considerando as covariáveis: sexo, idade no 1º delito, tipo de crime, motivo de saída e ano do primeiro delito. Com exceção da variável idade, as demais, que são categóricas, foram introduzidas no modelo por meio de variáveis *dummy*, com a primeira categoria de cada uma delas considerada como a "categoria de referência", sendo elas: sexo feminino, crime contra a pessoa, ano 2009 ou 2010 e motivo de saída = alvará de soltura.

Nesta primeira etapa, foi ajustado um modelo com todas as covariáveis, avaliando as que são significativas ao nível de significância de 0,10. A partir da Tabela 3, observam-se as estimativas associadas ao modelo de Cox, bem como os respectivos valores p, que mostram que a covariável "tipo de crime cometido" não apresentou efeito significativo (valor p > 0,10 para todas as categorias), e que a covariável "ano do primeiro delito" apresentou valor p significativo apenas para uma de suas categorias (ano de 2012).

Tabela 3 - Estimativas dos parâmetros associados ao modelo de Cox ajustado com todas as covariáveis indicadas na análise exploratória

Parâmetro	Coefficiente	Erro padrão	Valor p
Idade	-0,0446	0,01351	0,00096
Sexo			
Masculino	1,21308	0,54191	0,02519
Tipo de crime cometido			
Contra o patrimônio	-0,09168	0,32523	0,77802
Desarmamento	0,31092	0,68449	0,64966
Drogas	-0,15083	0,40419	0,70903
Outros motivos	-0,55252	0,39610	0,16304
Trânsito	-0,45192	0,63044	0,47348
Ano do primeiro delito			
2011	-0,43098	0,34955	0,21760
2012	-1,00143	0,43032	0,01996
2013	0,08063	0,41886	0,84734
2014	-0,20505	0,43064	0,63397
2015	-0,27048	0,64346	0,67423
Motivo de saída			
Pagamento de fiança ou outros	-1,48959	0,65188	0,02231
Transferência	0,29371	0,29398	0,31776

Fonte: O autor (2019).

A Tabela 4 apresenta as estimativas do modelo de Cox resultante do processo de seleção *stepwise*, que leva em consideração o valor do Critério de Informação de Akaike (AIC) e o valor p associado ao teste de Wald. Esse modelo foi considerado como o modelo final, apresentando as seguintes variáveis com efeito significativo: idade no 1º delito, sexo e motivo de saída do SECAT.

Tabela 4 - Estimativas dos parâmetros referentes ao modelo final de Cox ajustado aos dados de indivíduos do SECAT de Primeiro de Maio

Variável	Coefficiente	Erro-padrão	Valor p
Idade	-0,04215	0,01322	0,00143
Sexo			
Masculino	1,39205	0,51928	0,00735
Motivo de saída			
Pagamento de fiança ou outros	-1,52497	0,59755	0,01071
Transferência	0,25918	0,27868	0,35236

Fonte: O autor (2019).

Como o Critério de Informação de Akaike (AIC) é importante na escolha do modelo, espera-se que ele seja menor com a inclusão de covariáveis. A Tabela 5 mostra o AIC dos modelos somente com o intercepto e com as covariáveis, indicando que as covariáveis ajudam a explicar a variável resposta, uma vez que o AIC penaliza a inclusão de covariáveis desnecessárias no modelo.

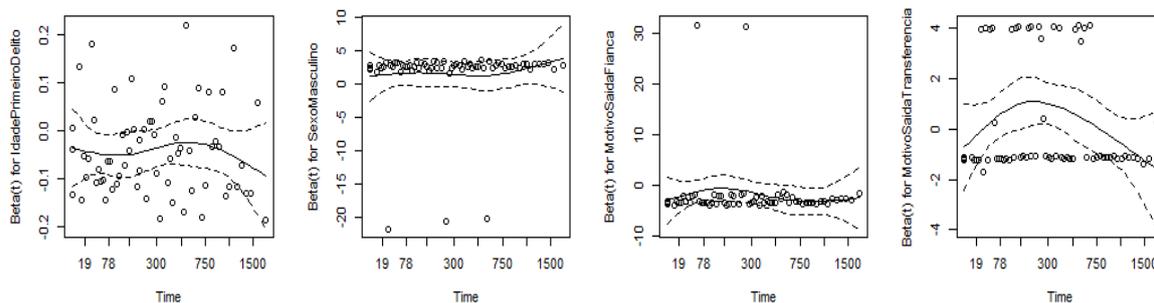
Tabela 5 – AIC dos modelos com e sem as covariáveis

Modelo de Cox	AIC
Somente com intercepto	743,02
Intercepto e covariáveis	718,44

Fonte: O autor (2019).

4.3 Análise de Resíduos

Na Figura 3, tem-se os gráficos dos resíduos padronizados de Schoenfeld, em que é possível notar a ausência de tendências muito acentuadas no decorrer do tempo, indicando que os coeficientes associados às covariáveis presentes no modelo não apresentaram efeito variando no tempo, o que indica evidências a favor suposição de proporcionalidade dos riscos.

Figura 3 - Resíduos padronizados de Schoenfeld do modelo de Cox ajustado

Fonte: O autor (2019).

Os resultados do coeficiente de correlação e dos testes associados aos resíduos padronizados de Schoenfeld, exibidos na Tabela 6, reforçam a ausência de séria violação da suposição de taxas de falha proporcionais, uma vez que os valores de ρ estão próximos de zero e os valores p são $> 0,23$, exceto para a categoria "transferência" da covariável motivo de saída. Com isso, os resultados mostram a não rejeição da suposição de riscos proporcionais.

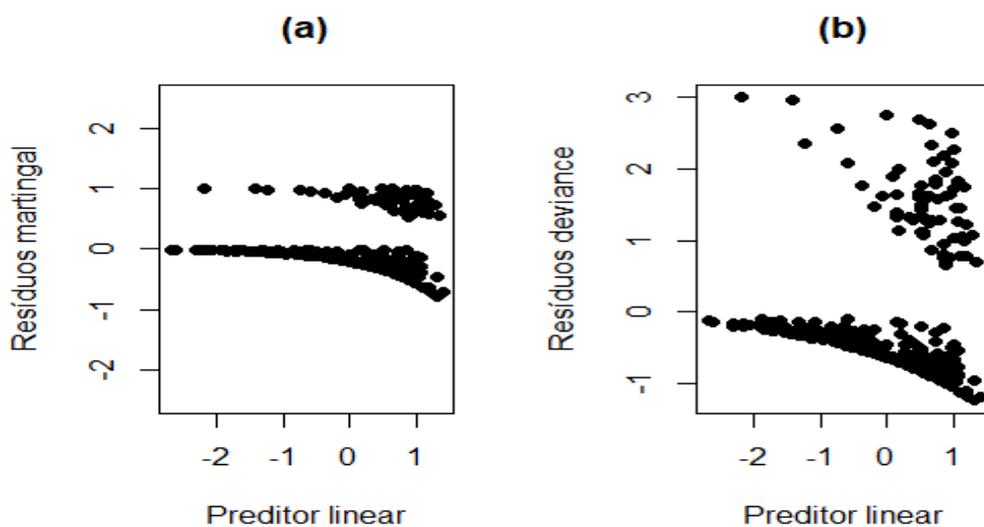
Tabela 6 – Testes da proporcionalidade dos riscos do modelo de Cox ajustado aos dados de indivíduos do SECAT de Primeiro de Maio

Parâmetro	ρ (ρ)	Chisq	Valor p
Idade	0,0123	0,0073	0,9316
Sexo	0,1282	1,1491	0,2837
Motivo de saída pagamento de fiança ou outros	-0,1436	1,4052	0,2359
Motivo de saída Transferência	-0,2066	2,8193	0,0931
Global	-	5,0935	0,2778

Fonte: O autor (2019).

Com o intuito de complementar o diagnóstico de qualidade do ajuste do modelo de Cox, tem-se na Figura 4 os gráficos dos resíduos *martingal* e *deviance* associados ao modelo. A partir da figura citada, fica visível que tais resíduos sugerem a ausência de pontos atípicos (*outliers*), pois os resíduos *martingal* e os resíduos *deviance* estão dispostos nos intervalos $(-2, 2)$ e $(-1, 3)$, respectivamente.

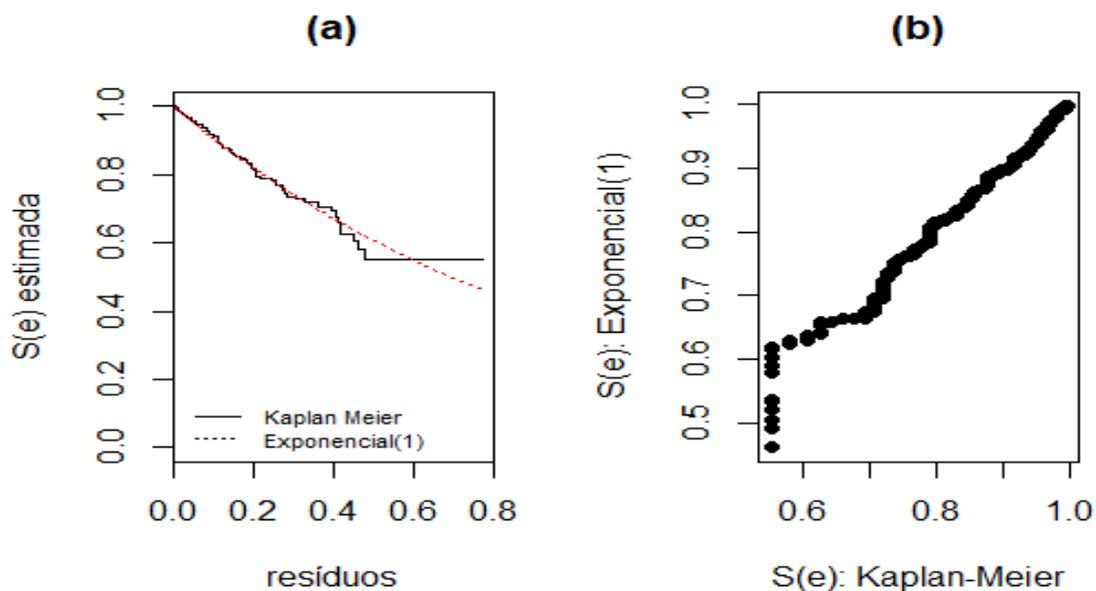
Figura 4 - Resíduos *martingal* e *deviance* associados ao modelo de Cox



Fonte: O autor (2019).

Quanto aos resíduos de Cox-Snell, nota-se, a partir dos gráficos mostrados na Figura 5, que eles seguem o comportamento de uma distribuição exponencial padrão, indicando um bom ajuste do modelo aos dados.

Figura 5 – Gráficos dos resíduos de Cox-Snell associados ao modelo de Cox



Fonte: O autor (2019).

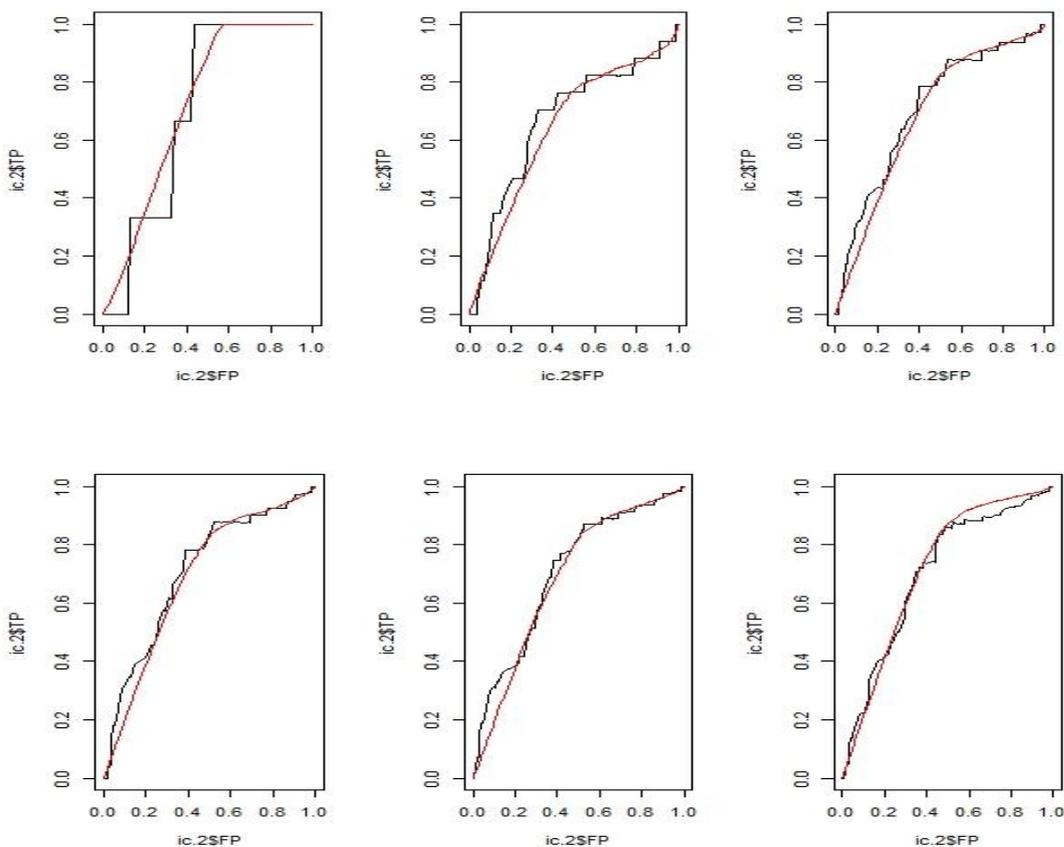
Para avaliar a qualidade de predição do modelo de Cox, foram obtidas as áreas $AUC(t)$ em alguns tempos de falha, com o auxílio do pacote *survivalROC* do *software R*. Na Tabela 7, podem ser observadas as estimativas das $AUC(t)$ nos 6 tempos $t \in [2, 1621]$ em que as áreas foram calculadas, as quais indicam qualidade de predição satisfatória do modelo. Na Figura 6, tem-se as curvas $ROC(t)$, para cada t considerado, sendo a linha vermelha estimada pelo método NNE e a preta pelo KM, ambas apresentando comportamento satisfatório.

Tabela 7 – $AUC(t)$ obtidas para avaliar a qualidade de predição do modelo de Cox

t	2	77	221	387	696	1621
$AUC(t) - NNE$	0,7188	0,6453	0,6850	0,6879	0,6854	0,7163
$AUC(t) - KM$	0,7065	0,6690	0,7059	0,7103	0,7053	0,7048

Fonte: O autor (2019).

Figura 6 – Curvas ROC em 6 tempos $t \in [2, 1621]$ associadas ao modelo de Cox



Fonte: O autor (2019).

4.4 Ajustando o Modelo de Cox com Fragilidade

Para avaliar se a inclusão de um componente aleatório (fragilidade) no modelo de Cox poderia trazer algum ganho, foi ajustado o modelo de Cox com fragilidade, considerando para cada indivíduo um efeito aleatório com distribuição Normal. A partir dos resultados dispostos na Tabela 8, nota-se valor $p = 0,41$ para o componente de fragilidade Normal, indicando heterogeneidade marginal não significativa, ou seja, o termo de fragilidade não trouxe ganho significativo quando adicionado ao modelo de Cox ajustado sem esse componente aleatório. Ainda, não foram observadas diferenças marcantes nas estimativas dos coeficientes obtidos pelos dois modelos (de Cox e de Cox com fragilidade Normal).

Tabela 8 – Estimativas obtidas para o modelo de Cox com fragilidade Normal

Parâmetro	Coefficiente	Erro padrão	Chisq	Valor p
Idade	-0,04327	0,01358	10,16	0,00140
Sexo: masculino	1,41511	0,52627	7,23	0,00720
Motivo de saída				
Fiança ou outros	-1,53538	0,60202	6,50	0,01100
Transferência	0,27619	0,28874	0,91	0,34000
Fragilidade Normal	-	-	13,09	0,41000

Fonte: O autor (2019).

Considerando, agora, efeito aleatório com distribuição Gama, foram obtidos os resultados dispostos na Tabela 9, os quais também indicaram não significância do componente de fragilidade Gama (valor $p = 0,78$).

Tabela 9 – Estimativas obtidas para o modelo de Cox com fragilidade Gama

Variáveis	Coefficiente	Erro padrão	Chisq	Valor p
Idade	-0,04215	0,01322	10,16	0,00140
Sexo: masculino	1,39206	0,51929	7,19	0,00720
Motivo de saída				
Fiança ou outros	-1,52498	0,59755	6,51	0,01100
Transferência	0,25919	0,27868	0,86	0,34000
Fragilidade Gama	-	-	0	0,78000

Fonte: O autor (2019).

Na Tabela 10, estão dispostas as $AUC(t)$ para 6 tempos $t \in [2, 1621]$ para os modelos de Cox com fragilidade Normal e Gama. Nota-se que os valores estão acima de 0,68 e não são muito distintos dos obtidos para o modelo de Cox sem fragilidade (Tabela 7), o que era esperado, tendo em vista que o componente de fragilidade não apresentou significância estatística.

Tabela 10 – $AUC(t)$ para avaliação da qualidade de predição dos modelos de Cox com fragilidade Normal e Gama ajustados aos dados do SECAT

Fragilidade	Tempos						
	2	77	221	387	696	1621	
Normal	$AUC(t) - NNE$	0,7778	0,6821	0,7267	0,7306	0,7314	0,7671
	$AUC(t) - KM$	0,8107	0,7470	0,7849	0,7881	0,7823	0,7889
Gama	$AUC(t) - NNE$	0,7187	0,6453	0,6850	0,6879	0,6854	0,7163
	$AUC(t) - KM$	0,7137	0,6761	0,7121	0,7169	0,7118	0,7128

Fonte: O autor (2019).

4.5 Análise Comparativa dos Modelos Ajustados

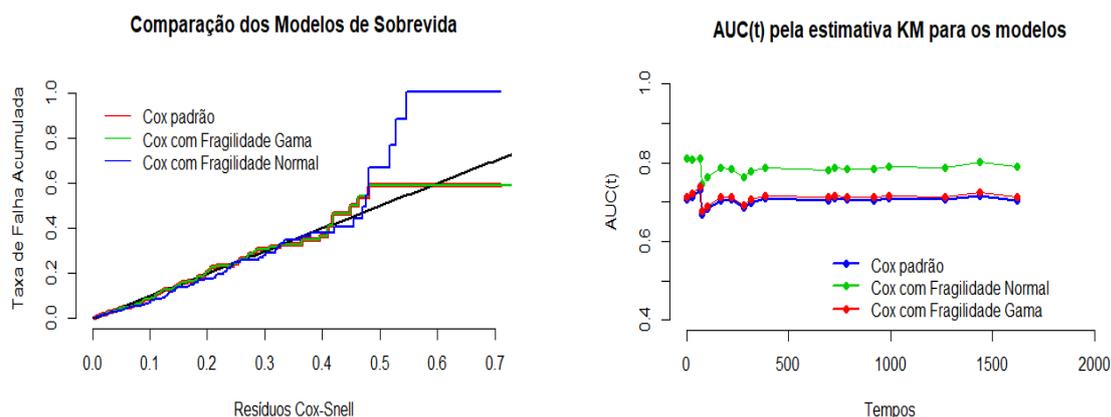
Esta seção apresenta uma comparação entre os três modelos de sobrevivência ajustados aos dados do SECAT de Primeiro de Maio. São eles, o modelo de Cox padrão e os modelos de Cox com fragilidade Gama e Normal.

Para proceder à comparação mencionada, foi analisado o gráfico da taxa de falha acumulada $\hat{\Lambda}(\hat{\epsilon}_i)$ versus $\hat{\epsilon}_i$, com $\hat{\epsilon}_i$ os resíduos de Cox-Snell, mostrado na Figura 7. Observando tal gráfico, nota-se que o modelo de Cox e o modelo de Cox com fragilidade Gama apresentaram ajustes muito semelhantes, com ambos apresentando uma proximidade mais acentuada com a reta, o que indica ajuste satisfatório desses dois modelos aos dados. Quanto à qualidade de predição, os valores das $AUC(t)$ calculadas em 18 tempos $t \in [2, 1621]$, também mostrados na Figura 7, indicaram qualidade de predição superior, embora não muito acentuada, do modelo de Cox com fragilidade Normal. Também é possível notar que os modelos de Cox e de Cox com fragilidade Gama apresentaram qualidade de predição bem semelhantes ao longo do tempo.

Apesar de o modelo de Cox com fragilidade Normal ter apresentado certa superioridade quanto à qualidade predição, o fato do componente de fragilidade

Normal não ter contribuído de forma significativa (valor $p = 0,41$), bem como o ajuste global deste modelo não ter sido o melhor (o que pode ser visualizado no gráfico à esquerda na Figura 7), optou-se por escolher o modelo de Cox padrão como sendo o mais apropriado dentre os três considerados.

Figura 7 – Resíduos $\hat{\epsilon}_i$ versus $\Lambda(\hat{\epsilon}_i)$ e AUC(t) em 18 tempos para os três modelos



Fonte: O autor (2019).

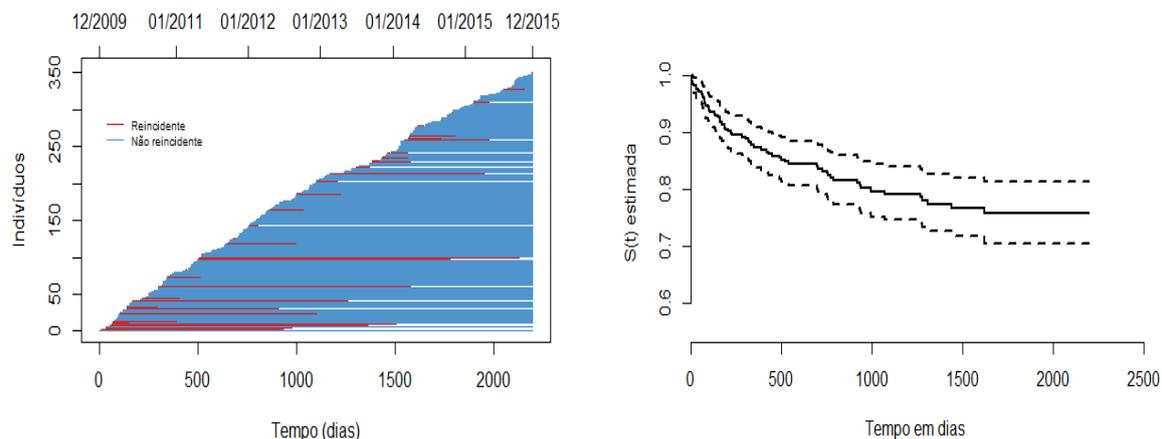
4.6 Modelo de Mistura com Fração de Imunes

Para verificar a viabilidade do modelo de mistura com fração de imunes, foram avaliadas as duas premissas mencionadas por Maller e Zhou (1996).

Quanto à primeira delas, a de que o tempo de seguimento dos indivíduos deve ser suficientemente longo, nota-se, a partir da Figura 8, que os indivíduos foram acompanhados por períodos de tempo que variaram bastante. Alguns foram acompanhados por períodos inferiores a 90 dias e outros por mais do que 2000 dias, indicando, desse modo, que os dados não satisfazem a primeira premissa para a utilização do modelo de mistura com fração de imunes.

No que se refere à segunda premissa, a de que a curva de sobrevivência apresente uma estabilização a partir de um determinado tempo, nota-se, a partir do gráfico à direita exibido na Figura 8, que a curva apresenta um comportamento decrescente, com uma "aparente" estabilização a partir do tempo 1600 dias. Tal estabilização, contudo, ocorre muito ao final da curva, o que pode gerar dúvidas quando a estar ou não de fato estabilizada a partir de 1600 dias.

Figura 8 – Representação da entrada e acompanhamento dos indivíduos (à esquerda) e curva de sobrevivência de Kaplan-Meier na ausência de covariáveis (à direita)



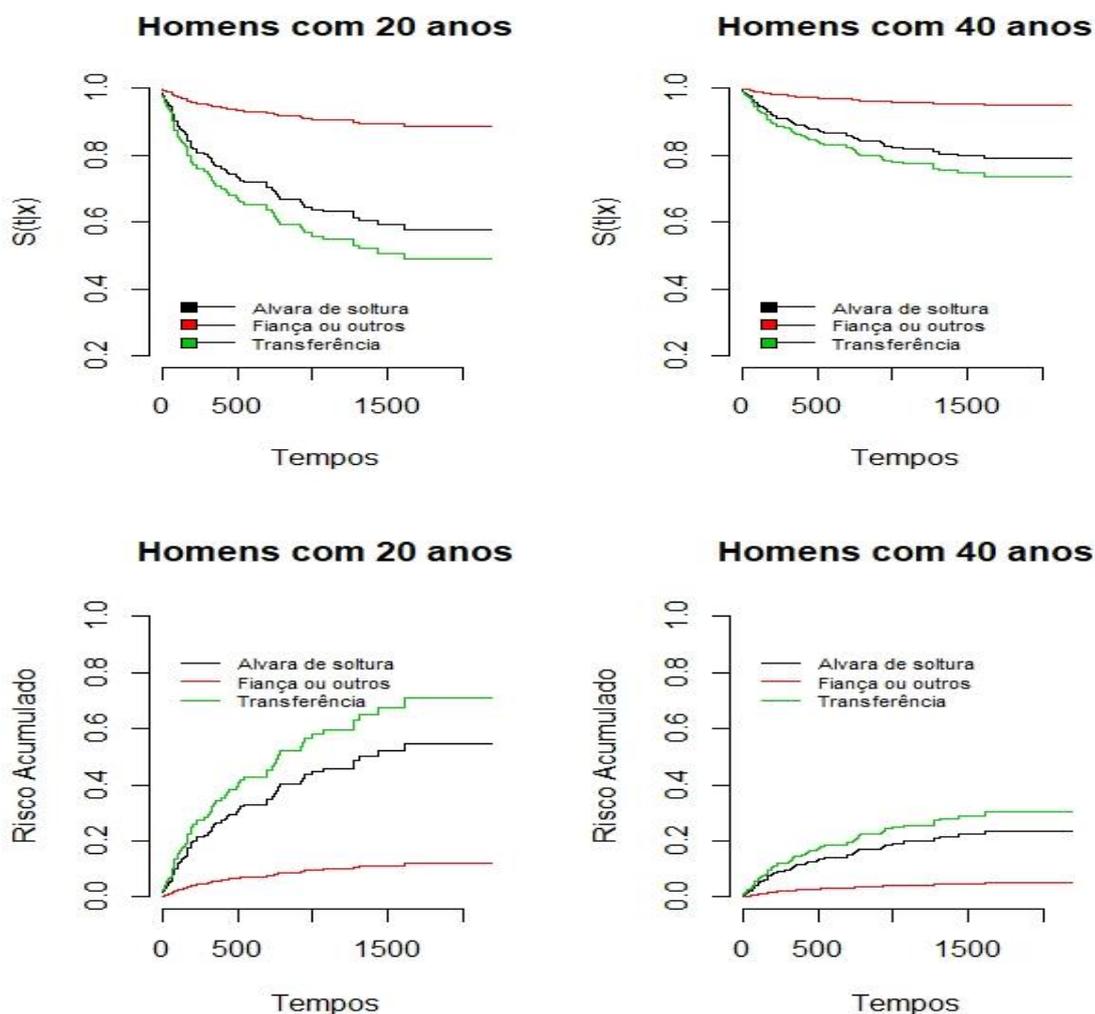
Fonte: O autor (2019).

Com base nas avaliações apresentadas, a conclusão foi a de que o modelo de mistura com fração de imunes não seria adequado para a análise dos dados e, sendo assim, o mesmo não foi utilizado neste trabalho. A escolha final foi, portanto, pelo modelo de Cox, cujas interpretações são apresentadas a seguir.

4.7 Interpretação dos Coeficientes do Modelo de Cox

Considerando indivíduos do sexo masculino com idades de 20 e 40 anos, tem-se na Figura 9 as curvas de sobrevivência e de risco acumulado estimadas a partir do modelo de Cox para os três motivos de saída (alvará de soltura, transferência e fiança/outros). Nota-se, para os motivos de saída "alvará de soltura" e "transferência", que as curvas não apresentaram diferenças muito acentuadas quando comparadas para indivíduos da mesma idade. Porém, entre indivíduos de idades diferentes, notou-se um decréscimo mais acentuado das curvas de sobrevivência para os homens de 20 anos do que para os de 40 anos, o que indica que o tempo até a reincidência ao crime é mais elevado para os homens de 40 anos. Ou seja, homens de 40 anos apresentaram probabilidade de sobreviver por mais tempo à reincidência criminal do que os de 20 anos. Dentre os três motivos de saída, "fiança/outros" foi o que apresentou probabilidade mais elevada de não ocorrência de reincidência, seja entre os homens de 20 ou de 40 anos de idade.

Figura 9 – Curvas de sobrevivência e de risco acumulado estimadas pelo modelo de Cox para homens com 20 e 40 anos de idade (por tipo de saída)

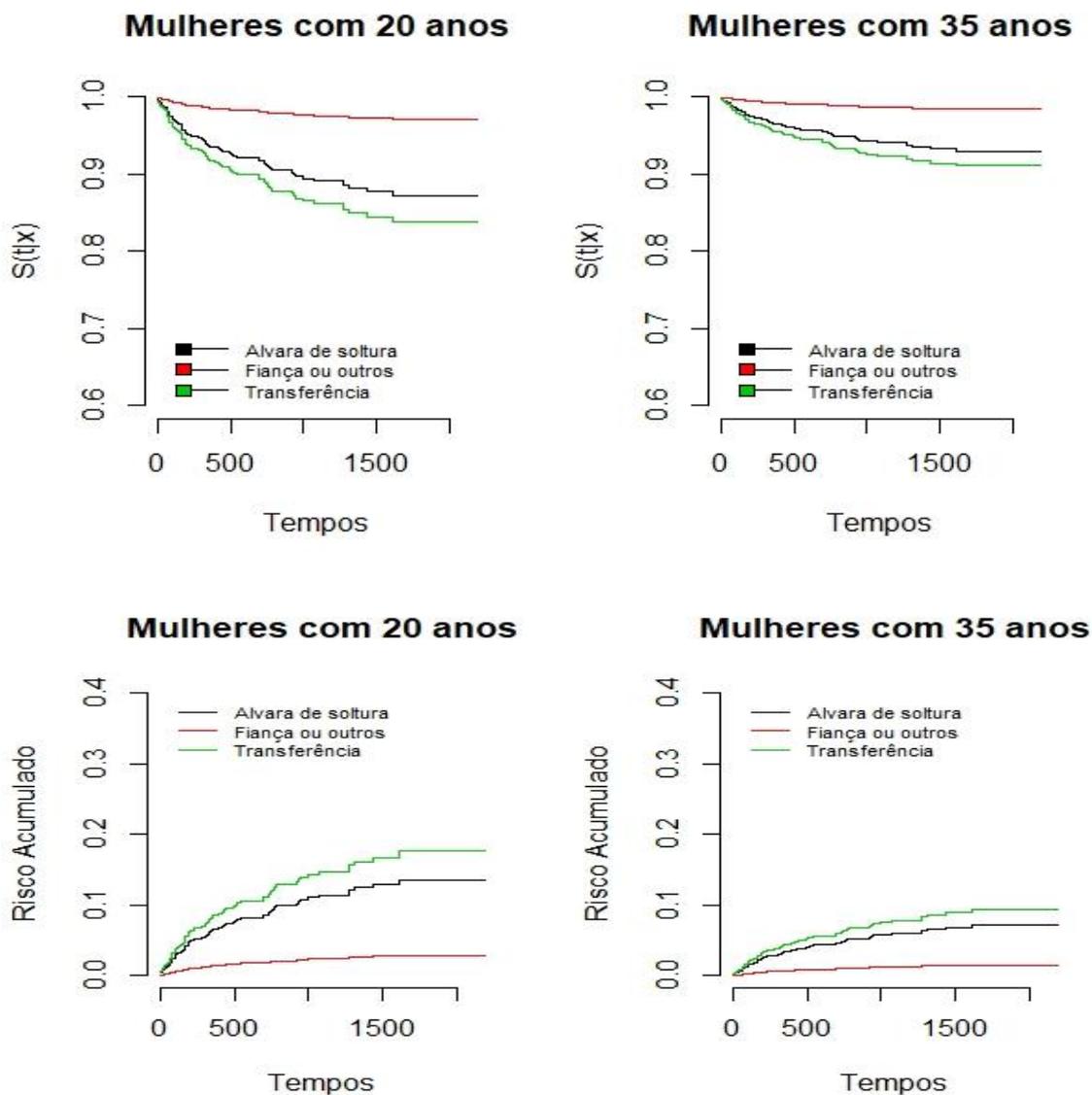


Fonte: O autor (2019).

Com base nas curvas de risco acumulado exibidas na Figura 9, bem como nas estimativas dispostas na Tabela 4, pode-se ainda concluir que se o motivo de saída de dois indivíduos do sexo masculino foi "transferência", mas um deles tem 20 anos de idade e o outro 40 anos, então o risco de reincidência ao crime do que tem 20 anos será de aproximadamente 2,32 vezes o do que tem 40 anos. Esse valor será o mesmo para homens que tenham diferença de 20 anos de idade, desde que o motivo de saída deles tenha sido o mesmo.

Fazendo, agora, o comparativo de mulheres com 20 e 35 anos de idade, nota-se, a partir das curvas exibidas na Figura 10, que há um decréscimo mais acentuado das curvas de sobrevivência das mulheres de 20 anos em relação às das com 35 anos, para os motivos de saída "alvará de soltura" e "transferência".

Figura 10 – Curvas de sobrevivência e de risco acumulado estimadas pelo modelo de Cox para mulheres de 20 e 35 anos em função do motivo de saída

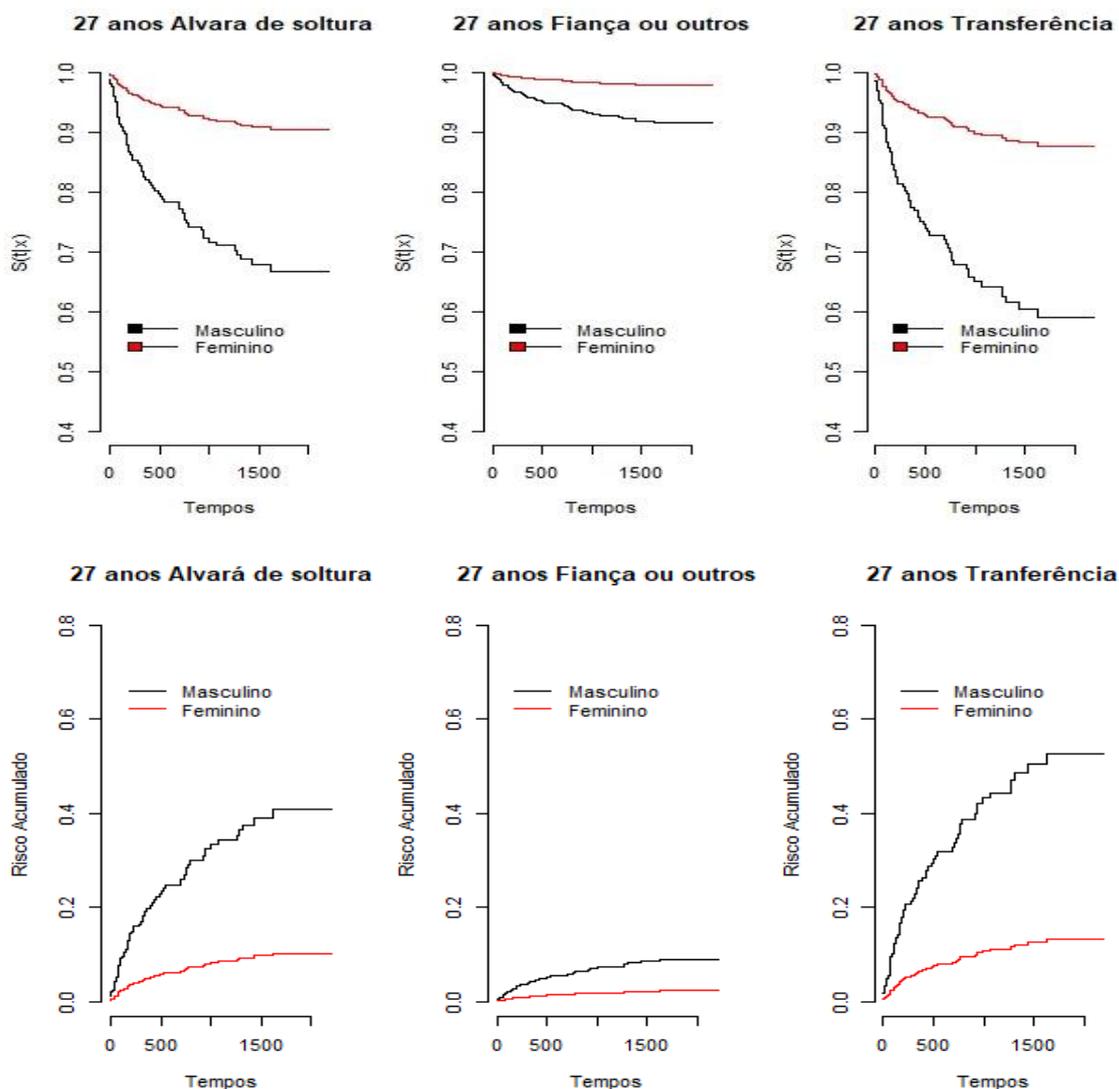


Fonte: O autor (2019).

Quanto às curvas de risco acumulado exibidas na Figura 10, é possível notar que as mulheres de 20 anos apresentaram risco de reincidência superior à das de 35 anos de idade. Ainda, das estimativas mostradas na Tabela 4, segue que as mulheres de 20 anos de idade apresentaram risco de reincidência em torno de 1,88 vezes o das com 35 anos, considerando que o motivo de saída delas tenha sido o mesmo. Nota-se que esse valor será o mesmo para mulheres que tenham diferença de 15 anos de idade, desde que o motivo de saída delas seja o mesmo.

Na Figura 11, tem-se as curvas de sobrevivência e de risco acumulado para indivíduos com 27 anos de idade, podendo-se observar um decréscimo mais acentuado das curvas de sobrevivência para homens com motivos de saída "transferência" e "alvará de soltura". Para o motivo de saída "fiança ou outros", as curvas de sobrevivência não apresentaram declive muito acentuado com o passar do tempo. Quando comparadas as curvas entre homens e mulheres (mantendo fixos a idade e motivo de saída), nota-se que as curvas de sobrevivência para as mulheres apresentaram decréscimos menos acentuados em relação às dos homens para todos os motivos de saída, o que indica que a probabilidade de não reincidência ao crime das mulheres foi maior do que à dos homens.

Figura 11 – Curvas de sobrevivência e de risco acumulado estimadas pelo modelo de Cox para indivíduos com 27 anos de idade (por sexo e tipo de saída)

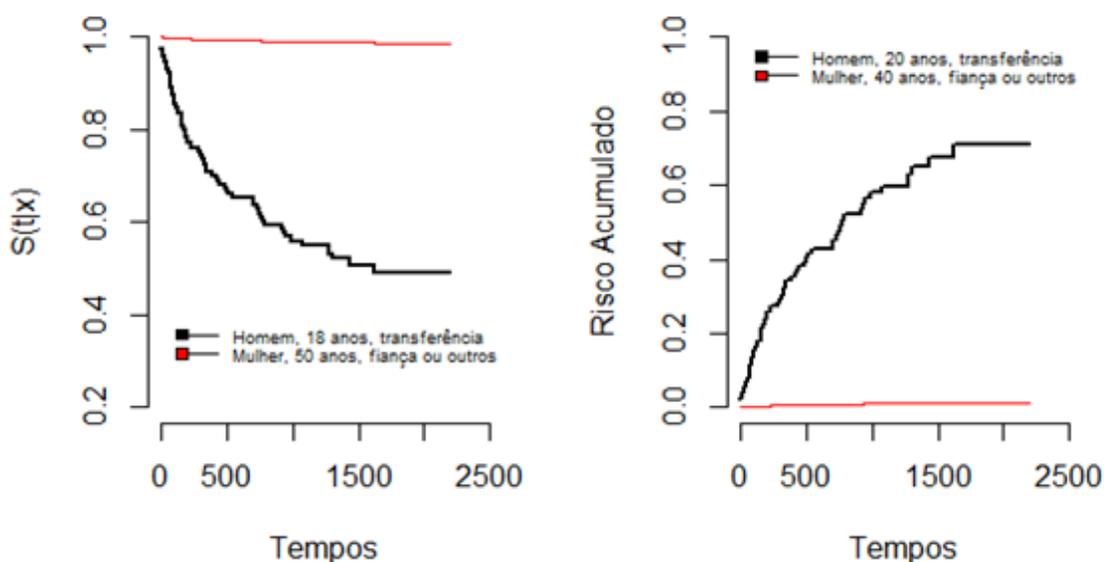


Fonte: O autor (2019).

Quanto às curvas de risco acumulado mostradas na Figura 11, nota-se que o risco de reincidência entre os homens foi maior, podendo-se concluir, a partir das estimativas mostradas na Tabela 4, que o risco de reincidência ao crime de homens com 27 anos foi 4,02 vezes o das mulheres com essa mesma idade, considerando que o motivo de saída deles tenha sido o mesmo. Por outro lado, comparando os motivos de saída entre si (para indivíduos do mesmo sexo e idade), tem-se, comparando o motivo "alvará de soltura" em relação à "fiança ou outros", que o risco de reincidência dos que saíram devido ao alvará de soltura foi 4,595 vezes o dos que saíram devido à fiança ou outros. Já para os que saíram devido à "transferência", o risco de reincidência foi 1,29 vezes o dos que saíram por "alvará de soltura e, por fim, o risco de reincidência dos que foram soltos por "transferência" foi 5,92 vezes o dos que foram soltos devido ao "pagamento de fiança ou outros".

Para finalizar, foram comparadas as curvas de sobrevivência e de risco acumulado de dois indivíduos com perfis distintos: indivíduo 1 do sexo masculino, com 20 anos de idade e motivo de saída "transferência", e indivíduo 2 do sexo feminino, com 40 anos de idade e motivo de saída "fiança". As curvas exibidas na Figura 12, mostram diferença visível entre as curvas desses indivíduos, sendo o risco de reincidência do indivíduo 1 em torno de 56 vezes o do indivíduo 2.

Figura 12 – Curvas de sobrevivência e de risco acumulado estimadas pelo modelo de Cox para dois indivíduos com perfis distintos



Fonte: O autor (2019).

5 CONSIDERAÇÕES FINAIS

A coleta de dados sobre o Sistema Prisional Brasileiro ainda é falha. O projeto do Departamento Penitenciário Nacional (DEPEN), intitulado "Dados Abertos", que consiste em unificar todos os dados penais dispostos no Brasil, ainda está em progresso. Ele significa mais acesso às informações que ajudarão a entender de várias maneiras o comportamento de indivíduos privados de liberdade e suas ações. Na base de dados utilizada neste trabalho, por exemplo, algumas variáveis não chegaram a ser analisadas devido a não se ter informação disponível para elas (por exemplo, profissão e escolaridade).

Enfim, as análises realizadas neste trabalho, com base nas informações que estavam disponíveis, tiveram como principal objetivo analisar o tempo até a reincidência ao crime de indivíduos privados de liberdade no SECAT de Primeiro de Maio, um município do interior do Paraná. Para isso, foram aplicadas técnicas estatísticas com o intuito de ajustar o melhor modelo aos dados e, a partir dele, extrair conclusões relevantes. Ao todo, foram ajustados três modelos no contexto de análise de sobrevivência, que foram discutidos no decorrer do trabalho.

Os modelos de regressão considerados foram: modelo de Cox e modelos de Cox com fragilidade Gama e Normal, os dois últimos foram ajustados para avaliar se a inclusão de um componente de fragilidade (efeito aleatório) no modelo de Cox poderia trazer alguma contribuição relevante, o que não ocorreu. Por fim, discutiu-se a viabilidade do uso do modelo de mistura com fração de imunes, tendo em vista o percentual elevado de não reincidentes registrado para os dados analisados ($\approx 81\%$). Após avaliação das premissas recomendadas por Maller e Zhou (1996) para uso do modelo citado, optou-se por não o utilizar, pois ele não atendeu as premissas recomendadas.

Ao final, o modelo de regressão de Cox, com as covariáveis sexo, idade no 1º delito e motivo de saída, foi o modelo selecionado dentre os considerados para a análise dos dados. Com base neste modelo, foram extraídas várias conclusões a respeito do risco de reincidência ao crime dos indivíduos, em função dos seus respectivos perfis. Foram comparados, por exemplo, i) indivíduos com idades distintas no primeiro delito, mas que apresentaram mesmo sexo e mesmo motivo de soltura; ii) indivíduos do mesmo sexo e com idades distintas no primeiro delito, mas com motivos de saída distintos; dentre outros.

De modo geral, os homens apresentaram risco maior de reincidência ao crime do que as mulheres. Risco maior de reincidência ao crime também foi observado quando o motivo de saída foi transferência, seguido dos motivos alvará de soltura e fiança/outros, nesta ordem. Os indivíduos mais jovens também apresentaram risco maior de reincidência ao crime.

Vale ressaltar, que os resultados obtidos para os dados analisados neste trabalho não devem ser inferidos para outras populações, tendo em vista as características específicas e diferentes de cada região, bem como dos indivíduos que as habitam, entre outros.

REFERÊNCIAS

- AKAIKE, H. A. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, London, v. 19, p.716-723, 1974.
- AKRITAS, M. G. Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics*, Beachwood, v.22, p. 1299-1327,1994.
- BARATTA, A. Criminologia crítica e política penal alternativa. *Revista de Direito Penal*. Rio de Janeiro, n. 23, p. 7-21, jul./dez. 1976.
- BRESLOW, N. Contribuição à discussão do artigo de D. R. Cox. *Journal of the Royal Statistical Society B*, Chichester, v. 34, p. 216-217, 1972.
- BRESLOW, N.; CROWLEY, J. A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, Beachwood, v.2, p. 437-453, 1974.
- CONSELHO NACIONAL DE JUSTIÇA (CNJ). Relatório Mensal do Cadastro Nacional de Inspeções nos Estabelecimentos Penais, 2017. Disponível em: <https://http://www.cnj.jus.br/inspecao_penal/mapa.php>. Acesso: 20 mai. 2017.
- CONSELHO NACIONAL DE JUSTIÇA (CNJ). *Levantamento de presos provisórios*. Disponível em: <<http://www.cnj.jus.br/noticias/cnj/84371-levantamento-dos-presos-provisorios-do-pais-e-plano-de-acao-distribunais>> Acesso: 15 nov. 2017.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Blucher, 2006. 392p.
- CORBIÈRE, F.; JOLY, P. A SAS macro for parametric and semiparametric mixture cure models. *Computer Method and Programs in Biomedicine*, v.83, n.2, p.173-180,2007
- COX, D. R. Partial Likelihood. *Biometrika*, v. 62, n. 2, p. 269-276, 1975.
- COX, D. R. Regression models and life tables. *Journal Royal Statistical Society, Series B*, v. 34, n. 2, p. 187-220, 1972.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society, Series B*, Chichester, v. 30, p. 248-275, 1968.
- EFRON, B. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, Alexandria, v.72, p. 557-565, 1977.
- GRAMBSCH, P. M.; THERNEAU, T. M. Proportionalhazard tests and diagnostics based on weighted residuals, *Biometrics*, Washington, v. 81, p. 515-526, 1994.
- HEAGERTY, P. J.; ZHENG, Y. Survival model predictive accuracy and ROC curves. *Biometrics*, Washington, v. 61, p. 92-105, 2005.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). *Reincidência Criminal no Brasil*. Rio de Janeiro: IPEA, 2015.

JUSTIÇA FEDERAL DO PARANÁ. *Transparência Pública*, 2016. Disponível em: <https://www2.trf4.jus.br/trf4/controlador.php?acao=pagina_visualizar&id_pagina=634>. Acesso: 15 nov. 2017.

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

LAWLESS, J. F. *Statistical models and methods for lifetime data*. New York: John Wiley, 1982. 580 p.

MALLER, R. A.; ZHOU, X. *Survival analysis with long-term survivors*. Chichester: John Wiley & Sons, 1996.

MANTEL, N.; HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, Washington, v. 22, p. 719-748, 1959.

NETO, N. *Sistema penitenciário brasileiro: a falibilidade da prisão no tocante ao seu papel ressocializador*. Disponível em: <<https://jus.com.br/artigos/24073/sistema-penitenciario-brasileiro-a-falibilidade-da-prisao-no-tocante-ao-seu-papel-ressocializador>>. Acesso: 27 out. 2017.

R CORE TEAM. *R: A language and environment for statistical computing*. Vienna, Austria, 2018. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.

SCHOENFELD, D. Partial residuals for the proportional hazard regression model. *Biometrika*, v. 69, p. 239-241, 1982.

SUPPORT MINITAB. *Fundamentos de regressão stepwise*, 2019. Disponível em: <<https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/basics-of-stepwise-regression/>>.

ZANON, M. H. P.; BARROS, E. A. C. Análise de sobrevivência aplicada a dados de reincidência ao crime. *Rev. Bras. Biom.*, v.36, n.3, p.527-542, 2018.

APÊNDICES

APÊNDICE A – Códigos em R utilizados para a análise dos dados

Pacotes necessários

```
library(lattice)
library(latticeExtra)
library(MASS)
library(survival)
```

dotplot com dados completo

```
my.settings <- list(
strip.background = list(col = c("gray88", "gray74", "gray48")),
strip.border = list(col = "black"),
superpose.symbol = list(pch = c(20, 20), col = c(4, 2), pch = 1.0), cex = 1.5)

dotplot(dados$id ~ dados$di + dados$TempoReincidencia ,
scales=list(y = list(tick.number=10, col='black', cex=1.2, alternating=1)),
ylim=c(0, length(dados$id) + 5 ), xlim=c(-20, 2250),
group = dados$Censura, data = dados, col = c(4, 2),
as.table = TRUE, layout = c(1, 1),
ylab = "Indivíduos",
xlab = "Dias \n", pch = c(20, 20), cex = 1.5,
main="Tempo até a Reincidência",
par.settings = my.settings,
par.strip.text = list(col="black", font = 2),
auto.key = list(text = c("Censura","Falha"),
space = "right",
columns = 1,
cex = 1.2),
between = list(x = c(.5, .5, 1.5), y = 1)) +
as.layer(segplot( dados$id ~ dados$di + dados$TempoReincidencia ,
as.table = TRUE, data = dados,
pch = c(20, 20), col = c(1),
layout = c(1, 1),axes=FALSE ))
```

dotplot com dados por sexo

Sexo Feminino

```
my.settings <- list(
strip.background = list(col = c("gray88", "gray74", "gray48")),
strip.border = list(col = "black"),
superpose.symbol = list(pch = c(20, 20), col = c(4, 2), pch = 1.0), cex = 1.5)

df<-subset(dados, subset= Sexo %in% c("Feminino"))
df$id<-c(1:dim(df)[1])
dotplot(id ~ di + TempoReincidencia,
ylim=c(-3, length(df$id) + 3 ), xlim=c(-100, 2250),
group = Censura, data = df, col = c(4, 2),
as.table = TRUE, layout = c(1, 1),
ylab = "Indivíduos",
xlab = "Dias \n", pch = c(20, 20), cex = 1.2,
main="Tempo de Reincidência ao Crime - Feminino",
```

```

par.settings = my.settings,
par.strip.text = list(col="black", font = 2),
auto.key = list(text = c("Não","Sim"),
space = "right",
columns = 1,
cex = 1.2),
between = list(x = c(.5, .5, 1.5), y = 1)) +
as.layer(segplot(id ~ di + TempoReincidencia,
as.table = TRUE, data = df,
pch = c(20, 20), col = c(1),
layout = c(1, 1) ))

```

Sexo Masculino

```

dm<-subset(dados, subset= Sexo %in% c("Masculino"))
dm$id<-c(1:dim(dm)[1])
dotplot(id ~ di + TempoReincidencia,
ylim=c(-3, length(dm$id) + 3 ), xlim=c(-10, 2250),
group = Censura, data = dm, col = c(4, 2),
as.table = TRUE, layout = c(1, 1),
ylab = "Individuos",
xlab = "Dias \n", pch = c(20, 20), cex = 1.2,
main="Tempo de Reincidência ao Crime - Masculino",
par.settings = my.settings,
par.strip.text = list(col="black", font = 2),
auto.key = list(text = c("Não","Sim"),
space = "right",
columns = 1,
cex = 1.2),
between = list(x = c(.5, .5, 1.5), y = 1)) +
as.layer(segplot(id ~ di + TempoReincidencia,
as.table = TRUE, data = dm,
pch = c(20, 20), col = c(1),
layout = c(1, 1)))

```

ano 2009

```

par(mfrow=c(2,2))
da2009<-subset(dados, subset= anosaida %in% c("2009"))
da2009$id<-c(1:dim(da2009)[1])
dotplot(id ~ di + TempoReincidencia,
ylim=c(-3, length(da2009$id)+ 3 ), xlim=c(-100, 2500),
group = Censura, data = da2009, col = c(4, 2),
as.table = TRUE, layout = c(1, 1),
ylab = "Individuos",
xlab = "Dias \n", pch = c(20, 20), cex = 1.2,
main=" Reincidência ao Crime em 2009",
par.settings = my.settings,
par.strip.text = list(col="black", font = 2),
auto.key = list(text = c("Não","Sim"),
space = "right",
columns = 1,
cex = 1.2),
between = list(x = c(.5, .5, 1.5), y = 1)) +
as.layer(segplot(id ~ di + TempoReincidencia,
as.table = TRUE, data = da2009,
pch = c(20, 20), col = c(1),
layout = c(1, 1)))

```

ano 2009 e 2010

```

da2016<-subset(dados, subset= anosaida %in% c("2009", "2010"))
da2016$id<-c(1:dim(da2016)[1])
dotplot(id ~ di + TempoReincidencia,
ylim=c(-3, length(da2016$id) + 3 ), xlim=c(-50, 2300),
group = Censura, data = da2016, col = c(4, 2),
as.table = TRUE, layout = c(1, 1),
ylab = "Individuos",
xlab = "Dias \n", pch = c(20, 20), cex = 1.2,
main="Reincidência ao Crime em 2009 e 2010",
par.settings = my.settings,
par.strip.text = list(col="black", font = 2),
auto.key = list(text = c("Não","Sim"),
space = "right",
columns = 1,
cex = 1.2),
between = list(x = c(.5, .5, 1.5), y = 1)) +
as.layer(segplot(id ~ di + TempoReincidencia,
as.table = TRUE, data = da2016,
pch = c(20, 20), col = c(1),
layout = c(1, 1)))

```

Ajuste do modelo - stepwise

```

modelo <- coxph(Surv(TempoReincidencia , Censura ) ~ 1, data = dados,
method = "breslow")
step(modelo,~ Sexo + IdadePrimeiroDelito + Tatuagem + Crimenovo + MotivoSaida
+ anoprimeirodelito, direction = "both", teste="Chisq")

```

ajustando modelo sem ano de delito para comparações

```

modelofinal <- coxph(Surv(TempoReincidencia , Censura ) ~ IdadePrimeiroDelito
+ Sexo + MotivoSaida, data = dados, method = "breslow")
summary(modelofinal)
anova(modelo,modelofinal)
cox.zph(modelofinal, transform="identity")
plot(cox.zph(modelofinal))

```

AUC(t)

```

AUC<-matrix(0,6,3)
require(survivalROC)
par(mfrow=c(1,6))

for(i in 1:6){
cutoff <- cut[i]
ic.1= survivalROC(Stime= dados$TempoReincidencia, status=dados$Censura,
marker = pi_cox ,predict.time = cutoff, method="NNE", lambda=0.02)
AUC[i,1]<-ic.1$AUC
ic.2= survivalROC(Stime=dados$TempoReincidencia , status=dados$Censura,
marker = pi_cox, predict.time = cutoff, method="KM")
AUC[i,2]<-ic.2$AUC; AUC[i,3]<-cut[i]
plot(ic.2$FP,ic.2$TP, type="l", xlim=c(0,1), ylim=c(0,1))
lines(ic.1$FP,ic.1$TP, type="l", xlim=c(0,1), ylim=c(0,1), col=2)
}
colnames(AUC)<-c("AUC_NNE", "AUC_KM", "t"); AUC

```

Análise dos Resíduos

Resíduos de Schoenfeld

```
resid(modelofinal, type="scaledsch")
cox.zph(modelofinal, transform="identity")
```

```
par(mfrow=c(1,4))
plot(cox.zph(modelofinal))
```

```
Ht<-basehaz(modelofinal, centered=F)
tempos<-Ht$time
H0<-Ht$hazard
S0<- exp(-H0)
round(cbind(tempos, S0, H0), digits=5)
```

Resíduos martingal

```
par(mfrow=c(1,2))
residmar <- resid(modelofinal, type = "martingale")
```

Resíduos deviance

```
residev<-resid(modelofinal, type="deviance")
```

Gráficos dos resíduos

```
pa<- modelofinal$linear.predictors
plot(pa, residmar, xlab="Preditor linear", ylab="Resíduo martingal",
      pch=16,ylim=c(-2.0,2.0))
plot(pa, residev, xlab="Preditor linear", ylab="Resíduo deviance" ,
      pch=16, ylim=c(-2.0,4.0))
```

Análise gráfica dos Resíduos de Cox-Snell

```
res <- dados$Censura - residmar
ekm <- survfit(Surv(res, dados$Censura)~1)
```

```
par(mfrow=c(1,2))
```

```
plot(ekm, mark.time=F, conf.int=F, xlab="resíduos", ylab="S(e) estimada")
res<-sort(res)
exp1<-exp(-res)
lines(res, exp1, lty=3)
legend(0.2, 0.2, lty=c(1,3), c("Kaplan Meier","Exponencial(1)"),
      lwd=1, bty="n", cex=0.7)
```

```
st<-ekm$surv
t<-ekm$time
sexp1<-exp(-t)
plot(st, sexp1, xlab="S(e): Kaplan-Meier",
      ylab= "S(e): Exponencial(1)", pch=16)
```

Gráfico das curvas ROC dos modelos

```

par(mfrow=c(1,1))
t<-c(2,28,66,77,102,169,221,281,315,387,696,727,788,918,994,1269,1437,1621)

AUC1<-c(0.7065,0.7123,0.7313,0.6690,0.6825,0.7051,0.7059, 0.6862, 0.6996,
0.7103, 0.7053, 0.7106, 0.7065,0.7056,0.7098, 0.7059,0.7166,0.7048) # Cox

AUC2<-c(0.8107,0.8078,0.8118,0.7470,0.7642,0.7860,0.7849,0.7646,0.7782,0.7881,
0.7823,0.7865,0.7852,0.7844,0.7903,0.7869,0.8008,0.7889) # Normal

AUC3<-c(0.7137,0.7227,0.7386,0.6761,0.6895,0.7120,0.7121,0.6927,0.7061,0.7169,
0.7118,0.7169,0.7136,0.7125,0.7173,0.7139,0.7247,0.7128) # Gama

plot(t,AUC1,ylim=c(0.4,1),ylab="AUC(t)",xlab="Tempos", pch=16, col=4,
     main = "AUC(t) pela estimativa KM para os modelos",
     axes=FALSE,xlim=c(0,2000))
axis(1)
axis(2)
lines(t,AUC1,lty=1,col=4,lwd=2)
points(t,AUC2,pch=16, col=3)
lines(t,AUC2, lty=1, col=3,lwd=1)
points(t,AUC3,pch=16, col=2)
lines(t,AUC3, lty=1, col=2,lwd=1)

legend(700,0.6,c("Cox padrão","Cox com Fragilidade Normal",
                "Cox com Fragilidade Gama"), col=c(4,3,2),
       pch=c(16,16,16), lty=c(1,1,1), bty="n",
       cex=1.0, lwd=2)

```