

Universidade de Brasília
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Acompanhamento e análise de processos em *Shiny*

Fernando Cesar Moreira Valle
16\0046386

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Fernando Cesar Moreira Valle

Acompanhamento e análise de processos em *Shiny*

Orientador:

Prof. Dr Eduardo Monteiro de Castro Gomes

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Brasília
2020

Resumo

Por meio de parceria com a instituição da 5ª vara da justiça federal, foi realizado um projeto capaz de criar visualizações, armazenamento de dados e análises estatísticas em tempo real com o manuseio de páginas Web. Dessa maneira, utilizou-se o software R com auxílio de alguns pacotes, destacando-se o ShinyDashboard, para estruturar um layout interativo e de simples manuseio para os advogados e usuários da instituição.

O sistema acima citado é composto de 10 (dez) abas moldadas no tema de bootstrap, sendo 2 (dois) delas referentes as análises estatísticas conceituadas na análise de sobrevivência, 1 (um) aba voltada para explicação e resumo das funcionalidades da página Web, 2 (dois) menus alusivos a análise exploratória dos dados e 5 (cinco) abas direcionadas para visualização, atualização e controle dos processos administrados. Além das estruturas já citadas, também foi adicionado como medida de segurança um identificador de usuário para liberação do manuseio e acesso a página Web, pois advogados e usuários do sistema terão a total possibilidade de atualização de etapas e prazos referentes aos pleitos classificados como de interesse.

Utilizando-se de um banco de dados composto por 78.260 processos com classificação PJe (Processo Judicial eletrônico), realizou-se um conjunto de análises teóricas e descritivas voltadas à disciplina de análise de sobrevivência, ou seja, efetuou-se os cálculos referentes a probabilidade de conclusão dos pleitos, tempo médio de duração das classes processuais (dias), tempo médio estimado de sobrevida dos processos (dias), entre outras medidas. Além disso, também foram consideradas as análises de pré-requisitos do modelo melhor ajustado em relação aos dados observados no trabalho.

Sumário

1	Introdução e Justificativa	7
2	Objetivos	8
2.1	Objetivos Específicos	8
3	Metodologia	9
3.1	Função de sobrevivência	9
3.1.1	Estimador de Kaplan-Meier	10
3.1.2	Estimador de Nelson-Aalen	12
3.2	Função taxa de risco	14
3.3	Função taxa de falha acumulada	15
3.3.1	Estimador de Kaplan-Meier	16
3.3.2	Estimador de Nelson-Aalen	16
3.4	Tempo médio de vida	17
3.5	Vida média residual	18
3.6	Modelo de Regressão de Cox	18
3.6.1	Estimação dos parâmetros no modelo de Cox	19
3.6.2	Interpretação dos parâmetros no modelo de Cox	20
3.6.3	Suposição de riscos proporcionais no modelo de Cox	21
3.7	Análise de adequação ao modelo ajustado	21
3.7.1	Resíduos de Cox-Snell	21
3.7.2	Resíduos de Schoenfeld	22
4	Análise exploratória	24
4.1	Análise do sistema	42
5	Conclusão	58
6	Cronograma	59
7	Referências	59

Lista de Figuras

Figura 1 – Função de taxa de falha	15
Figura 2 – Resíduos de Cox-Snell	22
Figura 3 – Resíduos de Schoenfeld	23
Figura 4 – Frequência das classes processuais	24
Figura 5 – Frequência das classes processuais	25
Figura 6 – Percentual de frequência das classes processuais no ano de 2017	26
Figura 7 – Percentual de frequência das classes processuais no ano de 2018	26
Figura 8 – Percentual de frequência das classes processuais no ano de 2019	27
Figura 9 – Frequência de observações censuradas e não censuradas nas etapas processuais	28
Figura 10 – Boxplot das classes processuais versus tempo de duração do pleito	30
Figura 11 – Função de sobrevivência	33
Figura 12 – Função taxa de risco acumulada	34
Figura 13 – Comparação entre a função de sobrevivência e alguns ajustes do modelo . .	35
Figura 14 – Comparação entre a função de sobrevivência e os melhores ajustes do modelo	36
Figura 15 – Análise de resíduos de Cox-Snell	42
Figura 16 – Tela referente a Identificação de Usuário	44
Figura 17 – Tela referente a Análise processual	45
Figura 18 – Tela referente ao Manual de uso	46
Figura 19 – Tela referente ao Quadro resumo	47
Figura 20 – Tela referente a Visualização dos dados	49
Figura 21 – Tela referente a impressão dos dados	50
Figura 22 – Tela referente a Seleção dos prazos	51
Figura 23 – Tela referente a Atualização dos prazos	52
Figura 24 – Tela referente a Criação de dados	53
Figura 25 – Tela referente a Atualização dos dados	54
Figura 26 – Tela referente ao Dashboard	56
Figura 27 – Tela referente aos Pré-requisitos do modelo	57

Lista de tabelas

Tabela 1	–	Medidas descritivas das variáveis presentes no Boxplot	31
Tabela 2	–	Qualidade de ajuste dos modelos em relação ao Deviance	36
Tabela 3	–	Resultados gerais apresentados pelos modelos de regressão	36
Tabela 4	–	Medidas do modelo de regressão na Distribuição de Weibull	38
Tabela 5	–	Medidas do modelo de regressão na Distribuição Exponencial	39
Tabela 6	–	Medidas do modelo de regressão na Distribuição Log-Logística	40
Tabela 7	–	Medidas do modelo de regressão na Distribuição Log-Normal	41
Tabela 8	–	Cronograma - 2/2019	59
Tabela 9	–	Cronograma - 1/2020	59

1 INTRODUÇÃO E JUSTIFICATIVA

Com a necessidade de realizar julgamentos de ações voltadas para órgãos da União, fundações e empresas públicas federais que se configurem como autoras ou rés, bem como intervenientes de qualquer natureza, gerou a necessidade da criação de um órgão responsável de incubir-se de tais competências (5^a Vara da Justiça Federal, 1967). No entanto, devido ao grande volume de demandas diárias dos diversos setores internos e a ampla diversidade de tipos processuais administrados diariamente pela instituição, verificou-se ao longo dos anos a dificuldade de controlar o tempo de vida das ações regidas por este órgão.

Dessa maneira, tendo em vista que os prazos limites para resolução dos pleitos são documentados nos artigos do Código de Processo Civil ¹ núm. 297, núm.º 508, núm. 536, núm. 802, núm. 896, núm. 1.065 e núm. 1.106, buscou-se criar um sistema de *dashboard* online que fosse capaz de informar os advogados e servidores da instituição sobre o tempo de vida existente dos pleitos e suas respectivas análises estatísticas, utilizando-se de gráficos e *infobox's* para análises descritivas dos processos institucionais. Além disso, também foram estudados *plots* referentes às análises de sobrevivência para averiguação das probabilidades de término em tempo hábil ou finalizações processuais fora da etapa de conclusão sequencial, sentença do pleito, essa que se caracteriza como a última etapa de uma sequência pré-existente de fases a serem seguidas pela instituição.

Com o intuito de melhorar o controle interno da instituição em relação aos prazos referentes a cada uma das datas relativas às atividades nos pleitos (Causa civil, Busca e apreensão, Cautela fiscal, Embargo à execução, Incidentes e etc), conjuntamente com o tempo de duração das etapas processuais pré-definidas pela instituição, delimitou-se um conjunto de procedimentos a serem analisados no software de programação R com o auxílio do pacote de visualizações Web denominado Shiny, para que assim ocorra uma amenização das quebras de prazos limites no decorrer das ações. Dessa maneira, visando um estudo dos dados na teoria de análise de sobrevivência, considerou-se um grupo de censuras (pleitos que até o atual momento de análise não obtiveram a conclusão de suas atividades, ou seja, são aqueles processos que possuem seus status em aberto) para todas as observações contidas no banco de dados disponibilizado pela 5^a Vara da Justiça Federal, com a intenção de produzir um conjunto de análises que sejam capazes de evidenciar o tempo médio decorrido pelo pleito, o tempo que ainda faltaria para seu término ou até mesmo a probabilidade desde ser concluído somente quando este chegasse na etapa de Conclusão da Sentença.

¹ "Art. 297 - ação ordinária - 15 dias; Art. 802 - medida cautelar - 5 dias; Art. 896 - consignação em pagamento - 15 dias; Art. 1.065 - restauração de autos - 5 dias; Art. 1.106 - procedimentos de jurisdição voluntária - 10 dias; Lei nº 4.717/65 - art. 7º - Ação Popular - 15 dias; Sobre embargos de declaração - art. 536 do CPC - 5 dias (não estão sujeitos a preparo); Sobre apelação - art. 508 - 15 dias; Embargos - art 740; Entre outros."(MANUAL DE ROTINAS E PROCEDIMENTOS INTERNOS, 2009)

2 OBJETIVOS

Com o intuito de auxiliar a 5ª Vara da Justiça Federal a informar os servidores e advogados da casa a respeito dos prazos de respostas remanescentes para elaboração e estruturação dos processos descritos como de interesse, foi desenvolvido com parceria do departamento de Estatística da UnB um sistema Web capaz de inteirar os usuários sobre as estatísticas e informações bases dos pleitos administrados pela instituição. Dessa maneira, utilizou-se integralmente o software de programação R com auxílio de pacotes, ressaltando-se o Shiny, para desenvolvimento e estruturação dos parâmetros necessários para compilação dos resultados.

Comparativamente com o método de controle anteriormente utilizado pela instituição, planilhas em Excel, o sistema desenvolvido tem como objetivo gerar um ganho considerável em capacidade de visualização pelos usuários: análises gráficas em tempo real e maior capacidade de controle dos prazos referentes as etapas processuais. Além disso, é uma possibilidade do usuário usufruir de um layout simples e dinâmico projetado com base numa página Web.

Dentre as análises estatísticas presentes no modelo, vale ressaltar os estudos efetuados em análise de sobrevivência. Pois, estes caracterizam as medidas analisadas mais relevantes a instituição, possuindo descrições sobre as projeções do tempo médio de duração de pleitos, probabilidades de conclusão, tempos médios restantes para finalizações dos processos que ainda se encontram com etapas em aberto e etc.

2.1 OBJETIVOS ESPECÍFICOS

- Realizar um conjunto de visualizações simples e diretas que sejam capazes de informar o usuário a cerca dos tempos restantes ou tardios para conclusão dos processos;
- Gerar gráficos que mostrem o desenvolvimento dos pleitos dentro da instituição de maneira interativa e eficiente;
- Realizar um estudo descritivo e analítico na área de análise de sobrevivência utilizando o banco de dados da Justiça Federal.
- Estruturar um breve manual de uso afim de inteirar novos usuários sobre as características descritivas de cada processo analisado, conjuntamente evidenciando os prazos delimitados para as conclusões dos pleitos;
- Evidenciar medidas de como prosseguir com a utilização do software e dos códigos estruturados, disponibilizando documentação e concedendo arquivos via Github;

3 METODOLOGIA

Os dados obtidos por meio de parceria com a 5ª Vara da Justiça Federal correspondem a cinquenta e cinco diferentes tipos de classes processuais com diferentes tempos de circulação, sendo estas administradas nos seis setores internos ao órgão (Secretaria, Gabinete, Central de mandados e Requerido). Dessa maneira, para elaborar uma análise estatística mais aprofundada em relação ao banco utilizou-se dos processos classificados como Pje (Processo Judicial Eletrônico), já que estes se classificariam como processos mais recentes, com melhores imputações de valores, e não precisariam de digitalização.

Posteriormente, verificou-se a necessidade de realização de uma limpeza nas variáveis relacionadas as etapas processuais devido ao fato destas possuírem repetições nas movimentações entre os períodos de início e término das fases. Dessa maneira, possibilitou-se a estruturação de censuras à direita ² para os pleitos que até o momento da extração dos dados (dia 02/12/2019) não haviam concluído a realização de suas atividades para aquelas etapas em desenvolvimento. Assim, já delimitado os termos que serão censurados, utilizou-se da teoria de análise de sobrevivência no ramo da estatística para realizar os cálculos referentes as métricas de interesse posteriormente disponibilizadas ao sistema por meio de infoboxes visualmente simples e de fácil compreensão para o usuário final.

As principais métricas em estudo são representadas por meio da *função de sobrevivência*, *função taxa de risco*, *função taxa de falha acumulada*, *tempo médio de vida*, *vida média residual*, *Modelo de Regressão de Cox*, *Resíduos de Cox-Snell* e *Resíduos de Schoenfeld*, sendo cada uma dessas descritas detalhadamente nos tópicos abaixo.

3.1 FUNÇÃO DE SOBREVIVÊNCIA

É uma das principais funções probabilísticas utilizadas para descrever estudos na área de análise de sobrevivência, dispondo de notação descrita por $S(t)$. Esta métrica é representada como uma função que admite uma forma similar a uma escada possuindo degraus nos tempos de falha observados de tamanho $\frac{d_j}{n}$, e é estimada com base nos dados obtidos em amostras e calculada com base no número de observações que não falharam ou sofreram o evento de interesse até o período de tempo t .

$$S(t) = \frac{\text{numero de observacoes que nao falharam ate o tempo } t}{\text{numero total de observacoes}}$$

²

Ocorre quando o tempo de falha é superior ao tempo registrado, ou seja, a informação faria com que o risco de ocorrência do evento de interesse fosse superestimado, pois o tempo até a ocorrência da falha é desconhecido. No entanto, sabe-se que o evento de interesse não ocorreu até o último momento observado.

Uma maneira mais eficaz e direta de estruturar o estimador da função de sobrevivência é dado por:

$$\hat{S}(t) = \frac{n_j - d_j}{n}, \text{ para todo } t_{(j)} \leq t < t_{(j+1)}, j = 0, 1, \dots, k \quad (1)$$

Onde n_j representa o número de observações sobre risco no tempo $t_{(j)}$ (contido) e d_j retrata o número de observações que sofreram censura no tempo $t_{(j)}$, $j = 1, 2, \dots, k$

O estimador também pode ser definido como a probabilidade de um objeto em estudo não falhar até um determinado período de tempo $t_{(j)}$, ou seja, a probabilidade da observação analisada não ter sofrido censura até o tempo $t_{(j)}$. Uma das técnicas amplamente utilizadas na estimação de $S(t)$ são os Estimadores de Kaplan-Meier e Nelson-Aalen, sendo ambos métodos não paramétricos de estimação.

3.1.1 Estimador de Kaplan-Meier

O estimador não paramétrico de Kaplan-Meier proposto por Kaplan e Meier (1958), também denominado como estimador limite-produto, se define como a medida mais popular para aferição da Função de Sobrevivência, estruturando assim a estimativa de $\hat{S}_{KM}(t)$ por meio de:

$$\hat{S}_{KM}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right) \quad (2)$$

Onde n_j representa o número de observações sobre risco no tempo $t_{(j)}$ (contido) e d_j retrata o número de observações que sofreram censura no tempo $t_{(j)}$, $j = 1, 2, \dots, k$

Como consequência do estimador de Kaplan-Meier ser uma adaptação da função de sobrevivência empírica, sua formação leva como base uma sequência de passos dependentes das anteriores, ou seja, para que uma observação sobreviva até o período de tempo $t_{(j)}$ é necessário que esta tenha sobrevivido até o período $t_{(j-1)}$, isto é, (RUPERT et al., 1998)

$$S(t) = S(t_{(0)}) \prod_{j:t_{(j)} \leq t} \left(1 - P(t_{(j-1)} < T \leq t_{(j)} \mid T \geq t_{(j)})\right), \text{ para } t \geq 0 \quad (3)$$

Para obter os valores do estimador na Função de Sobrevivência considerando $S(t_{(0)}) = 1$, basta utilizar a estimativa do parâmetro de \hat{q}_j , ou seja, efetuar os cálculos em relação a chance da observação sofrer o evento de interesse no intervalo de tempo $(t_{(j-1)}, t_{(j)}]$, considerando que esta ainda se encontrava em risco no momento $t_{(j)}$, ou seja,

$$\hat{S}_{KM}(t) = \frac{d_j}{n_j} \quad (4)$$

Vale salientar que naturalmente, o estimador de Kaplan-Meier se reduz à função de Sobrevivência empírica se não existirem censuras. O estimador também apresenta esta forma em estudos envolvendo mecanismos de censuras do tipo I ou II que não assumem $\hat{S}_{KM}(t) = 0$.

Devido ao fato da função de sobrevivência de Kaplan-Meier está suscetível a variações amostrais que devem ser descritas em termos de sua variância, pode-se aplicar três diferentes modelos de intervalos de confiança ao estudo.

- **Intervalo de Confiança simétrico:**

A variância assintótica da função de sobrevivência é conhecida como fórmula de GreenWold, e pode ser obtida por meio da expressão matemática (2). Este cálculo foi efetuado em Kalbfleisch e Prentice (1980, p. 12-14):

$$\hat{Var}(\hat{S}_{KM}(t)) = \hat{S}_{KM}^2(t) \sum_{j:t(j) \leq t} \frac{d_j}{n_j(n_j - d_j)} \quad (5)$$

Devido ao fato de $\hat{S}_{KM}(t)$ convergir assintoticamente para uma distribuição normal com média $S(t)$ e variância acima explicitada, obtém-se um intervalo $100(1 - \alpha)\%$ de confiança para $S(t)$ por meio de:

$$\left[\hat{S}_{KM}(t) - Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{S}_{KM}(t))}; \hat{S}_{KM}(t) + Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{S}_{KM}(t))} \right] \quad (6)$$

Entretanto, para os valores contidos nos extremos de t , este intervalo de confiança pode expor limite inferior negativo ou limite superior maior que 1. Nesse caso, o problema é resolvido utilizando uma transformação em $\hat{S}_{KM}(t)$.

- **Intervalo de Confiança Log:**

Uma forma alternativa de efetuar um intervalo de confiança para a função de sobrevivência é estruturá-lo com base na estimativa do risco acumulado, apresentando assim sua variância estimada obtida por meio de: (MARILIA et al, 2019)

$$\hat{Var}(\hat{H}_{KM(t)}) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j(n_j - d_j)} \quad (7)$$

Onde n_j representa o número de observações sobre risco no tempo $t_{(j)}$ (contido) e d_j retrata o número de observações que sofreram censura no tempo $t_{(j)}$, $j = 1, 2, \dots, k$

Assim para todo t fixo, possui-se um intervalo $100(1 - \alpha)\%$ de confiança por meio da equação:

$$\left[e^{-Z_{(1-\frac{\alpha}{2})}\sqrt{\widehat{Var}(\hat{H}_{KM}(t))}}\hat{S}_{KM}(t); e^{Z_{(1-\frac{\alpha}{2})}\sqrt{\widehat{Var}(\hat{H}_{KM}(t))}}\hat{S}_{KM}(t) \right] \quad (8)$$

Onde a função de risco acumulada descrita por $\hat{K}_{KM}(t)$ representa a soma das probabilidades de um evento de interesse falhar até um periodo de tempo t .

- **Intervalo de Confiança Log-log:**

Baseia se na aplicação do logaritmo na função de risco acumulado, ou uma transformação "log-log" na função de sobrevivência de Kaplan-Meier, isto é:

$$\hat{L}_{KM}(t) = \log [-\log(\hat{S}_{KM}(t))] \quad (9)$$

Ocasionalmente assim uma variância dada por,

$$\widehat{Var}(\hat{L}_{KM}(t)) = \frac{1}{\left(\log(\hat{S}_{KM}(t))\right)^2} \sum_{j:t(j) \leq t} \frac{d_j}{n_j(n_j - d_j)} \quad (10)$$

Logo, para todo t fixo, assume-se um intervalo $100(1 - \alpha)\%$ de confiança por meio da fórmula

$$\left[\hat{S}_{KM}(t)^{\exp\left(Z_{(1-\frac{\alpha}{2})}\sqrt{\widehat{Var}(\hat{L}_{KM}(t))}\right)}; \hat{S}_{KM}(t)^{\exp\left(-Z_{(1-\frac{\alpha}{2})}\sqrt{\widehat{Var}(\hat{L}_{KM}(t))}\right)} \right] \quad (11)$$

3.1.2 Estimador de Nelson-Aalen

Proposto por Nelson-Aalen (1972), o estimador da função de sobrevivência baseia-se na relação entre a função de sobrevivência e a função de risco acumulado. Dessa maneira, utilizando

a equação $S(t) = \exp[-H(t)]$ assume-se que o estimador para função de sobrevivência é obtido por meio de: (COLOSIMO et al., 2006)

$$\hat{S}_{NA}(t) = \exp(-\hat{H}_{NA}(t)) = \exp\left(-\sum_{j:t(j) \leq t} \frac{d_j}{n_j}\right) \quad (12)$$

Vale salientar que o estimador $\hat{S}_{NA}(t)$ e o de Kaplan-Meier possuem na maioria das vezes suas estimativas muito próximas de $S(t)$. Bohoris(1994) demonstrou que $\hat{S}_{NA}(t) \geq \hat{S}_{KM}(t)$ para todo t . Ou seja, o estimador de Nelson-Aalen é maior ou igual que o de Kaplan-Meier.

Considerando-se o fato da função de sobrevivência de Nelson-Aalen está sujeita a variações amostrais que devem ser descritas em termos de sua variância, pode-se descrever três diferentes modelos de intervalos de confiança ao estudo.

- **Intervalo de Confiança simétrico:**

A variância assintótica do estimador na função de sobrevivência de Nelson-Aalen é obtida por meio da equação (Aalen e Johansen, 1978):

$$\hat{Var}(\hat{S}_{NA}(t)) = \hat{S}_{NA}^2(t) \sum_{j:t(j) \leq t} \frac{d_j}{n_j^2} \quad (13)$$

Onde n_j representa o número de observações sobre risco no tempo $t_{(j)}$ (contido) e d_j retrata o número de observações que sofreram censura no tempo $t_{(j)}$, $j = 1, 2, \dots, k$

Considerando-se que $\hat{S}_{NA}(t)$ converge assintoticamente para uma distribuição normal com média $S(t)$ e variância acima explicitada, obtém-se um intervalo $100(1 - \alpha)\%$ de confiança para $S(t)$ por meio de:

$$\left[\hat{S}_{NA}(t) - Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{S}_{NA}(t))}; \hat{S}_{NA}(t) + Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{S}_{NA}(t))} \right] \quad (14)$$

- **Intervalo de Confiança Log:**

Considerando que a função de risco pode ser escrita por meio da relação entre a função de sobrevivência e o risco acumulado pela fórmula $H(t) = -\log[S(t)]$, possui-se uma variação assintótica obtido por: (COLOSIMO et al., 2006)

$$\hat{Var}(\hat{H}_{NA}(t)) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j^2} \quad (15)$$

Logo, para todo t fixo, um intervalo $100(1 - \alpha)\%$ de confiança para a função de sobrevivência é dada por:

$$\left[e^{-Z_{(1-\frac{\alpha}{2})}\sqrt{\hat{Var}(\hat{H}_{NA(t)})}}\hat{S}_{NA}(t); e^{Z_{(1-\frac{\alpha}{2})}\sqrt{\hat{Var}(\hat{H}_{NA(t)})}}\hat{S}_{NA}(t) \right] \quad (16)$$

- **Intervalo de Confiança Log-log:**

Baseia-se na aplicação do logaritmo na função de risco acumulado, ou uma transformação "log-log" na função de sobrevivência de Nelson-Aalen, isto é:

$$\hat{L}_{NA}(t) = \log [-\log(\hat{S}_{NA}(t))] \quad (17)$$

Gerando-se assim em uma variância obtida por,

$$\hat{Var}(\hat{L}_{NA}(t)) = \frac{1}{\left(\log(\hat{S}_{NA}(t))\right)^2} \sum_{j:t(j) \leq t} \frac{d_j}{n_j^2} \quad (18)$$

Logo, para todo t fixo, obtêm-se um intervalo $100(1 - \alpha)\%$ de confiança para a função de sobrevivência por meio da fórmula

$$\left[\hat{S}_{NA}(t)^{\exp\left(Z_{(1-\frac{\alpha}{2})}\sqrt{\hat{Var}(\hat{L}_{NA}(t))}\right)}; \hat{S}_{NA}(t)^{\exp\left(-Z_{(1-\frac{\alpha}{2})}\sqrt{\hat{Var}(\hat{L}_{NA}(t))}\right)} \right] \quad (19)$$

3.2 FUNÇÃO TAXA DE RISCO

É a probabilidade do objeto em estudo falhar entre os intervalos de tempo $[t_1, t_2)$ na função de sobrevivência sendo que este não falhou num período de tempo anterior a t_1 , dividida pela probabilidade da função de sobrevivência em t_1 vezes o comprimento do intervalo. Assim, a taxa de falha é expressa por:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} \quad (20)$$

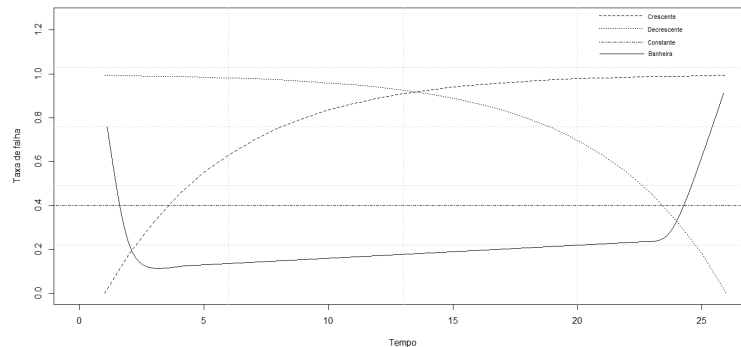
De maneira geral, pode-se assumir o intervalo como sendo $[t, t + \Delta t)$. Acarretando assim que a equação passe a assumir a forma de:

$$\lambda(t) = \frac{S(t_1) - S(t_2)}{\Delta t S(t)} \quad (21)$$

Considerando-se $\Delta(t)$ como um valor muito pequeno, $\lambda(t)$ representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o instante t . Observe que as taxas de falhas são números que assumem seus valores positivos, mas sem limite superior. A função de taxa de falha $\lambda(t)$ é bastante utilizada para enunciar a distribuição do tempo de vida das observações. (COLOSIMO et al., 2006)

Graficamente, a taxa de risco de um objeto pode vir a assumir quatro formulações diferentes entre si, sendo estas: Função crescente (indicando um aumento na taxa de falha ao longo do tempo, este comportamento representa um efeito progressivo de envelhecimento), Função constante (evidencia que a taxa não se altera com o passar do tempo, ou seja, uma constante), Função banheira (representa uma taxa de falha que é frequentemente alta inicialmente, baixa no centro e alta novamente no final da vida, ou seja, o modelo divide a vida operacional de um elemento em três estágios) e Função decrescente (demonstra uma redução da taxa de falha à medida que o tempo passa). (COLOSIMO et al., 2006)

Figura 1 – Função de taxa de falha



3.3 FUNÇÃO TAXA DE FALHA ACUMULADA

É um procedimento útil na estimação não-paramétrica e na seleção de modelos mais apropriados para determinados conjuntos de dados. A função $H(t)$ fornece o risco acumulado do indivíduo no tempo t , caso esta seja definida como uma variável aleatória contínua, é descrita por

$$H(t) = \int_0^t \lambda(u) du, t \geq 0 \quad (22)$$

A função de Taxa de falha acumulada, $H(t)$, não tem interpretação direta como visto em outras medidas mas pode ser útil na avaliação de uma função mais relevante, que seria a taxa de falha ($\lambda(t)$). Isto ocorre substancialmente na estimação não-paramétrica onde $\lambda(t)$ apresenta um

estimador com propriedades ótimas e difíceis de serem estimadas. (COLOSIMO et al., 2006)

Uma das técnicas amplamente usufruídas na estimação de $H(t)$ são os estimadores de Kaplan-Meier e Nelson-Aalen. Sendo ambos métodos não paramétricos de estimação.

3.3.1 Estimador de Kaplan-Meier

Considerando-se que a função de sobrevivência de uma variável contínua pode vir a ser expressa em relação a função taxa de falha acumulada por meio da equação: $S(t) = \exp(-H(t))$, acata-se que o estimador de Kaplan-Meier para a função de risco acumulado é obtido por meio de:

$$\hat{H}_{KM}(t) = \log [\hat{S}_{km}(t)] \quad (23)$$

Em que $\hat{S}_{KM}(t)$ é a função de sobrevivência estimada em relação ao estimador de Kaplan-Meier.

Devido ao fato da função taxa de risco acumulada em Kaplan-Meier está sujeita a variações amostrais presentes em sua variância, pode-se descrever intervalos de confiança em estudo no modelo Log.

• Intervalo de Confiança Log:

Como visto anteriormente, a função da taxa de risco acumulado pode ser descrita em relação à função de sobrevivência pela equação (23). Dessa forma, respeitando-se a variância do Estimador de Kaplan-Meier expressa em (7), obtém-se um intervalo $100(1 - \alpha)\%$ de confiança para $H(t)$ por meio de:

$$\left[\hat{H}_{KM}(t) - Z_{1-(\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{H}_{KM}(t))}; \hat{H}_{KM}(t) + Z_{1-(\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{H}_{KM}(t))} \right] \quad (24)$$

3.3.2 Estimador de Nelson-Aalen

Nelson (1972) propôs um estimador não paramétrico para a função de risco acumulado, $H(t)$, cujas propriedades assintóticas foram comprovadas por Aalen (1978). Este estimador recebe a denominação na literatura de Nelson-Aalen e é definido por: (COLOSIMO et al., 2006)

$$\hat{H}_{NA}(t) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j} \quad (25)$$

Onde n_j representa o número de observações sobre risco no tempo $t_{(j)}$ (contido) e d_j retrata o número de observações que sofreram censura no tempo $t_{(j)}$, $j = 1, 2, \dots, k$

Devido ao fato da função taxa de risco acumulada de Nelson-Aalen está sujeita a variações amostrais presentes em sua variância, pode-se descrever os intervalos de confiança em estudo no modelo Log.

- **Intervalo de Confiança Log:**

Considere que a variância assintótica da função de risco acumulado do estimador de Nelson-Aalen é obtida por meio da equação:

$$\hat{Var}(\hat{H}_{NA}(t)) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j^2} \quad (26)$$

Dessa maneira, caracteriza-se para todo t fixo um intervalo de confiança de $100(1 - \alpha)\%$ em relação a $H_{NA}(t)$ obtido por:

$$\left[\hat{H}_{NA}(t) - Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{H}_{NA}(t))}; \hat{H}_{NA}(t) + Z_{(1-\frac{\alpha}{2})} \sqrt{\hat{Var}(\hat{H}_{NA}(t))} \right] \quad (27)$$

3.4 TEMPO MÉDIO DE VIDA

É a representação da área gerada abaixo do gráfico da *Função de sobrevivência*, ou seja, o tempo médio que o objeto sobrevive sem presenciar a censura.

Considerando que T (tempo) é uma variável aleatória discreta que assume valores não negativos, então o R -ésimo momento, se existir, pode ser descrito por: (A. KAUFMANN, 1997)

$$E(T^r) = \sum_{j:t(j) \leq t}^{\infty} t^r p(t), \text{ para todo } r \geq 0 \quad (28)$$

Sendo $p(t)$ obtido por meio da relação $p(t) = S(t - 1) - S(t)$ se t for maior que zero.

Em particular, o primeiro e segundo momento da variável aleatória T podem ser expressos respectivamente por meio de sua relação com a função de sobrevivência: (A. KAUFMANN, 1997)

$$E(T) = \sum_{t=0}^{\infty} S(t) \quad (29)$$

e

$$E(T^2) = E(T) + 2 \sum_{t=1}^{\infty} tS(t) \quad (30)$$

3.5 VIDA MÉDIA RESIDUAL

É a representação do tempo médio de vida restante para os elementos de interesse no estudo. A vida média residual no tempo t é a área sob a curva de sobrevivência à direita do ponto t , dividido pelo valor da *Função de sobrevivência* neste ponto mais o valor da *Distribuição de probabilidade* também nesse ponto .

$$V(t) = E[T - t | T \geq t] = \frac{1}{p(t) + S(t)} \sum_{k=t}^{\infty} S(k), \text{ para todo } t = 0, 1, 2, \dots \quad (31)$$

Considerando que $p(t)$ é obtido por meio da relação $p(t) = S(t-1) - S(t)$ se t for maior que zero.

3.6 MODELO DE REGRESSÃO DE COX

O modelo de regressão de Cox, proposto por Cox em 1972, é sem dúvida um dos mais populares na análise de sobrevivência. Este modelo possibilita que a análise dos tempos de vida até a ocorrência da censura seja realizada considerando-se as covariáveis de interesse no estudo. Cox, assim como outros autores, propõe a modelagem dos dados de sobrevivência, na presença de covariáveis, por meio da função de risco. (GIOLO et al., 2018)

Considera-se, nesse modelo, que os tempos t_i , $i = 1, \dots, n$, são independentes entre si e que o risco do elemento i é disponibilizado por:

$$\alpha_i(t|x_i) = \alpha_0(t) \exp\{\beta' x_i\} \quad (32)$$

em que α_0 é estruturado como a função de risco base, ou seja, o risco de um elemento com covariáveis $x = 0$, β é denominado como sendo o vetor de dimensão p dos coeficientes de regressão desconhecidos e x_i é representado como um vetor de dimensão p de covariáveis observadas para o elemento i . (GIOLO et al., 2018)

O modelo de Cox, definido em (32), é conhecido como sendo semi-paramétrico por considerar que as covariáveis agem multiplicativamente no risco pela relação $g(x, \beta) = \exp\{\beta' x_i\}$ e por acatar $\alpha_0(t)$ arbitrário, ou seja, por não assumir nenhuma estrutura paramétrica em relação

à $\alpha_0(t)$. A denominação desse modelo como sendo de riscos proporcionais vem do fato da razão entre as funções de risco de dois indivíduos serem dadas por, (GIOLO et al., 2018)

$$\frac{\alpha(t|x_i)}{\alpha(t|x_l)} = \frac{\alpha_0(t)\exp\{\beta'x_i\}}{\alpha_0(t)\exp\{\beta'x_l\}} = \exp\{\beta'(x_i - x_l)\} \quad (33)$$

Os pré-requisitos básicos para o uso dos riscos proporcionais de Cox são, portanto, que as taxas de falhas atuem de maneira proporcionais. Em outras palavras, significa que se um indivíduo exibe no início do estudo um risco de falhar igual a três vezes o risco de outro elemento, então esta taxa representará o mesmo valor para qualquer tempo t do período de acompanhamento. (GIOLO et al., 2018)

As funções de taxa de falha base acumulada bem como a correspondente função de sobrevivência são também de interesse e estas relacionam-se com a função de risco base por, respectivamente,

$$H_0(t) = \int_t^0 \alpha_0(u)du \quad (34)$$

e

$$S(t|x) = \exp[-H_0(t)\exp(\beta'x)] / (S_0(t))^{\exp\{\beta'x\}} \quad (35)$$

em que,

$$S_0(t) = \exp\{-H_0(t)\} \quad (36)$$

3.6.1 Estimação dos parâmetros no modelo de Cox

O modelo de Cox é representado pelos coeficientes $\beta's$ que devem ser estimados a partir de observações presentes nas amostras. A aparição de componentes não-paramétricos impossibilita, contudo, o uso do método da máxima verossimilhança. (GIOLO et al., 2018)

Para a estimação dos vetores de parâmetros desconhecidos $\beta's$, Cox (1975) apresentou, então, a função de verossimilhança parcial, $L(\beta)$. Para a realização desta função de verossimilhança, considere uma amostra de n elementos em que $k \leq n$ diferentes falhas ocorrem nos

tempos t_1, \dots, t_k e que a chance condicional da i -ésima observação vir a sofrer uma falha no tempo t , tendo como base as observações sob risco em t_i , é: (GIOLO et al., 2018)

$$\frac{\alpha_i(t_i|x_i)}{\sum_{j \in R(t_i)} \alpha_j(t_i|x_j)} = \frac{\alpha_0(t_i) \exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \alpha_0(t_i) \exp\{\beta' x_j\}} = \frac{\exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \exp\{\beta' x_j\}} \quad (37)$$

em que $R(t_i) = R_i$ é o conjunto de todos os indivíduos ainda em risco no tempo t_i .

Assim, a função de verossimilhança a ser utilizada é obtida por meio de:

$$L(\beta) = \prod_{i=1}^n \left(\frac{\exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \exp\{\beta' x_j\}} \right)^{\delta_i} \quad (38)$$

Os elementos censurados entram na função de verossimilhança parcial $L(\beta)$ por meio do conjunto de risco $R(t_i)$ e colaboram, para esta função, somente enquanto encontram-se em risco. Os parâmetros β 's são obtidos maximizando-se o logaritmo da função $L(\beta)$, isto é, resolvendo-se o sistema de equações escores (GIOLO et al., 2018)

$$0 = \frac{\partial l(\beta)}{\partial \beta_k} = \sum_{i=1}^n \delta_i \left[x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} \exp\{\beta' x_j\}}{\sum_{j \in R(t_i)} \exp\{\beta' x_j\}} \right], \quad k = 1, \dots, p \quad (39)$$

em que $l(\beta) = \log L(\beta)$.

3.6.2 Interpretação dos parâmetros no modelo de Cox

Os coeficientes β 's no modelo de regressão de Cox, servem para medir os efeitos das covariáveis em relação a taxa de falha considerando-se que, uma covariável pode acelerar, ou desacelerar, a função de risco. (GIOLO et al., 2018)

A propriedade de riscos proporcionais é utilizada para averiguar os coeficientes estimados. Observe que a taxa de falha para dois indivíduos (i e l) que possuem os mesmos valores e suas covariáveis, exceto para a p -ésima delas, é obtida por: (GIOLO et al., 2018)

$$\frac{\alpha(t|x_i)}{\alpha(t|x_l)} = \frac{\exp\{\beta_p x_{ip}\}}{\exp\{\beta_p x_{lp}\}} = \exp\{\beta_p (x_{ip} - x_{lp})\} \quad (40)$$

Esta razão é, na realidade, uma razão de riscos e, assim, se por exemplo, x_p é a covariável dicotômica sexo em que $x_{ip} = 1$ (masculino) e $x_{lp} = 0$ (feminino), tem-se que o risco de falha dos elementos presentes no sexo masculino é $\exp(\beta_p)$ vezes o risco de falha dos elementos presentes no sexo feminino, mantendo-se constante as demais covariáveis. (GIOLO et al., 2018)

3.6.3 Suposição de riscos proporcionais no modelo de Cox

O modelo de Cox é utilizado em situações que a suposição de riscos proporcionais é legítima, isto é, para situações em que as linhas das funções de riscos não se cruzam.

Para verificar tal suposição pode-se utilizar, dentre outras maneiras, de técnicas gráficas que consistem basicamente em:

1. Realizar a divisão dos dados em j estratos distintos segundo as j categorias de alguma covariável em interesse.
2. Estimar $\hat{H}_{0j}(t_i)$ para cada estrato j adquirindo as curvas de $\log \hat{H}_{0j}(t)$ contra t , ou $\log(t)$.

Se as hipóteses forem válidas, as curvas de $\log \hat{H}_{0j}(t)$ contra t , ou $\log(t)$, devem possuir diferenças constantes no tempo, ou seja, devem ser aproximadamente paralelas. Estas curvas podem ser obtidas para cada covariável em estudo podendo-se, desse modo, localizar indícios de qual covariável encontraria-se causando a possível violação da suposição de riscos proporcionais. Se a covariável for de natureza contínua é indicado agrupá-la em um pequeno número de categorias. (GIOLO et al., 2018)

3.7 ANÁLISE DE ADEQUAÇÃO AO MODELO AJUSTADO

Uma avaliação da qualidade do modelo ajustado é parte essencial na análise dos dados. No modelo de regressão linear usualmente utilizado, uma análise gráfica dos resíduos é fundamental para esta finalidade. Nos modelos de regressão apresentados anteriormente, a definição de resíduos não é tão evidente e, desse modo, diversos resíduos têm sido propostos na literatura para averiguar um ajuste ao modelo. (COLOSIMO et al., 2006)

Técnicas gráficas, que utilizam dos diferentes resíduos propostos são, em particular, bastante utilizadas para examinar os diversos aspectos do modelo. Um desses aspectos é o de validar, por meio dos resíduos, a distribuição dos erros. Estas técnicas, no entanto, como observado por Klein e Moeschberger (1997), devem ser usufruídas como um meio de recusar modelos evidentemente inapropriados e não para "provar" que um particular modelo paramétrico está correto, mesmo porque, em muitas situações, dois ou mais modelos paramétricos podem prover ajustes razoáveis bem como estimativas próximas dos valores de interesse. (COLOSIMO et al., 2006)

3.7.1 Resíduos de Cox-Snell

Em regressão linear é comum a verificação da adequação do modelo ajustado por meio da análise de gráficos de resíduos. Em sobrevivência, devido a presença de censuras bem como os resíduos não seguirem uma distribuição Normal e poderem ser consideravelmente assimétricos,

desenvolveu-se o Resíduo de Cox-Snell para testar a qualidade do ajuste proposto.

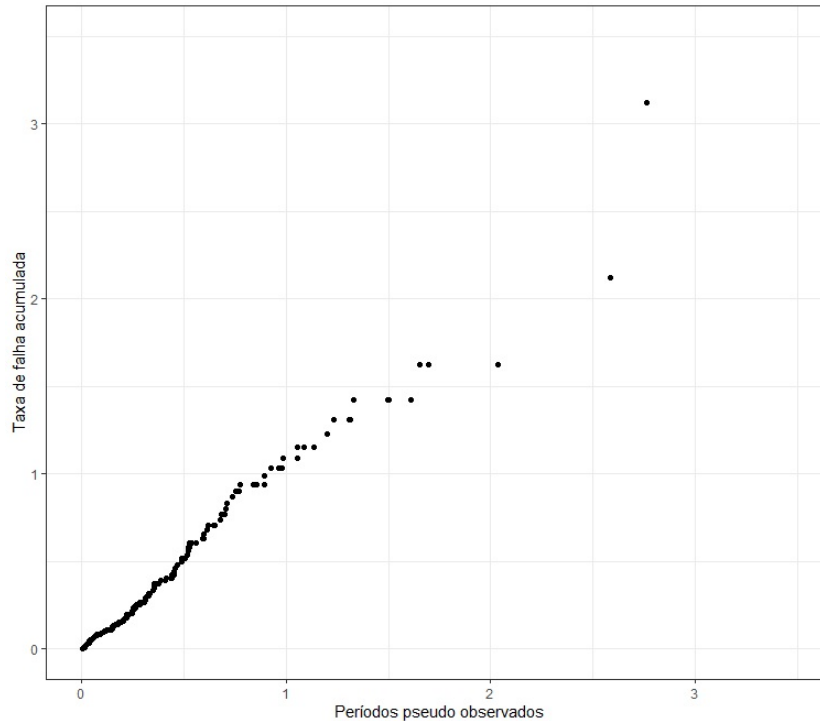
Para todos os n processos em análise ($i = 1, 2, 3, \dots, n$), os Resíduos de Cox-Snell (1968), para o modelo de Cox, são caracterizados por:

$$e_i = \hat{H}_0(t_i) \exp\left(\sum_{k=1}^p x_{ip} \hat{\beta}_k\right) \quad (41)$$

e podem ser utilizados para analisar a qualidade da modelagem no ajuste de Cox. Se o modelo estiver bem ajustado, os e_i 's podem ser visualizados como uma amostra censurada de uma distribuição exponencial padrão e, então, o gráfico de, por exemplo, $\hat{H}(e_i)$ contra e_i necessitaria ser algo próximo de uma reta. (GIOLO et al., 2018)

Os resíduos de Cox-Snell são eficientes para analisar a qualidade do ajuste global no modelo de Cox. Uma desvantagem de seu uso, contudo, é que eles não evidenciam o momento em que o gráfico de $\hat{H}(e_i)$ contra e_i apresenta não linearidade, o tipo de falha detectado pelo modelo. (GIOLO et al., 2018)

Figura 2 – Resíduos de Cox-Snell



3.7.2 Resíduos de Schoenfeld

O resíduo de Schoenfeld é a diferença entre os valores observados de covariáveis em um elemento contido no tempo (t_i) e os valores esperados desde elemento também no período de

tempo (t_i) dado o grupo de risco $R_{(t_i)}$. Existirão tantos vetores de resíduos quanto covariáveis ajustadas no modelo, sendo estes definidos unicamente nos períodos de ocorrência do evento. Considerando-se o plot de resíduos padronizados de Schoenfeld contra o tempo é verificável a ocorrência ou não de proporcionalidade, ou seja, se as suposições de riscos proporcionais forem satisfeitas não deverá existir nenhuma propensão sequencial no gráfico ($H_0 : p = 0$).

Levando em consideração os $k(k \leq n)$ períodos distintos de falha t_1, \dots, t_k , um vetor de resíduos de Schoenfeld é adquirido em cada tempo observado de falha. Assim, se o elemento i é verificado falhar, o correspondente resíduo é obtido por meio de: (GIOLO et al., 2018)

$$r_i = x_i - \frac{\sum_{j \in R_{(t_i)}} x_j e^{\hat{\beta}_{xj}}}{\sum_{j \in R_{(t_i)}} e^{\hat{\beta}_{xj}}} \quad (42)$$

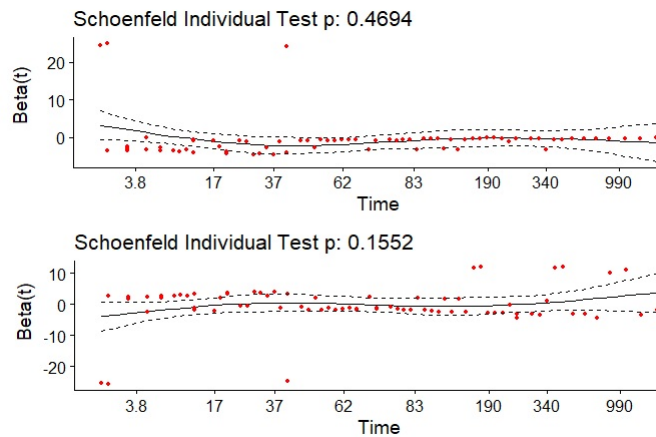
Note que este é um vetor com um componente para cada covariável. Condicional a uma única falha no conjunto de risco $R_{(t_i)}$, o valor esperado da covariável, desta falha, é obtido pelo termo $\frac{\sum_{j \in R_{(t_i)}} x_j e^{\hat{\beta}_{xj}}}{\sum_{j \in R_{(t_i)}} e^{\hat{\beta}_{xj}}}$ evidenciado em (42) e, dessa maneira, a análise de r_i como um resíduo é coerente. Também, $\sum_i r_i = 0$, como usual para resíduos e, se o modelo é apropriado, não necessitaria existência de tendências no tempo. Os resíduos de Schoenfeld são, em geral, usados para analisar as suposições de riscos proporcionais. (GIOLO et al., 2018)

Uma forma graduada dos resíduos de Schoenfeld é representada por:

$$r_i^* = d \times i(\beta)^{-1} \times r_i \quad (43)$$

Em que $i(\beta)$ é a informação e d é o número de falhas observadas.

Figura 3 – Resíduos de Schoenfeld

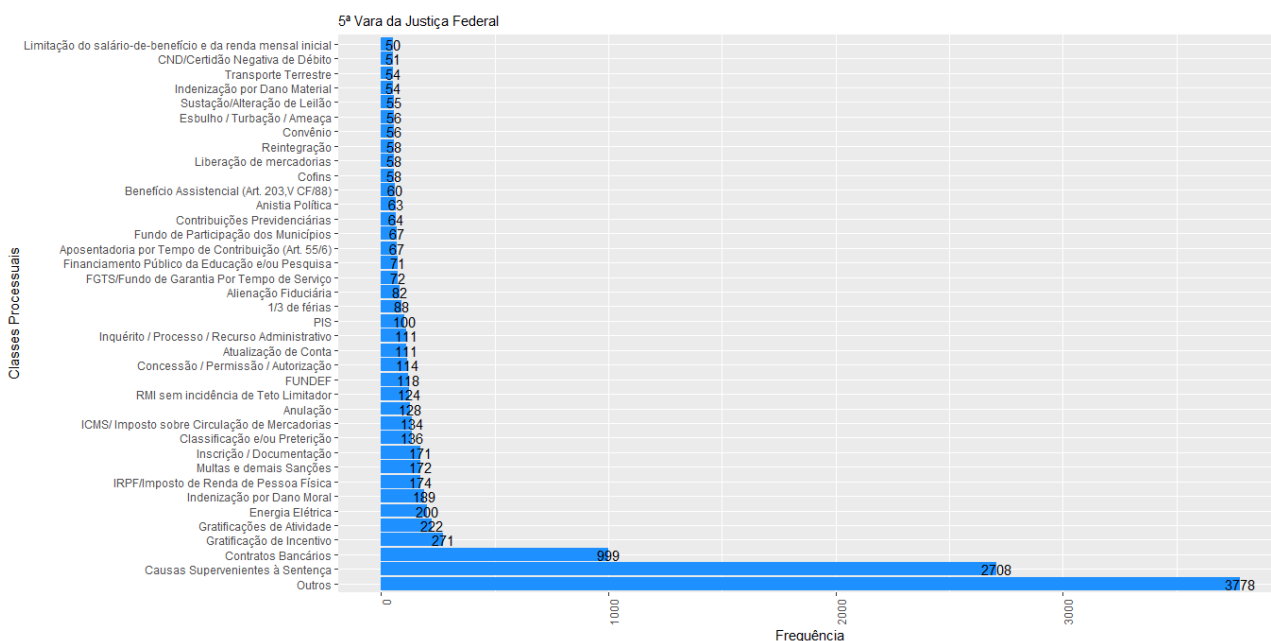


4 ANÁLISE EXPLORATÓRIA

A análise exploratória a seguir é referente ao banco de dados obtido em parceria com a 5ª vara da justiça federal no dia 02/12/2019 e consta inicialmente com 220.363 (duzentos e vinte mil trezentos e sessenta e três) observações separadas em duas classificações distintas, sendo estas Processual e Pje. Dessa maneira, foram selecionadas apenas as classificações Pje que não incluíssem no mesmo dia duplicadas em suas etapas processuais, contabilizando-se 65.539 (sessenta e cinco mil quinhentos e trinta e nove) termos plausíveis de uso para averiguação da qualidade das medidas estatísticas efetuadas e dos ajustes aos modelos propostos sequentemente.

Inicialmente na análise dos dados, foram observadas as frequências das classes processuais para saber como funcionaram as demandas da instituição entre os anos de 2014 a 2019. Observações com contagens inferiores a 50 unidades sofreram ajustes de nomenclatura para "Outros" afim de reduzir a poluição gráfica e otimizar a compreensão dos resultados propostos.

Figura 4 – Frequência das classes processuais

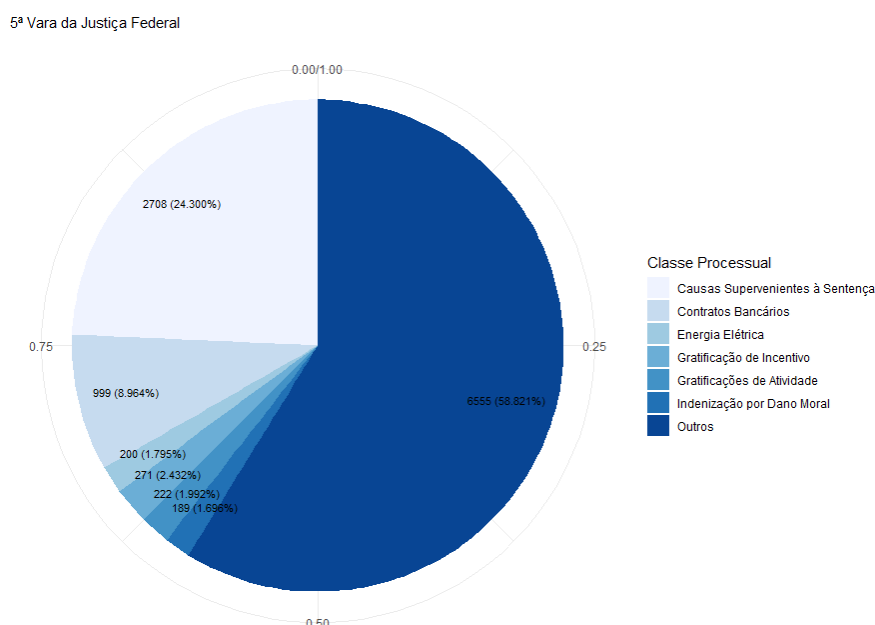


Na **Figura 4** obtém-se um grande destaque nas classes processuais compostas por Causas Supervenientes à Sentença e Contratos Bancários, sendo estes dois equivalentes a 3.707 (três mil setecentos e sete) casos distintos, ou seja, aproximadamente 33% entre todos os processos analisados pela instituição. No entanto, mesmo realizando a soma do segundo e terceiro maior número contabilizado de casos ainda não se supera os 34% referentes a nomenclatura "outros" presentes no gráfico, informando assim que a maior participação da 5ª vara da justiça federal é referente a pequenas classes de processos que necessitaram de julgamentos da instituição ao longo dos

últimos 5 anos, mas que não obtiveram retornos frequentes.

Para uma visualização percentual mais limpa e composta por um menor número de variáveis, maior alocação de classes processuais na nomenclatura "Outros", gerou-se o gráfico presente na **Figura 5**. Sendo este composto pelas variáveis Causas supervenientes à sentença (24.3%), Contratos bancários (8.96%), Energia elétrica (1.79%), Gratificação de incentivo (2.43%), Gratificação de Atividades (2%), Indenização por dano moral (1.7%) e Outros (58.82%) respectivamente.

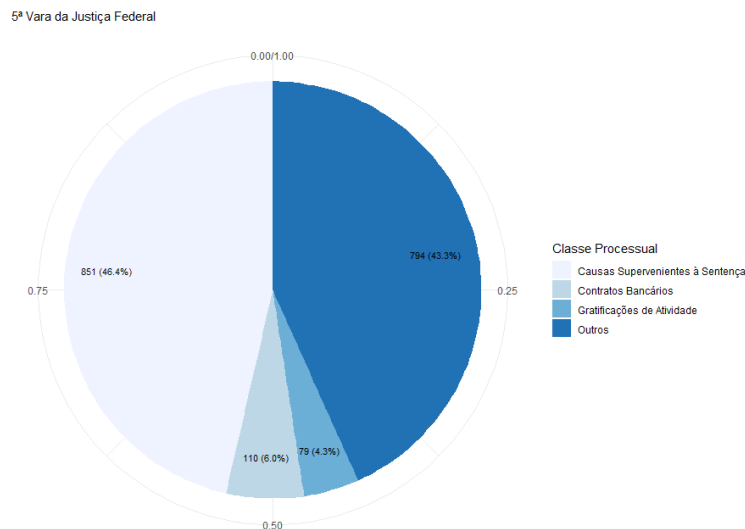
Figura 5 – Frequência das classes processuais



Após as informações obtidas pelos gráficos nas **Figuras 4 e 5**, gerou-se o interesse em desmembrar ainda mais a análise realizada para poder usufruir de uma visualização detalhada em relação aos três anos mais influentes (2017, 2018 e 2019) na instituição em relação à classificação Pje, ou seja, nas **Figuras 6, 7 e 8** foram gerados gráficos de setores possuindo as separações das frequências observadas pelas classes processuais mais influentes em cada um dos três anos acima referenciados.

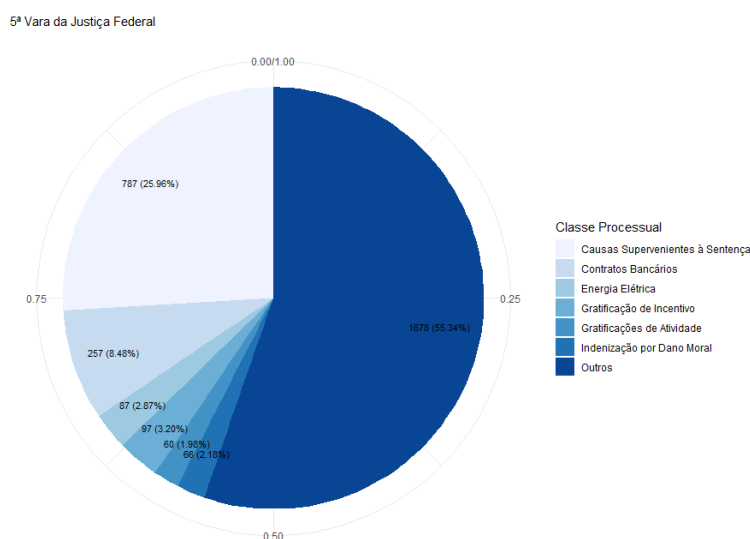
Dessa maneira, consta-se que no ano de 2017 assim como observado na **Figura 4**, a prevalência de frequência das observações são obtidas nas classes processuais "Causas Supervenientes à Sentença" (46.4%) e "Contratos Bancários" (6%), no entanto, ainda considerando o ano de 2017 segundo a **Figura 6** observa-se que o percentual de frequência de ambas as classes são consideravelmente maiores que as presentes na **Figura 4**, contabilizando cerca de 52.4% do valor total presente.

Figura 6 – Percentual de frequência das classes processuais no ano de 2017



No ano subsequente de 2018, já se considera uma maior abrangência na variedade do número de frequências existentes para as diversas classes processuais, embora se tenha permanecido a prevalência dos mesmos grupos de categorias descritas para o ano anterior, contudo ocorre uma redução significativa do percentual existente na soma de ambas para 34.44%.

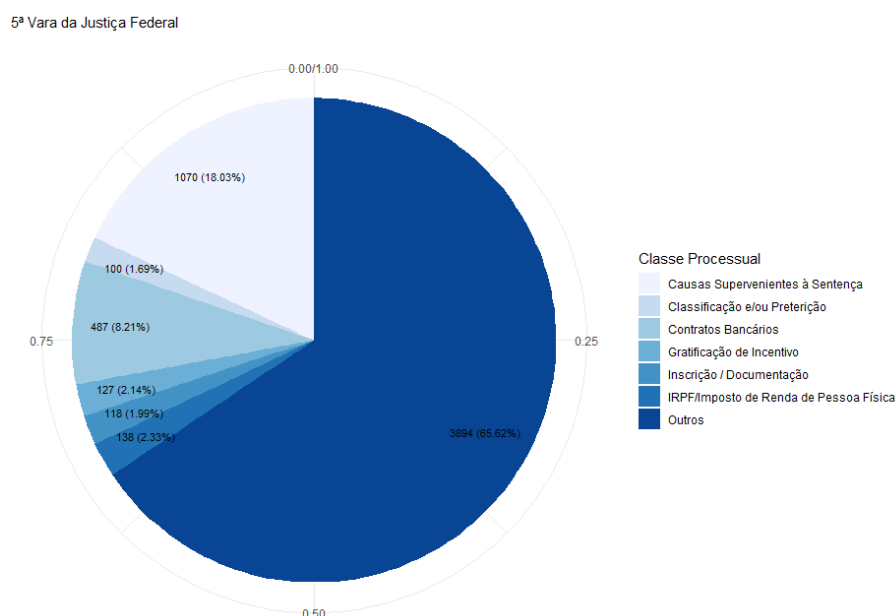
Figura 7 – Percentual de frequência das classes processuais no ano de 2018



Para finalizar com o ano de 2019, a **Figura 8** apresenta uma diversidade ainda maior entre as participações das diversas classes, mas ainda considerando a prevalência dos mesmos destaques dos anos anteriores, reduzindo-se apenas seu percentual para 26.24% do valor total

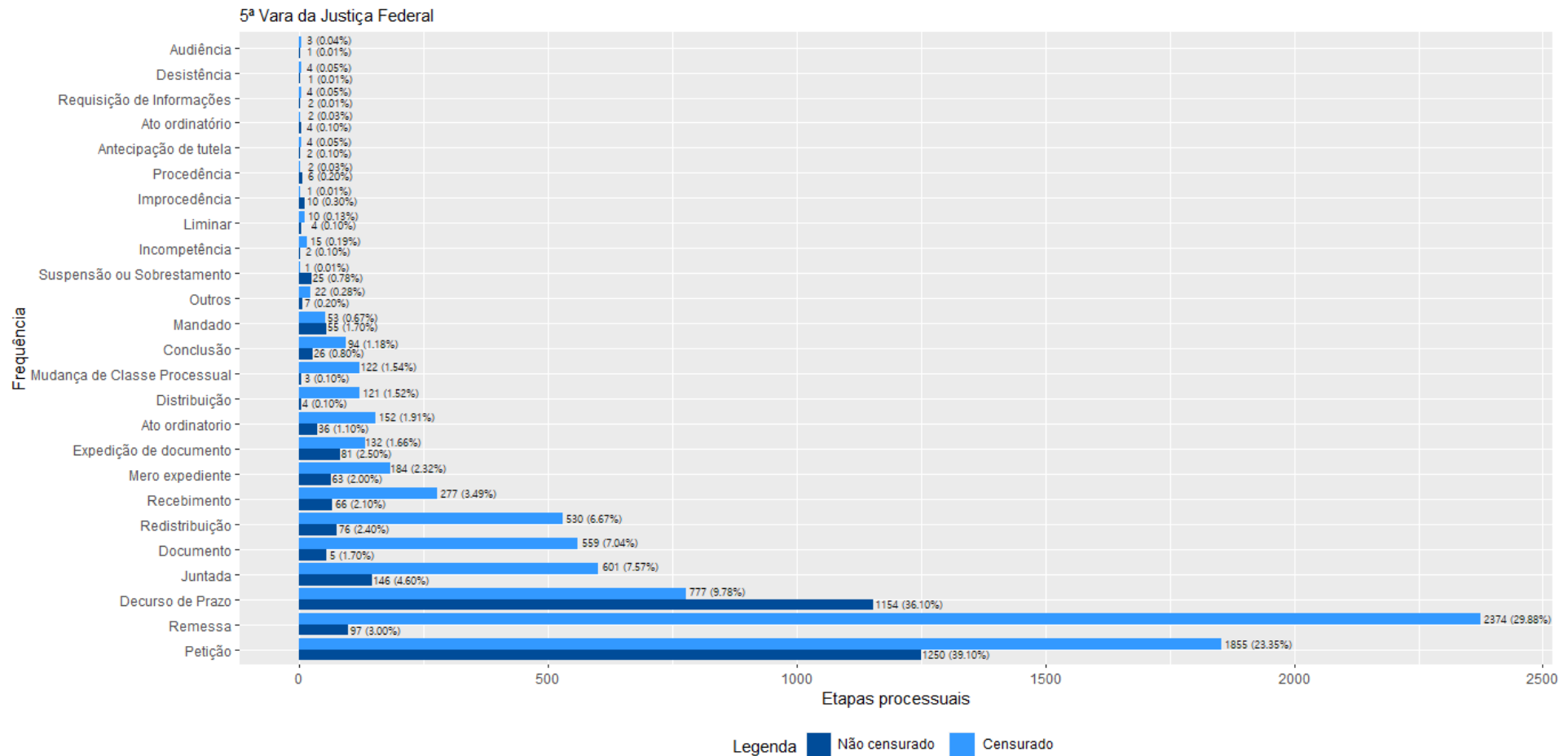
apresentado para o ano de 2019.

Figura 8 – Percentual de frequência das classes processuais no ano de 2019



Voltando-se para a análise do funcionamento das etapas realizadas pela instituição em cada um dos processos efetuados, considerou-se duas maneiras distintas de avaliação dos resultados. Primeiramente, evidenciou-se a frequência de dados não censurados existentes para cada uma das etapas, com o objetivo de mostrar o número de vezes que estes estágios foram concluídos dentre as diversas classes processuais e posteriormente o número de frequências censuradas no banco, ou seja, o número de vezes que as diversas fases ainda se encontram com status aberto até o dia 02/12/2019 para as distintas classes de pleitos, **Figura 9**.

Figura 9 – Frequência de observações censuradas e não censuradas nas etapas processuais



Dessa maneira pode-se dizer que em relação as etapas não censuradas, "Decurso de prazo" (1154 observações - 36.10%) e "Petição" (1250 termos - 39.10%) sozinhas equivalem a 2.404 (duas mil quatrocentas e quatro) frequências de estágios com status ainda em aberto até o dia de coleta das informações, ou seja, são essas às duas principais etapas onde a maior parte dos processos da instituição se encontram em andamento.

Já em relação à frequências de etapas censuradas, destaca-se uma maior ocorrência em ordem crescente das etapas em "Petição" e "Remessa" com respectivamente 1.855 e 2.374 casos distintos. Dessa maneira, vale ressaltar que sozinhas essas duas atividades foram responsáveis por 4.229 (quatro mil duzentas e vinte e nove) observações dentre os 7.890 (sete mil oitocentos e noventa) casos, representando cerca de 53.59% do valor total referenciado pelo sistema.

Com o intuito de compreender a evolução das classes processuais em relação ao tempo, foi efetuado o gráfico de Box-plot, **Figura 10**, que evidencia os valores outliers da distribuição. Além disso, salientam-se as posições aproximadas do primeiro quartil, média, mediana e terceiro quartil de cada uma das variáveis presentes nas classes processuais.

Assim, de maneira meramente visual ao se considerar apenas a **Figura 10**, é possível acatar que os valores mais elevados das médias em relação ao tempo em dias se encontram nas variáveis "FUNDEF", "Contratos Bancários", "Anistia Política" e "Indenização por Dano Material". Já se considerando a amplitude interquartílica dos dados com o intuito de descrever uma possível desaglomeração de valores na duração das classes processuais analisadas, destacam-se as variáveis "Reintegração", "FUNDEF", "ICMS", "Convênio" e "Outros". Enquanto em contrapartida, os valores mais aglomerados estão em "Benefício Assistencial (Art. 203, V CF/88)" e "Financiamento Público da Educação e/ou Pesquisa".

Na **Tabela 1** abaixo encontram-se descritas todas as medidas descritivas utilizadas para elaboração do gráfico de Boxplot referente as classes processuais, **Figura 10**.

Figura 10 – Boxplot das classes processuais versus tempo de duração do pleito

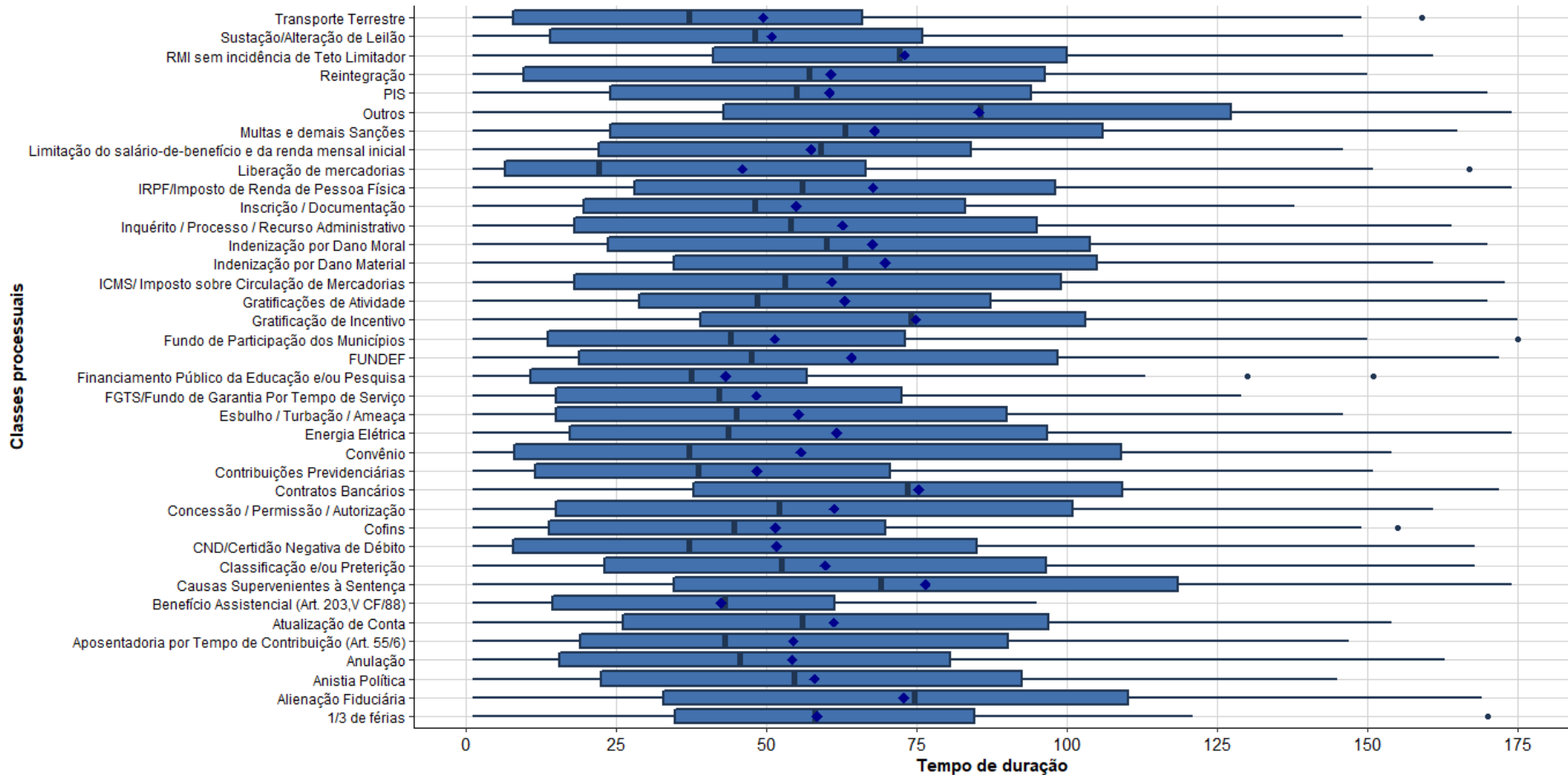


Tabela 1 – Medidas descritivas das variáveis presentes no Boxplot

	Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
1/3 de férias	1.00	10.25	40.50	53.03	74.50	190.00
Alienação Fiduciária	1.00	3.25	42.00	53.84	100.00	189.00
Anistia Política	1.00	6.50	50.00	77.08	100.50	651.00
Anulação	1.0	2.0	14.0	41.7	65.0	274.0
Aposentadoria por Tempo de Contribuição (Art. 55/6)	1.0	2.5	16.0	42.4	48.0	218.0
Atualização de Conta	1.00	22.00	50.00	61.41	98.00	269.00
Benefício Assistencial (Art. 203,V CF/88)	1.0	2.0	18.0	31.0	47.5	183.0
Causas Supervenientes à Sentença	1.00	3.00	34.00	49.76	70.00	443.00
Classificação e/ou Preterição	1.00	2.75	31.50	50.74	74.00	651.00
CND/Certidão Negativa de Débito	1.00	4.00	21.00	56.49	86.00	407.00
Cofins	1.00	4.25	42.00	43.91	67.00	155.00
Concessão / Permissão / Autorização	1.00	2.00	8.00	42.25	93.00	294.00
Contratos Bancários	1.00	33.00	65.00	78.73	101.00	651.00
Contribuições Previdenciárias	1.00	6.75	38.50	73.19	122.00	412.00
Convênio	1.00	2.00	7.50	47.29	58.25	379.00
Energia Elétrica	1.0	3.0	8.5	42.9	59.0	288.0
Esubulho / Turbação / Ameaça	1.00	3.00	13.50	42.88	66.25	275.00
FGTS/Fundo de Garantia Por Tempo de Serviço	1.00	4.75	27.00	47.78	62.50	651.00
Financiamento Público da Educação e/ou Pesquisa	1.00	2.00	7.00	40.69	42.00	651.00
FUNDEF	1.00	14.25	69.00	103.30	156.00	635.00
Fundo de Participação dos Municípios	1.00	2.50	19.00	45.97	73.00	178.00
Gratificação de Incentivo	1.00	14.50	49.00	66.81	100.00	269.00
Gratificações de Atividade	1.00	3.00	34.00	55.64	86.25	304.00
ICMS/ Imposto sobre Circulação de Mercadorias	1.0	4.0	27.5	47.9	67.0	375.0
Indenização por Dano Material	1.00	34.25	65.50	76.52	102.25	249.00
Indenização por Dano Moral	1.00	9.00	45.00	60.63	86.00	598.00
Inquérito / Processo / Recurso Administrativo	1.00	4.00	26.00	58.18	81.50	651.00
Inscrição / Documentação	1.00	2.00	18.00	54.69	65.50	651.00
IRPF/Imposto de Renda de Pessoa Física	1.00	7.00	33.00	51.24	90.00	220.00

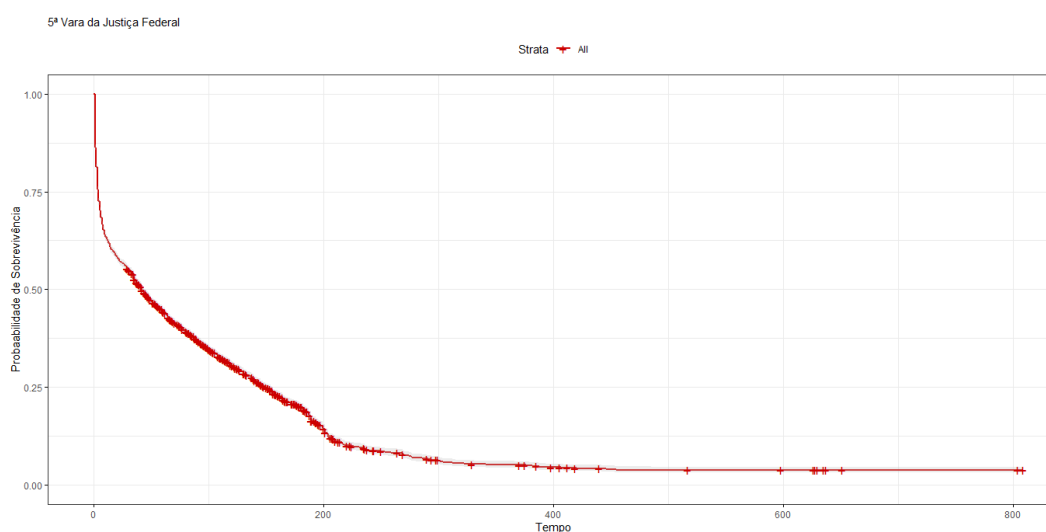
	Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo
Liberação de mercadorias	1.00	1.00	5.50	50.81	49.25	433.00
Limitação do salário-de-benefício e da renda mensal inicial	1.00	19.50	40.50	50.06	79.25	146.00
Multas e demais Sanções	1.00	3.00	31.00	63.07	105.00	651.00
Outros	1.00	4.00	35.00	61.42	88.75	809.00
PIS	1.00	4.00	35.00	55.87	85.00	651.00
Reintegração	1.00	4.00	20.50	53.29	93.75	270.00
RMI sem incidência de Teto Limitador	1.00	31.75	59.00	65.60	93.00	274.00
Sustação/Alteração de Leilão	1.00	2.00	10.00	33.62	57.50	146.00
Transporte Terrestre	1.00	1.00	5.50	30.91	40.25	224.00

A função de sobrevivência representada pela **Figura 11** serve para comparar os tempos de falha segundo variáveis qualitativas. No exemplo abaixo, pode-se comparar o comportamento do tempo até o arquivamento do processo de acordo com outras variáveis de interesse, como classe processual e etapa processual.

Observa-se na **Figura 11** que a função se inicia num determinado momento no tempo, com 100% dos processos ainda com status em aberto, nos permitindo calcular qual a percentagem desses processos permanecem em aberto em relação a outros momentos ao longo do tempo, ou seja, serve para evidenciar o percentual de chance dos pleitos serem concluídos antes de seu arquivamento.

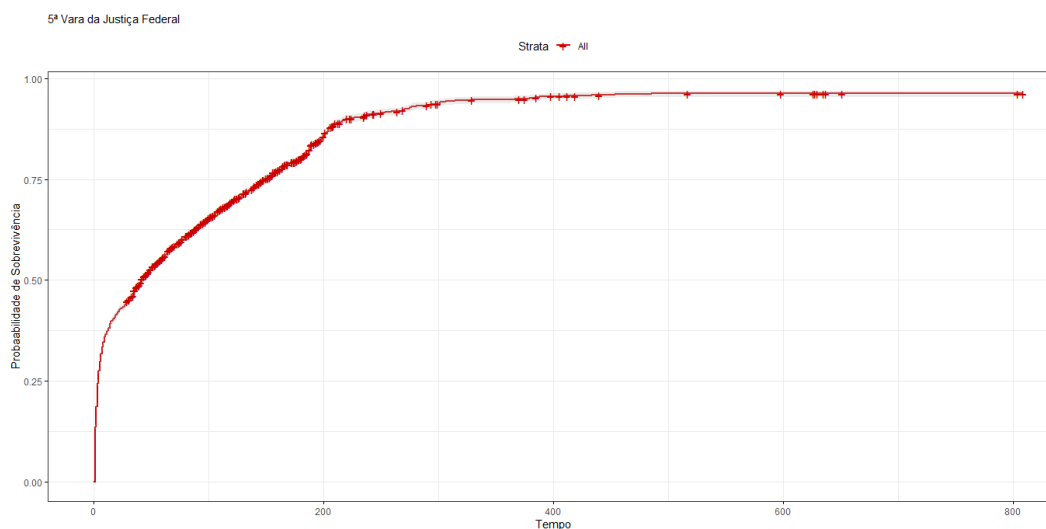
Já a função da taxa de risco acumulada presente na **Figura 12** mostra a proporção de processos no todo que inicialmente encontram-se com status em aberto ou que não sofreram ainda o evento de interesse. Dessa maneira, este vem a sofrer o evento de interesse, arquivamento, em algum momento durante um período de tempo definido.

Figura 11 – Função de sobrevivência



Com base nos valores obtidos no gráfico de sobrevivência, observa-se um rápido decaimento das probabilidades dos pleitos permanecerem em aberto ao longo do tempo. Dessa maneira, com cerca de 50% do período de duração médio dos processos em dias, os pleitos já apresentam aproximadamente 6% de probabilidade. Já com 81% do tempo, essa probabilidade cai para algo próximo de 5%.

Figura 12 – Função taxa de risco acumulada



Já em relação ao gráfico da Função taxa de risco acumulada, observa-se que com 50% do período de duração médio dos processos em dias, existe uma probabilidade de falha na conclusão do pleito antes da ocorrência de seu arquivamento de aproximadamente 92%. Já com 81% do tempo, esse percentual sobe para quase 94%.

Abaixo nas **Figuras 13 e 14** foram efetuados os cálculos dos modelos paramétricos de falhas aceleradas (**AFT**), este que leva em consideração a existência de tempos de sobrevivência desacelerados ou acelerados por um fator que é constante quando comparado a diferentes níveis de covariáveis.

Utilizando-se do ajuste **AFT** por meio do comando **Survreg**, realizou-se a adequação do modelo de regressão a 4 (quatro) distribuições distintas de interesse (Weibull, Exponencial, Log-logística e Log-normal) representados no gráfico presente na **Figura 13**. Dessa maneira, considerou-se a abrangência da aplicabilidade dos modelos acima citados para elaboração dos 4 (quatro) modelos paramétricos descritos abaixo:

- **Modelo 1** apresenta em seus interceptos apenas as relações entre as etapas processuais;
- **Modelo 2** possui em seus interceptos apenas as relações entre as classes processuais;
- **Modelo 3** dispõe em seus interceptos uma junção entre as categorias de classes processuais e etapas processuais, considerando-se os valores duplicados removidos;
- **Modelo 4** retém em seus interceptos uma relação de *cross over* entre as categorias de classes processuais e etapas processuais.

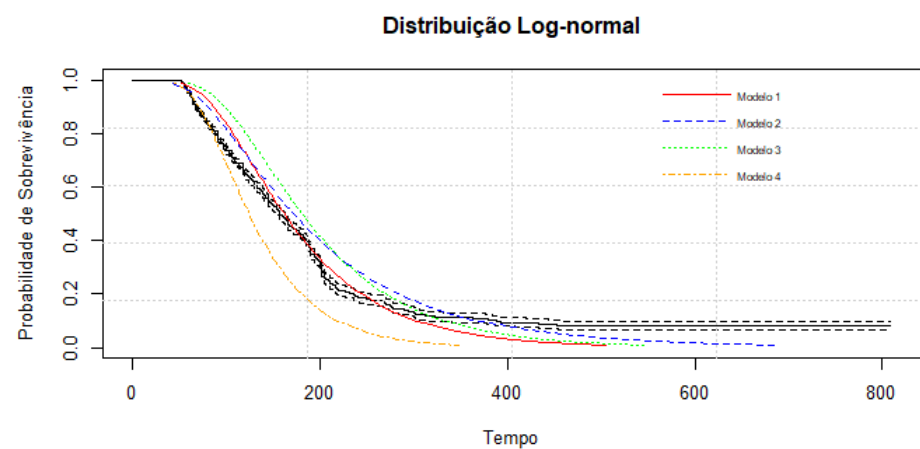
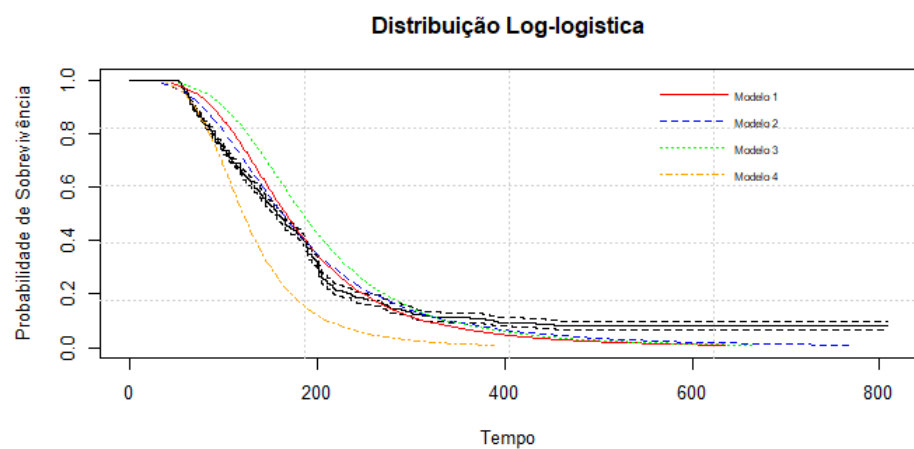
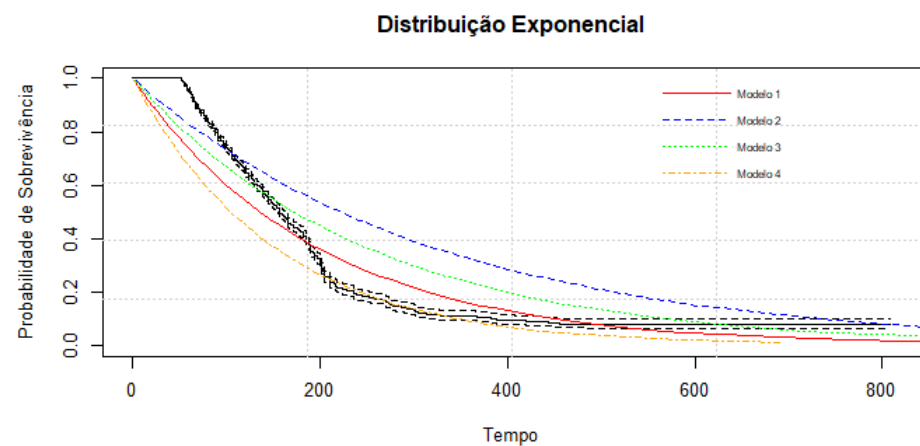
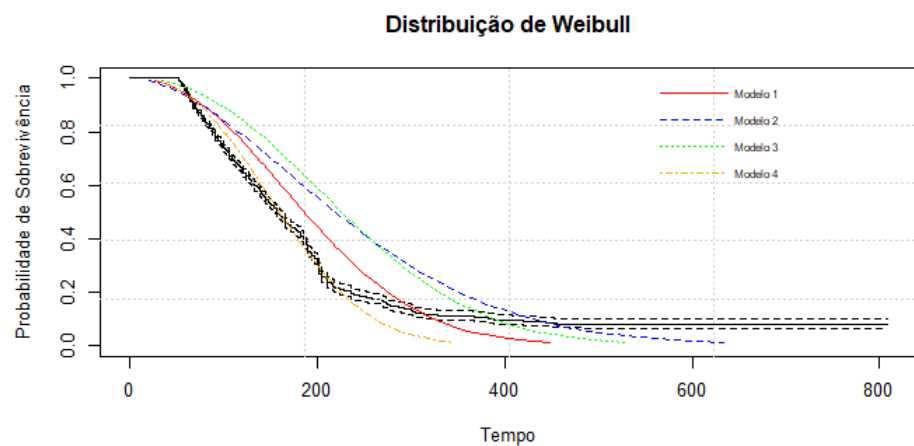


Figura 13 – Comparação entre a função de sobrevivência e alguns ajustes do modelo

Observando os resultados obtidos pela **Figura 13**, selecionou-se o modelo 1 como melhor opção de ajuste aos dados apresentados, já que este possui uma maior aproximação de enquadramento em relação aos valores evidenciados pela função de sobrevivência em todas as distribuições.

Posteriormente utilizando-se do intuito de gerar uma visualização mais robusta para selecionar quais distribuições dentre o modelo selecionado utilizam de um melhor ajuste, plotou-se o gráfico presente na **Figura 14**. Este que em conjunto com as **Tabelas 2, 3, 4, 5, 6 e 7** evidenciam a distribuição Exponencial como aquela que se destaca num melhor ajuste de resultados obtidos para esse conjunto de dados em estudo.

Figura 14 – Comparação entre a função de sobrevivência e os melhores ajustes do modelo

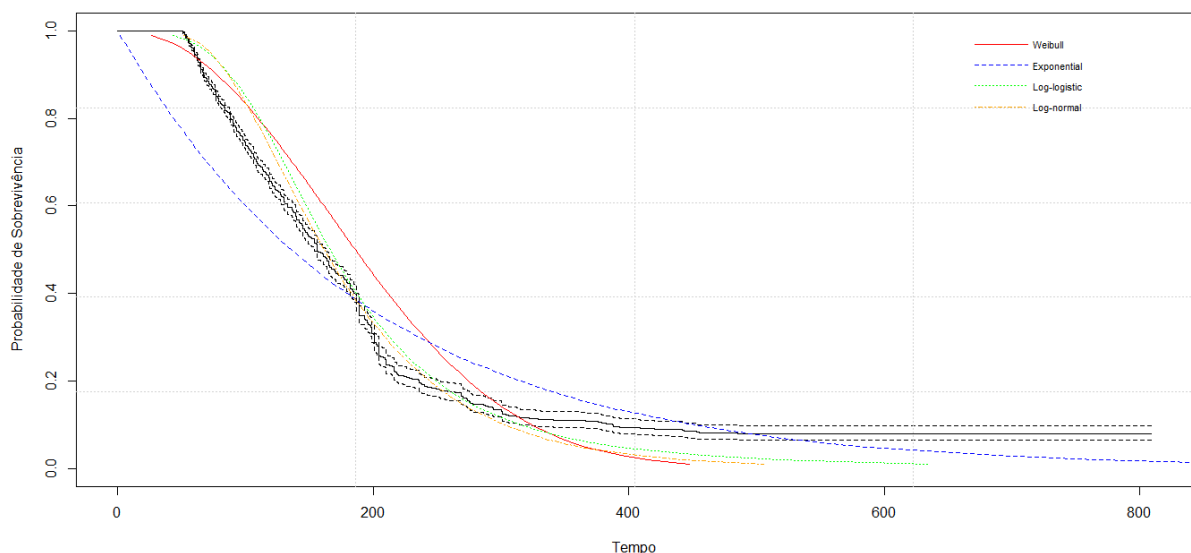


Tabela 2 – Qualidade de ajuste dos modelos em relação ao Deviance

	Weibull	Exponential	Log-logistic	Log-normal
Deviance	4777.141	3409.105	4611.491	4141.815

Tabela 3 – Resultados gerais apresentados pelos modelos de regressão

	Loglik (modelo)	Loglik (intercepto)	Qui-quadrado	P-value	Interações de Newton-Raphson
Weibul	-11833	-13422.7	3179.43 (44 gl)	0	22
Exponencia	-12593.9	-13877.8	2567.87 (44 gl)	0	21
Log-logística	-11721.7	-12997.2	2551.07 (44 gl)	0	19
Log-normal	-11661.6	-12977.9	2632.54 (44 gl)	0	17

O motivo pelo qual se selecionou a distribuição de probabilidade exponencial visualizada na **Figuras 14** foi respectiva a análise visual e devido o resultado obtido pelo deviance e pela razão de verossimilhança, esta que consiste na razão entre o logaritmo da verossimilhança de um modelo mais simples (que se deseja testar) e outro mais generalizado. Uma vez que esta razão segue uma distribuição qui-quadrada, pode-se ter seu p-valor calculado pela distribuição considerando o número de graus de liberdade como a diferença do número de parâmetros.

$$TRV = 2 \cdot \log \left(\frac{L(\theta)_{\text{simples}}}{L(\theta)_{\text{generalizado}}} \right) \quad (44)$$

O modelo de verossimilhança acima proposto serve para testar a hipótese de que a variável independente é relevante na predição da variável resposta, ou seja,

$$H_0 : \beta_1 = \beta_2 = \dots = 0 \quad \text{vs} \quad H_a : \beta_1 \neq \beta_2 \neq \dots \neq 0$$

Sob hipótese nula, a estatística do teste segue uma distribuição qui-quadrado com o número de graus de liberdade igual ao número de restrições impostas pela hipótese nula.

Com base nas informações presentes inicialmente na **Tabelas 2**, considera-se que a variável independente (Classe Processual) utilizada na estruturação do modelo 2 foi consideravelmente relevante na predição da variável resposta assumindo-se um alfa de 95%.

Dessa maneira, também considerando-se os valores presentes nas **Tabelas 3, 4, 5 e 6** observa-se a possibilidade de quantificar a influencia de cada classe processual em relação a variável resposta. Além disso, também pode-se considerar o resultado obtido no Log(Scale) para averiguar que as distribuições em Weibull, Log-Logística e Log-Normal realmente não teriam resultados significativos de P-value que comprovassem um ajuste adequado aos dados, ou seja, assim como mencionado anteriormente o ajuste realizado com a distribuição exponencial é aquele que melhor se adequa ao conjunto de dados existente.

Tabela 4 – Medidas do modelo de regressão na Distribuição de Weibull

	Valor	Std. Error	Z	P-value
(Intercept)	15.2904	556.5059	0.0275	0.9781
Ausência de pressupostos processuais	-0.9102	0	-Inf	<2e-16
Baixa Definitiva	-0.6112	0	-Inf	<2e-16
Cancelamento de Distribuição	-0.9102	0	-Inf	<2e-16
Antecipação de tutela	-9.9334	556.5061	-0.0178	0.9858
Ato ordinatório	-10.2425	556.5059	-0.0184	0.9853
Audiência	-10.363	556.5059	-0.0186	0.9851
Decurso de Prazo	-10.2927	556.5059	-0.0185	0.9852
Distribuição	-9.8703	556.5059	-0.0177	0.9858
Documento	-9.8954	556.5059	-0.0178	0.9858
Expedição de documento	-10.1243	556.5059	-0.0182	0.9855
Incompetência	-10.3556	556.506	-0.0186	0.9852
Juntada	-10.4857	556.5059	-0.0188	0.985
Mandado	-10.2568	556.5059	-0.0184	0.9853
Mero expediente	-10.2031	556.5059	-0.0183	0.9854
Mudança de Classe Processual	-10.5661	556.5059	-0.019	0.9849
Não-Acolhimento de Embargos de Declaração	-10.9019	556.506	-0.0196	0.9844
Outros	-11.0562	556.5061	-0.0199	0.9841
Perempção, litispendência ou coisa julgada	-10.9209	556.5061	-0.0196	0.9843
Petição	-10.2033	556.5059	-0.0183	0.9854
Recebimento	-10.4048	556.5059	-0.0187	0.9851
Redistribuição	-10.292	556.5059	-0.0185	0.9852
Remessa	-10.39	556.5059	-0.0187	0.9851
Requisição de Informações	-10.4087	556.5061	-0.0187	0.9851
Desistência	-0.9102	0	-Inf	<2e-16
Improcedência	-0.9102	0	-Inf	<2e-16
Julgamento em Diligência	-0.839	0	-Inf	<2e-16
Suspensão ou Sobrestamento	1.0095	0	Inf	<2e-16
Log(scale)	-0.7679	0.016	-47.8724	<2e-16

Tabela 5 – Medidas do modelo de regressão na Distribuição Exponencial

	Valor	Std. Error	Z	P-value
(Intercept)	27.0998	1309.9111	0.0207	0.9835
Ausência de pressupostos processuais	-0.9102	0	-Inf	<2e-16
Baixa Definitiva	-0.6113	0	-Inf	<2e-16
Cancelamento de Distribuição	-0.9102	0	-Inf	<2e-16
Antecipação de tutela	-21.4162	1309.9115	-0.0163	0.987
Ato ordinatório	-21.9523	1309.9115	-0.0168	0.9866
Audiência	-22.1139	1309.9112	-0.0169	0.9865
Decurso de Prazo	-22.1843	1309.9111	-0.0169	0.9865
Distribuição	-21.8589	1309.9111	-0.0167	0.9867
Documento	-21.8216	1309.9111	-0.0167	0.9867
Expedição de documento	-22.0145	1309.9111	-0.0168	0.9866
Incompetência	-21.9494	1309.9113	-0.0168	0.9866
Juntada	-22.3761	1309.9111	-0.0171	0.9864
Mandado	-22.0844	1309.9111	-0.0169	0.9865
Mero expediente	-22.1457	1309.9111	-0.0169	0.9865
Mudança de Classe Processual	-22.4427	1309.9111	-0.0171	0.9863
Não Acolhimento de Embargos de Declaração	-22.7116	1309.9113	-0.0173	0.9862
Outros	-22.8657	1309.9115	-0.0175	0.9861
Perempção, litispendência ou coisa julgada	-22.7304	1309.9115	-0.0174	0.9862
Petição	-22.0765	1309.9111	-0.0169	0.9866
Recebimento	-22.1327	1309.9111	-0.0169	0.9865
Redistribuição	-22.1632	1309.9111	-0.0169	0.9865
Remessa	-22.2168	1309.9111	-0.017	0.9865
Requisição de Informações	-21.6445	1309.9115	-0.0165	0.9868
Desistência	-0.9102	0	-Inf	<2e-16
Improcedência	-0.9102	0	-Inf	<2e-16
Julgamento em Diligência	-0.8833	0	-Inf	<2e-16
Suspensão ou Sobrestamento	0.9009	0	Inf	<2e-16

Tabela 6 – Medidas do modelo de regressão na Distribuição Log-Logística

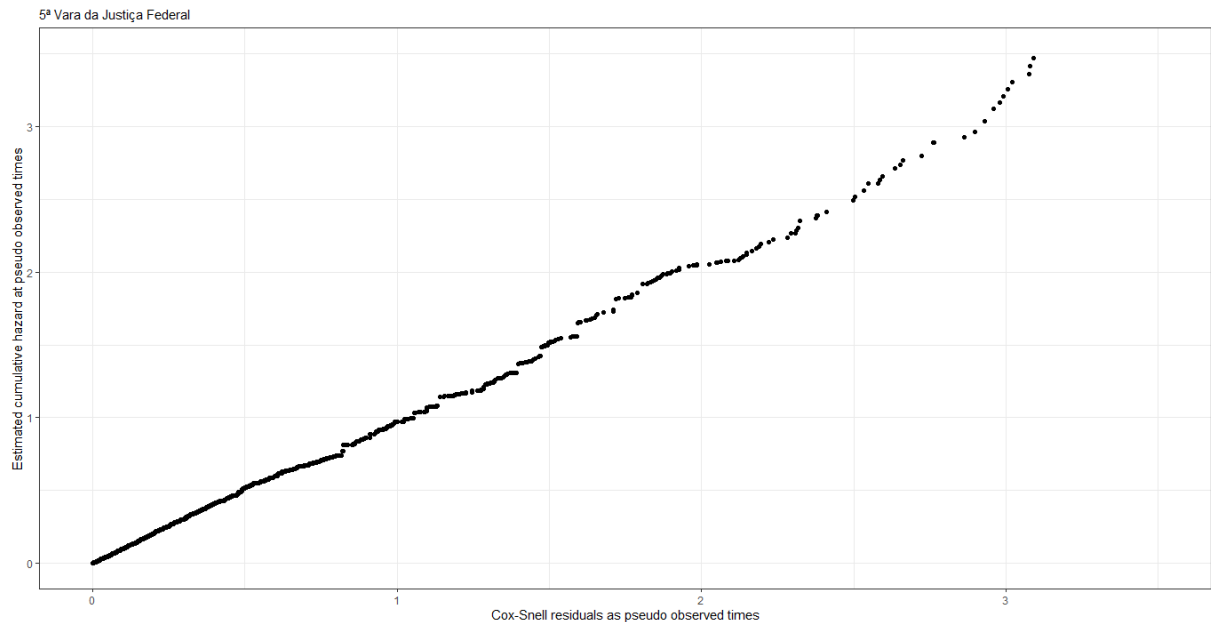
	Valor	Std. Error	Z	P-value
(Intercept)	11.4169	729.3824	0.0157	0.9875
Ausência de pressupostos processuais	-0.9102	0	-Inf	<2e-16
Baixa Definitiva	-0.6111	0	-Inf	<2e-16
Cancelamento de Distribuição	-0.9102	0	-Inf	<2e-16
Antecipação de tutela	-6.3304	729.3825	-0.0087	0.9931
Ato ordinatório	-6.2694	729.3825	-0.0086	0.9931
Audiência	-6.8144	729.3824	-0.0093	0.9925
Decurso de Prazo	-6.7364	729.3824	-0.0092	0.9926
Distribuição	-6.36	729.3824	-0.0087	0.993
Documento	-6.3036	729.3824	-0.0086	0.9931
Expedição de documento	-6.4534	729.3824	-0.0088	0.9929
Incompetência	-6.5559	729.3824	-0.009	0.9928
Juntada	-6.7774	729.3824	-0.0093	0.9926
Mandado	-6.79	729.3824	-0.0093	0.9926
Mero expediente	-6.7007	729.3824	-0.0092	0.9927
Mudança de Classe Processual	-6.9098	729.3824	-0.0095	0.9924
Não Acolhimento de Embargos de Declaração	-7.0288	729.3824	-0.0096	0.9923
Outros	-7.1828	729.3825	-0.0098	0.9921
Perempção, litispendência ou coisa julgada	-7.0474	729.3825	-0.0097	0.9923
Petição	-6.5521	729.3824	-0.009	0.9928
Recebimento	-6.8211	729.3824	-0.0094	0.9925
Redistribuição	-6.7051	729.3824	-0.0092	0.9927
Remessa	-6.7634	729.3824	-0.0093	0.9926
Requisição de Informações	-6.6736	729.3825	-0.0091	0.9927
Decurso de Prazo	-0.4392	859.9968	-5e-04	0.9996
Desistência	-0.9102	0	-Inf	<2e-16
Improcedência	-0.9102	0	-Inf	<2e-16
Julgamento em Diligência	-0.7467	0	-Inf	<2e-16
Suspensão ou Sobrestamento	1.2185	0	Inf	<2e-16
Log(scale)	-1.2319	0.0177	-69.7306	<2e-16

Tabela 7 – Medidas do modelo de regressão na Distribuição Log-Normal

	Valor	Std. Error	Z	P-value
(Intercept)	8.1379	522.548	0.0156	0.9876
Ausência de pressupostos processuais	-0.9102	0	-Inf	<2e-16
Baixa Definitiva	-0.6106	0	-Inf	<2e-16
Cancelamento de Distribuição	-0.9102	0	-Inf	<2e-16
Antecipação de tutela	-3.0147	522.5481	-0.0058	0.9954
Ato ordinatório	-2.9904	522.5482	-0.0057	0.9954
Audiência	-3.4549	522.548	-0.0066	0.9947
Decurso de Prazo	-3.427	522.548	-0.0066	0.9948
Distribuição	-3.0947	522.548	-0.0059	0.9953
Documento	-3.0519	522.548	-0.0058	0.9953
Expedição de documento	-3.1885	522.548	-0.0061	0.9951
Incompetência	-3.2459	522.5481	-0.0062	0.995
Juntada	-3.4941	522.548	-0.0067	0.9947
Mandado	-3.4642	522.548	-0.0066	0.9947
Mero expediente	-3.3929	522.548	-0.0065	0.9948
Mudança de Classe Processual	-3.6001	522.548	-0.0069	0.9945
Não Acolhimento de Embargos de Declaração	-3.7498	522.5481	-0.0072	0.9943
Outros	-3.9038	522.5482	-0.0075	0.994
Perempção, litispendência ou coisa julgada	-3.7685	522.5482	-0.0072	0.9942
Petição	-3.2878	522.548	-0.0063	0.995
Recebimento	-3.491	522.548	-0.0067	0.9947
Redistribuição	-3.4188	522.548	-0.0065	0.9948
Remessa	-3.4561	522.548	-0.0066	0.9947
Requisição de Informações	-3.342	522.5481	-0.0064	0.9949
Decurso de Prazo	-0.0976	542.8299	-2e-04	0.9999
Desistência	-0.9102	0	-Inf	<2e-16
Improcedência	-0.9102	0	-Inf	<2e-16
Julgamento em Diligência	-0.5611	0	-Inf	<2e-16
Suspensão ou Sobrestamento	1.443	0	Inf	<2e-16
Log(scale)	-0.7138	0.0155	-46.0684	<2e-16

Após a seleção do melhor ajuste conjuntamente com sua distribuição, gera-se o interesse de qualificar os resíduos para averiguar a qualidade do modelo proposto. Dessa maneira, abaixo na **Figura 15** utilizou-se o gráfico de resíduos de Cox-Snell para realização dessa atividade.

Figura 15 – Análise de resíduos de Cox-Snell



Considerando que a interpretação de um bom ajuste no gráfico acima se deve quando os (e_i 's) podem ser olhados como uma amostra censurada de uma distribuição exponencial padrão e, então, o gráfico de, $H(e_i)$ versus (e_i) cria um *plot* que deveria ser aproximadamente uma reta, então pode-se considerar com base no *plot* acima que o modelo 1 conjuntamente com a distribuição exponencial não apresenta maiores problemas em relação aos seus resíduos.

4.1 ANÁLISE DO SISTEMA

O sistema foi desenvolvido por meio do software R com auxílio de pacotes, tendo destaque principalmente para o Shiny³ e ShinyDashboard, que possibilitaram a modelagem da página Web em um layout de Bootstrap que fosse simples e agradável ao usuário final do produto.

A página de acesso representada pela **Figura 16** possibilita a descrição da aba de login de usuário (User Id), sendo este existente devido uma necessidade de segurança com os dados públicos presentes no sistema e a possibilidade dos advogados e/ou usuários da instituição realizarem

3

Shiny é um pacote R que facilita a criação de aplicativos interativos via Web. Você pode hospedar aplicativos independentes em uma página online ou incorporá-los aos documentos do R Markdown ou criar painéis interativos via Browser. Também é possível obter uma extensão de seus aplicativos com temas CSS, widgets html e ações JavaScript.

atualizações e adições de novos pleitos ao banco de dados. Também foram adicionados 3 (três) ícones ao topo da tela que servem para encaminhar o usuário para sites externos previamente selecionados e para advertir sobre novas atualizações ou informações a respeito da instituição.

Após a inserção do login e senha para liberação de acesso ao sistema, ocorre a possibilidade do usuário usufruir de 10 diferentes abas alocadas no canto superior esquerdo do Browser. A primeira opção representada pela **Figura 17** disponibiliza uma visualização dos prazos pré definidos pela instituição para elaboração de cada uma das atividades processuais, segundo cada setor interno. Dessa maneira, o usuário pode manter-se atento ao tempo sobressalente ou faltante para conclusão de sua etapa atualmente administrada ao processo. Também é possível ainda na mesma aba por meio de filtros, a seleção dos meses, anos e classes processuais de interesse para análises descritivas dos números de pleitos com classificação aberta ou fechada, número de categorias dentre os status, com alocação de valores em Ok, Alerta, Atrasado, que possuem relação com as classes processuais fixadas anteriormente. Além disso, existem gráficos de bolhas e setores que são referentes ao controle de tempo para determinadas etapas processuais e descrições do percentual de tempo demandado pela instituição para conclusão dos processos elaborados, respectivamente.

Na segunda opção de menu, Manual de Uso, visualizada na **Figura 18** é disponibilizada uma descrição prévia sobre as funcionalidades do sistema como um todo, além de possuir uma análise sobre a aplicabilidade existente em cada uma das dez abas presentes no modelo. Também é evidenciando nesse menu uma interpretabilidade de todos os gráficos apresentados no sistema, ou seja, é descrito como interpretar e analisar as saídas dos plots, sendo que estes foram expostos na localidade inferior da página web. Com relação aos gráficos que foram referenciados acima, é possível uma visualização completa destes por meio de auxílio do pacote *Plotly* que possibilita uma interação em tempo real com os valores expostos no gráfico, além de permitir um conjunto de funcionalidades extras que beneficiariam numa comparação direta e eficiente entre valores apresentados nas variáveis, além disso, também é possível a realização de um download simples e prático para armazenamento dos gráficos caso seja de interesse do usuário.

A **Figura 19**, Quadro Resumo, serve para ilustrar o Dashboard desenvolvido e utilizado de maneira dinâmica para referenciar o controle geral de todos os processos finalizados ou em andamento presente na instituição até o atual momento. Inicialmente no topo da página, existem infoboxes capazes de informar as medidas quantitativas dos pleitos segundo suas classificações ou status, posteriormente ainda na mesma aba, foram plotados dois gráficos para salientar um maior detalhamento das informações referentes a cada classe processual já existente no banco, vale ressaltar, que a informação presente no quarto gráfico, gráfico de setores, possui como intuito destacar quais classes de pleitos estão gerando uma maior demanda de tempo em dias da instituição.

5ª Vara da Justiça Federal

Login

Username

Password

Log in



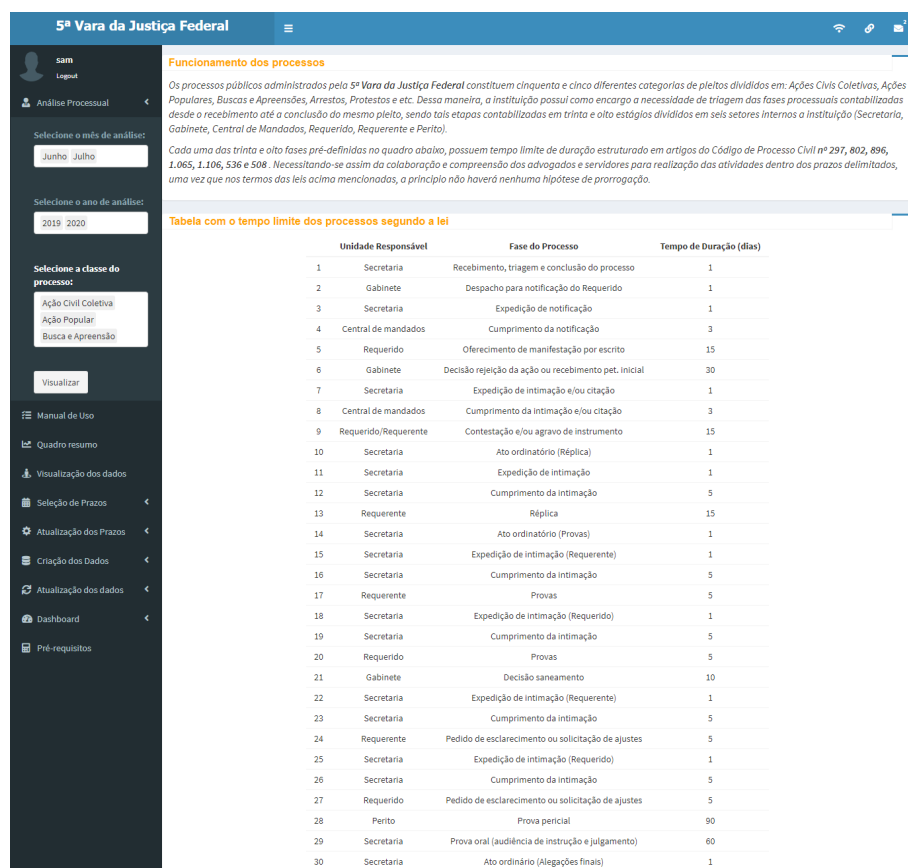


JUSTIÇA FEDERAL



UnB

Figura 16 – Tela referente a Identificação de Usuário



31	Secretaria	Expedição de intimação (Requerente)	1
32	Secretaria	Cumprimento da intimação	5
33	Requerente	Alegações finais	15
34	Secretaria	Expedição de intimação (Requerido)	1
35	Secretaria	Cumprimento da intimação	5
36	Requerido	Alegações finais	15
37	Secretaria	Recebimento e conclusão do processo	1
38	Gabinete	Sentença	30



Figura 17 – Tela referente a Análise processual

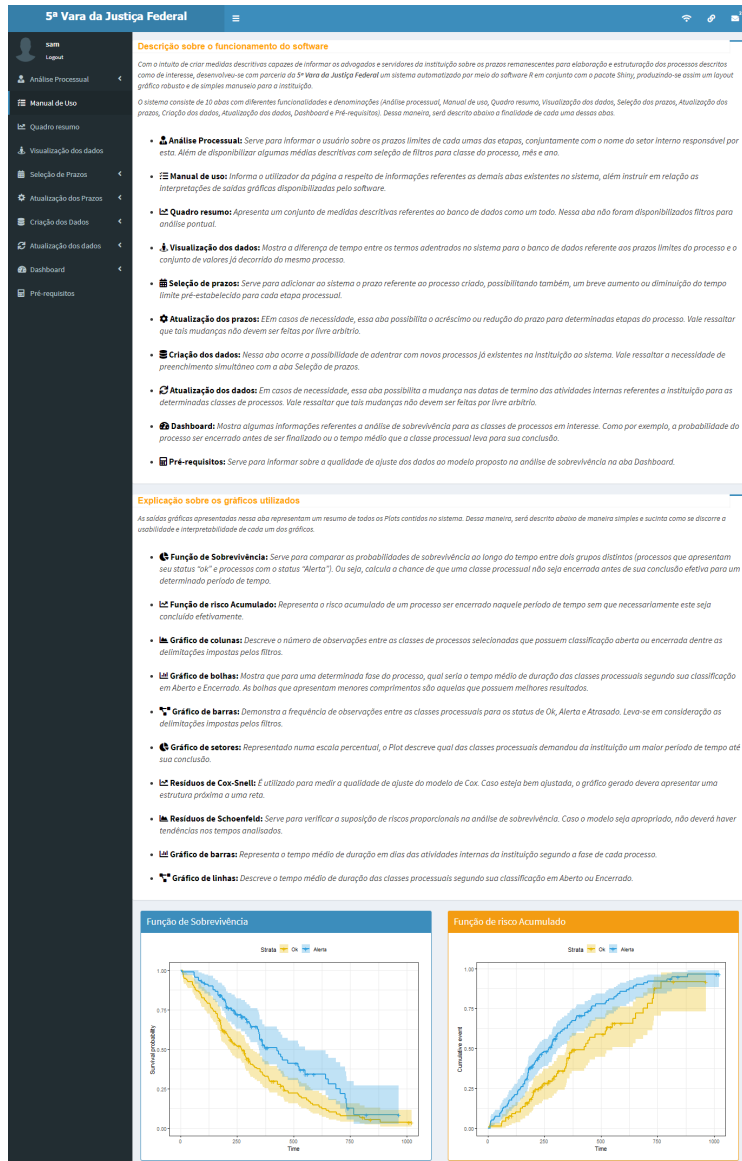


Figura 18 – Tela referente ao Manual de uso

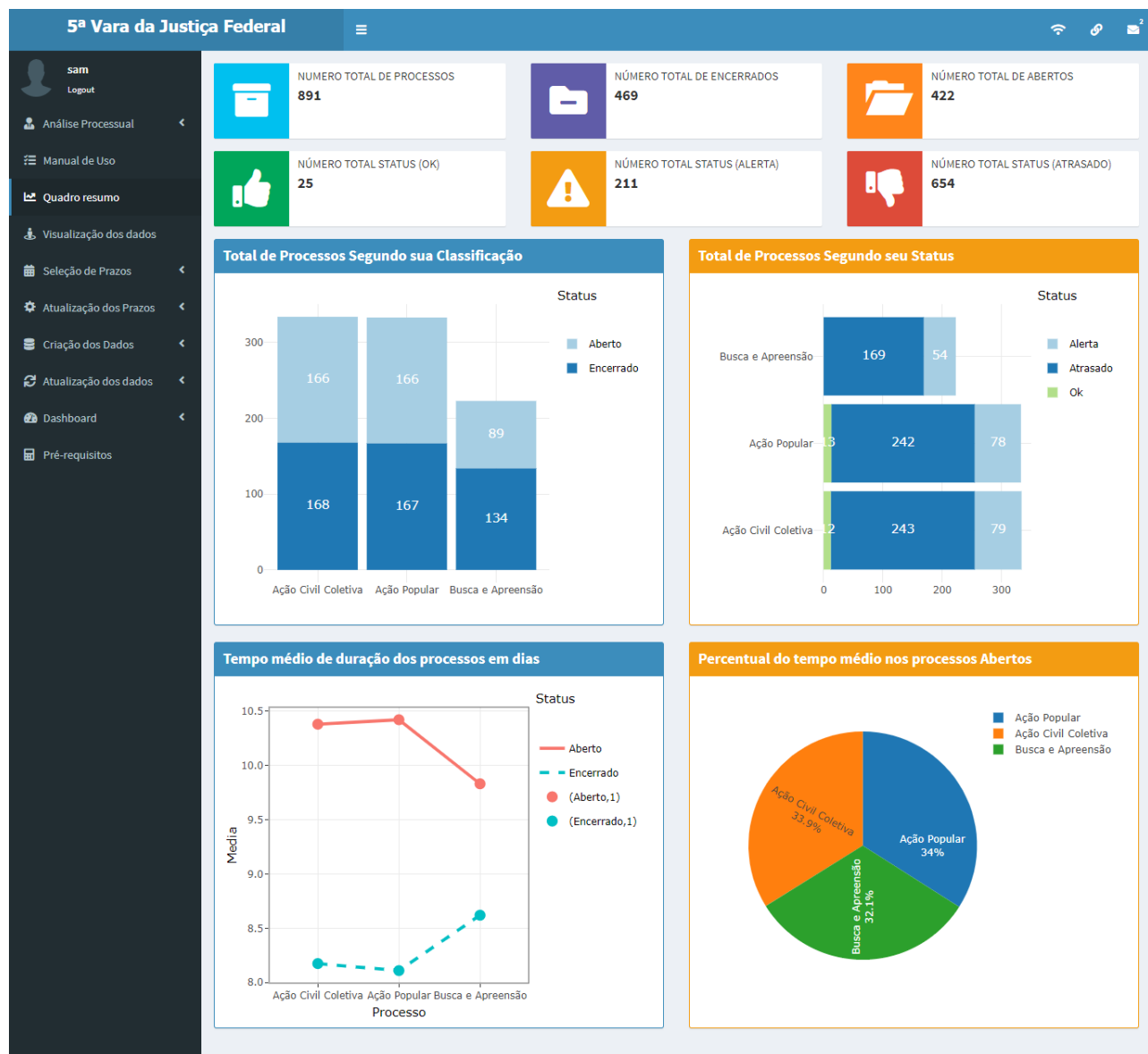


Figura 19 – Tela referente ao Quadro resumo

O quarto menu da página Web, **Figura 20**, utiliza uma estrutura de tabela dinâmica para evidenciar uma visualização das relações entre as diferenças de valores em dias já decorridos dos pleitos e os prazos de duração dos mesmos, ou seja, o sistema calcula de maneira automatizada o tempo faltante ou excedente para a realização das atividades nas etapas de interesse, caso estas ainda não estejam com seus status concluídos. De maneira geral, a aba referenciada serve para englobar e expor os períodos em dias já decorridos ou faltantes de todos os processos presentes no banco de dados e viabilizar a possibilidade de o usuário ter acesso de maneira direta aos dados por meio de impressão (**Figura 21**), cópia de observações que foram enquadradas no layout da página na tabela (existe a possibilidade de seleção do número de observações que se deseja dispor para cada página) ou download via CSV, EXCEL ou PDF.

Vale ressaltar que as opções de obtenção dos dados citados no parágrafo anterior também estão presentes em todas as tabelas subsequentes do sistema, ou seja, encontram-se disponíveis nas abas número cinco, seis, sete e oito.

Na **Figura 22**, Seleção dos Prazos, o usuário tem a possibilidade de adentrar com os novos prazos referentes as novas etapas que serão cumpridas pela instituição de maneira automática (apenas explicitando número do processo e apelido do pleito) ou selecionar de maneira manual os valores desejados utilizando-se de um input em scroll lateral que se referencia aos períodos de duração a serem seguidos em algumas ou em várias diferentes etapas do processo. Caso o utilizador do sistema necessite efetuar algumas mudanças em relação aos prazos selecionados anteriormente, também é possível a utilização da aba número 6, **Figura 23**, para esta ação, ou seja, adição de etapas e mudanças de valores / períodos são realizadas de maneira simples e direta por meio do próprio usuário em relação à página web.

Posteriormente à adição dos prazos, é necessário ao usuário a imputação dos dados brutos de cada um dos processos referenciados, destacando-se assim apenas a carência de seleção dos períodos de início de cada uma das etapas presentes no processo, **Figura 24**, ou seja, o usuário não necessita memorizar datas de etapas anteriores para efetuar o cadastramento do processo ao sistema. Dessa maneira, serão disponibilizados aos utilizadores inputs em formatos de calendários, Scrolls e textos para escrita e seleção das informações básicas para manipulação, análises e descrição dos pleitos. Assim como ocorrido anteriormente com a **Figura 23**, foi implementado uma aba, **Figura 25**, capaz de realizar atualizações de dados para etapas anteriores a atual ou mudanças nas datas de etapas futuras, já que essas por default admitem a alocação de valores zero.

5ª Vara da Justiça Federal

Copy Print Download Search:

Logout

Análise Processual

Manual de Uso

Quadro resumo

Visualização dos dados

Seleção de Prazos

Atualização dos Prazos

Apelido do Processo:

aaaa

Selecione a Fase:

Recebimento, triagem e conclusão do processo

Prazo de duração da etapa

0 1 2 3 4 5

Atualizar

Criação dos Dados

Nº Processo	Apelido	Recebimento, triagem e conclusão do processo	Despacho para notificação do Requerido	Expedição de notificação	Cumprimento da notificação	Oferecimento de manifestação por escrito	Decisão rejeição da ação ou recebimento pet. inicial	Expedição de intimação e/ou citação	Cumprimento da intimação e/ou citação	
892	1123XXX	aaaa	1	1	1	3	15	30	1	3

Previous 1 Next

Figura 23 – Tela referente a Atualização dos prazos

5ª Vara da Justiça Federal

sam

Logout

Análise Processual

Manual de Uso

Quadro resumo

Visualização dos dados

Seleção de Prazos

Atualização dos Prazos

Criação dos Dados

Número do Processo

Enter text....

Apelido do Processo

Enter text...

Classificação do processo:

Aberto

Selecione a classe do processo:

Ação Civil Coletiva

Data de Início

2020-07-13

Selecione a Fase

Recebimento, triagem e conclusão do processo

Data do termino

2020-07-13

Criar

CopyPrintDownloadShow10▼entries

Search:

Nº Processo	Apelido	Classificação	Status	Classe do Processo	Data Inicial	Recebimento, triagem e conclusão do processo	Despacho para notificação do Requerido	Expedição de notificação	Cumprimento da notificação	Oferecimento de manifestação por escrito
1	1000XXX	Apelido878	Encerrado	Alerta	Ação Civil Coletiva	2018-06-06	0	4	1	4
2	1001XXX	Apelido879	Encerrado	Alerta	Ação Civil Coletiva	2018-07-10	4	1	5	5
3	1002XXX	Apelido880	Encerrado	Alerta	Ação Civil Coletiva	2017-10-13	4	3	1	2
4	1003XXX	Apelido881	Encerrado	Alerta	Ação Civil Coletiva	2019-05-20	1	1	1	5
5	1004XXX	Apelido882	Encerrado	Alerta	Ação Civil Coletiva	2017-10-25	4	2	1	3
6	1005XXX	Apelido883	Encerrado	Alerta	Ação Civil Coletiva	2018-08-17	4	5	2	4
7	1006XXX	Apelido884	Encerrado	Alerta	Ação Civil Coletiva	2019-02-06	1	1	4	1
8	1007XXX	Apelido885	Encerrado	Alerta	Ação Civil Coletiva	2019-08-26	5	2	2	4
9	1008XXX	Apelido886	Encerrado	Alerta	Ação Civil Coletiva	2017-08-10	1	1	1	1

Previous

1

2

3

4

5

...

90

Next

Figura 24 – Tela referente a Criação de dados

Figura 25 – Tela referente a Atualização dos dados

Na **Figura 26**, Dashboard, é possível o usuário do sistema selecionar por meio de filtros os meses a serem analisados, os anos de interesse, a classificação do pleito e a classe processual que para si possui relevância, podendo-se assim obter as medidas estatísticas que estão disponíveis na área da análise de sobrevivência e que para o estudo são úteis. Dessa maneira, inicialmente logo no topo da página, são evidenciadas medidas descritivas como: número de processos existentes, função de sobrevivência, taxa de risco acumulada, taxa de risco, tempo médio de sobrevivência e a vida média residual para todos os processos selecionados pelos filtros existentes na lateral esquerda do browser. Vale ressaltar que os infoboxes apresentados são dinâmicos em cores e desenhos para assim oferecer uma simplicidade maior na compreensão dos valores expostos.

Posteriormente ainda na figura acima referenciada, são gerados os gráficos para as análises das funções de sobrevivência e risco acumulado ao longo do tempo. Ressaltando-se a importância visual do primeiro gráfico (Função de Sobrevivência) devido à necessidade de averiguar os pré-requisitos existentes no modelo de Cox relacionados a existência de sobressaliência entre as linhas dos três status processuais representados no plot. Além disso, também são realizados dois gráficos adicionais, sendo eles: gráfico de barras horizontais referente ao prazo médio em dias para execução de cada uma das atividades internas da instituição e o gráfico de setores que evidencia o tempo médio utilizado por cada área interna da 5ª vara da justiça federal para realização dos trabalhos, possibilitando assim, uma visão geral de quais setores necessitam de apoio técnico ou melhoria na produtividade.

A última aba da página Web, **Figura 27**, é composta por dois testes estatísticos voltados para a área de análise de sobrevivência, sendo estes utilizados para verificar a qualidade do ajuste da função de sobrevivência ao modelo proposto em estudo (**Surv(Tempo de vida, Grupo de censura)-etapa processual**), ou seja, o modelo apresentado leva em consideração a presença de interceptos em cada uma das etapas processuais presentes no banco.

Dessa maneira utilizou-se dos resíduos de Cox-Snell, representado por um gráfico de pontos, para tentar medir a qualidade da modelagem proposta. Assim, caso o plot se ajuste a uma reta, este indicaria uma ótima qualidade de modelagem dos dados, ou caso contrário, ocorreria uma dispersão de pontos. Já o segundo teste referenciado aos resíduos de Schoenfeld, que também é representado por meio de um plot, serve para evidenciar a qualidade de ajuste. Caso o modelo de Schoenfeld não apresente nenhuma tendência de linearidade ($H_0: \rho = 0$), admite-se uma boa qualidade na adequação dos dados ao modelo proposto na função de sobrevivência.

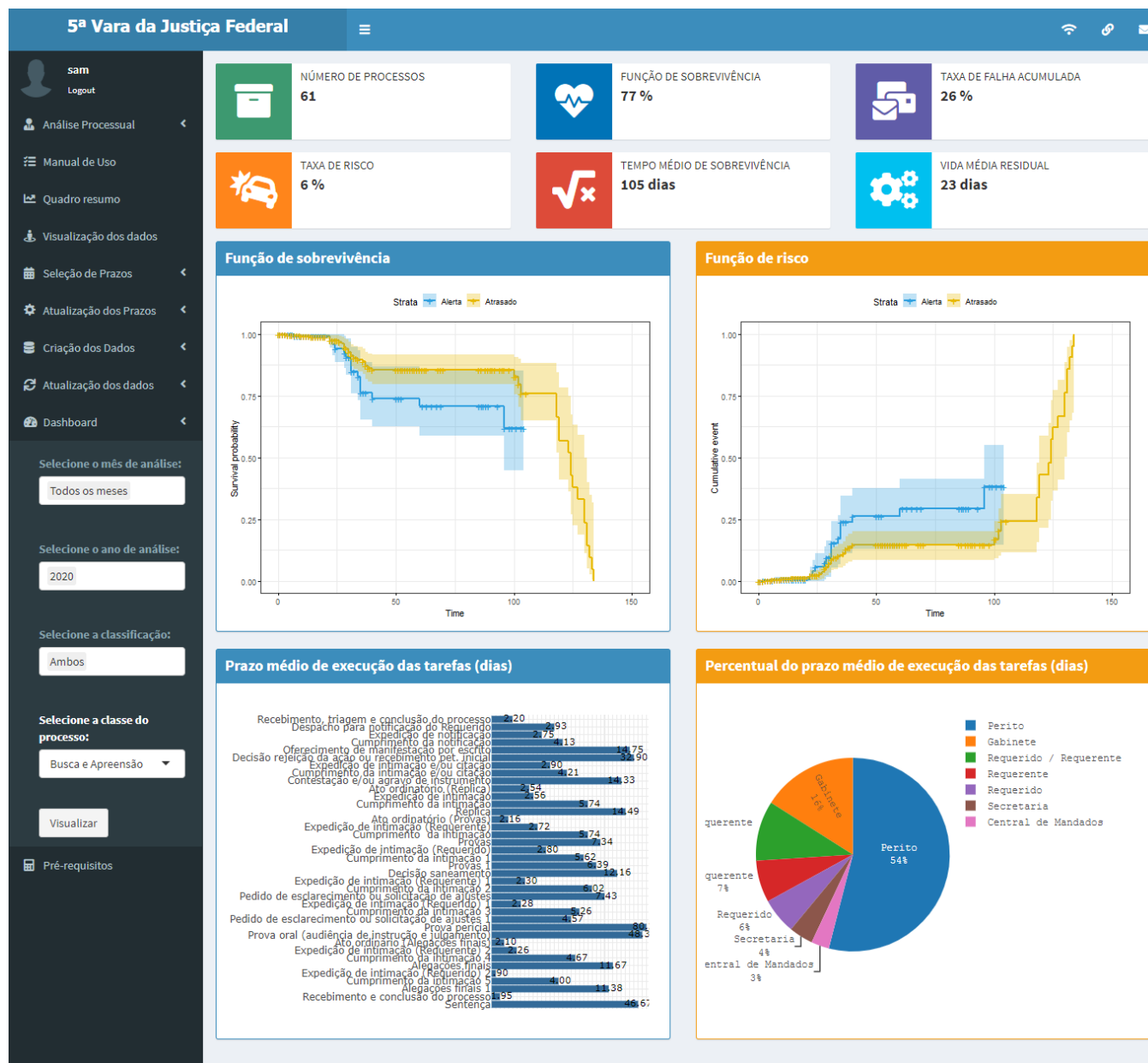


Figura 26 – Tela referente ao Dashboard

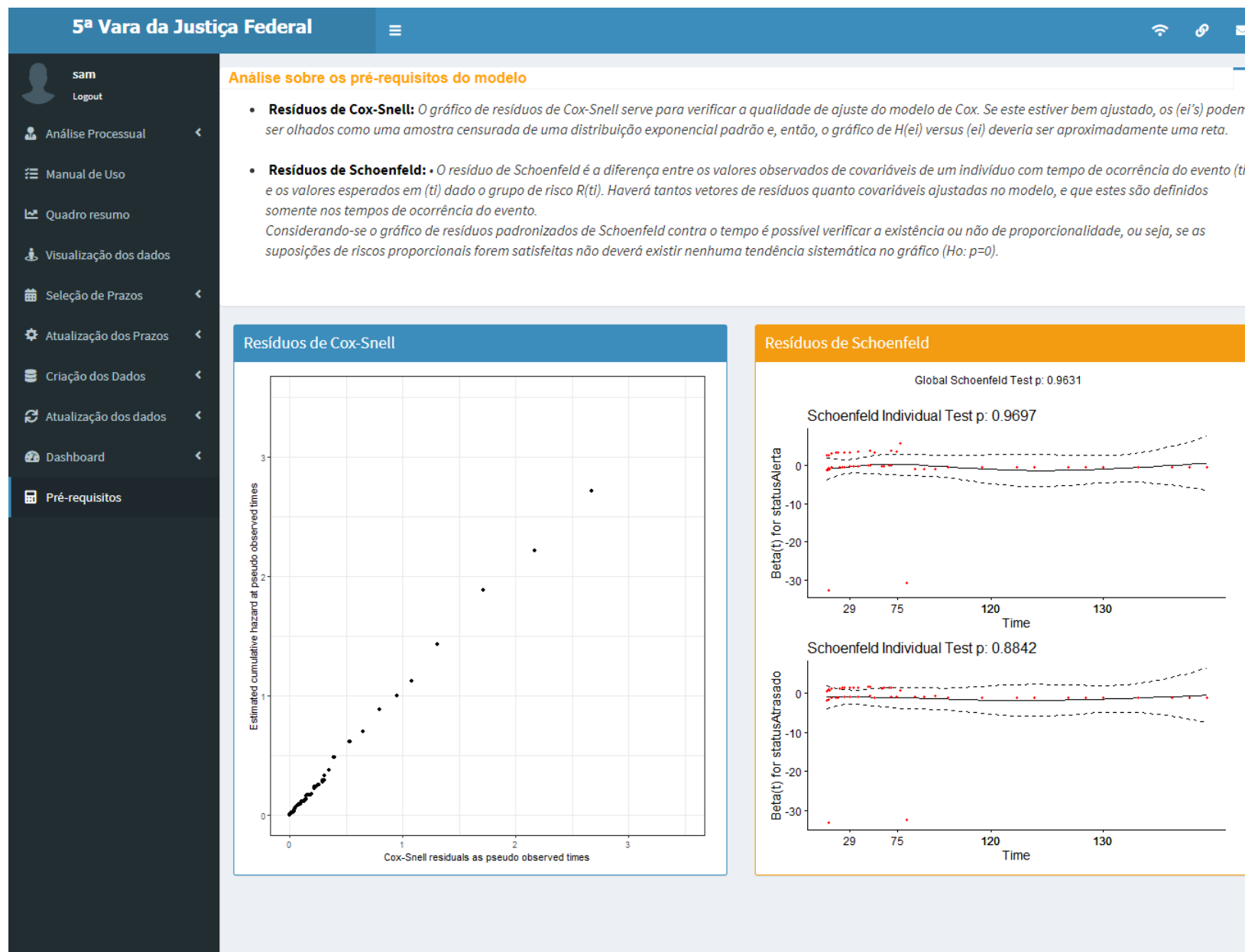


Figura 27 – Tela referente aos Pré-requisitos do modelo

5 CONCLUSÃO

O trabalho acima desenvolvido foi realizado com o intuito de otimizar o controle interno dos processos elaborados pela instituição, sendo estes obtidos por meio da parceria entre o departamento de estatística e a 5ª vara da justiça federal. Dessa maneira, observou-se um avanço técnico relacionado a capacidade de visualização e análise em tempo real quando comparado ao antigo modelo de análise existente na instituição (escrita e atualização manual de pleitos por meio de planilhas em excel).

O sistema desenvolvido foi programado utilizando-se o software R com auxílio de alguns pacotes, destacando-se as bibliotecas Shiny e ShinyDashboard, que permitiram a realização de uma página web interativa que possuísse separações simples de layout em Bootstrap com capacidade de auxiliar diretamente o usuário na hora de selecionar as abas de interesse. Dessa maneira, estruturou-se dez abas que possuíssem funcionalidades distintas (análises exploratórias do banco, descrição das informações, visualização dos dados e análise de pré-requisitos estatísticos) para assim poder englobar todas as necessidades apresentadas em reuniões pelos membros da instituição. Mas caso ocorra necessidades futuras de uma manipulação diferente dos dados, também existe a possibilidade de obtenção do banco por meio de download pelas abas internas.

Complementarmente as atividades voltadas a análise de sobrevivência desenvolvida na página Web, também foi elaborado um estudo aprofundado nessas mesmas métricas com base nos dados presentes na classificação Pje para os anos de 2014 a 2019. Tendo estes evidenciado que processos que possuíram tempo de duração superior ou igual a seiscentos e vinte e cinco dias apontam probabilidade de conclusão antes do arquivamento quase nulo, já em contrapartida, por meio do gráfico da Taxa de falha acumulada obtém-se chance de falha quase de 100% após o mesmo período. Além disso, vale ressaltar que grande parte das classes processuais analisadas pela **Figura 24** sofreram a ação de interesse antes de trezentos dias.

Como medida para melhorar a compreensão sobre a distribuição e comportamento dos dados, também foi realizada uma regressão paramétrica do modelo para selecionar se o melhor ajuste é referente a distribuição Exponencial, Weibull, Log-Logística ou Log-Normal. Dessa maneira, considerando-se os resultados apresentados na **Figura 14** e **Tabela 2** evidencia-se que o ajuste Exponencial enquadra-se melhor para o conjunto de dados analisados com parceria da 5ª Vara da Justiça Federal, o que já seria esperado já que as variáveis utilizadas em análise na função de sobrevivência e regressão se caracterizam num modelo de contagens.

6 CRONOGRAMA

Tabela 8 – Cronograma - 2/2019

Atividades	Jul	Ago	Set	Out	Nov	Dez
Escolha do tema a ser abordado						
Estudo do tema						
Estudo de técnica						
Elaboração de relatório parcial						
Entrega e correção do relatório parcial						

Tabela 9 – Cronograma - 1/2020

Atividades	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set
Desenvolvimento dos modelos em estudo								
Análise dos Resultados								
Elaboração do relatório Final								
Entrega do relatório final ao prof. orientador								
Correção do relatório final								
Apresentação do relatório final								
Entrega do relatório final a banca								

7 REFERÊNCIAS

COLOSIMO, E.A.; GIOLO, S.R. ANÁLISE DE SOBREVIVÊNCIA APLICADA. SÃO PAULO: EDGARD BLUCHER, 2006.

J.F. LAWLESS, STATISTICAL MODELS AND METHODS FOR LIFETIME DATA, JOHN WILEY SONS, NEW YORK, 1982.

COX, D.R. REGRESSION MODELS AND LIFE TABLES. JOURNAL OF ROYAL STATISTICAL SOCIETY. SERIES B, V.39, P.1-38, 1972.

COX, D.R. PARTIAL LIKEHOOD. BIOMETRIKA, V.62, N.2, P.269-276, MAR. 1975.

GIOLO, S. R. MODELOS DE ANÁLISE DE SOBREVIVÊNCIA PARA EXPERIMENTOS DOSE-RESPOSTA. CAMPINAS: DISSERTAÇÃO DE MESTRADO, 1994.

GIOLO, S. R. VARIÁVEIS LATENTES EM ANÁLISE DE SOBREVIVÊNCIA E CURVAS DE CRESCIMENTO. PIRACICABA: TESE DE DOUTORADO, 2003.

HOUGAARD, P. ANALYSIS OF MULTIVARIATE SURVIVAL DATA. NEW YORK: SPRINGER VERLAG, 2000.

KAPLAN, E. L.; MEIER, P. NON-PARAMETRIC ESTIMATION FROM INCOMPLETE OBSERVATIONS. JOUR-NON-PARAMETRIC STATISTICAL ASSOCIATION, V.53, P.547-481, 1958.

KLEIN, J. P, MOESCHBERGER, M. SURVIVAL ANALYSIS. NEW YORK: SPRINGER VERLAG, 1997.

LINDSEY, J.C.; RYAN, L.M. TUTORIAL IN BIOESTATISTICS: METHODS FOR INTERVAL-CENSORED DATA. STATISTICS IN MEDICINE. V.17, P.219-238, 1998.

LAWLESS, J. F. STATISTICAL MODELS AND METHODS FOR LIFETIME DATA. NEW YORK: JOHN WILEY SONS, 1981.

LEE, Elisa. STATISTICAL METHODS FOR SURVIVAL DATA ANALYSIS. BELMONT: LIFETIME LEARNING PUBLICATIONS, 1980.

LOUZADA NETO, F., MAZUCHELI, J., ACHCAR, J.A. INTRODUÇÃO À ANÁLISE DE SOBREVIVÊNCIA E CONFIABILIDADE. MARINGÁ: MINICURSO APRESENTADO NA III JORNADA REGIONAL DE ESTATÍSTICA, 2002.

BRASIL. PODER JUDICIÁRIO JUSTIÇA FEDERAL, CORREGEDORIA-REGIONAL DA JUSTIÇA FEDERAL DA 2ª REGIÃO. MANUAL DE ROTINAS E PROCEDIMENTOS INTERNOS. BRASÍLIA, DF, 2009.

KONRATH ET AL.; ANDRÉA ET AL., DESENVOLVIMENTO DE APLICATIVOS WEB COM R E SHINY: INOVAÇÕES NO ENSINO DE ESTATÍSTICA, ABAKÓS, BELO HORIZONTE, MAIO DE 2018.

PACHECO ET AL.; ANTONIO ET AL., APRENDENDO R, ESCOLA NACIONAL DE SAÚDE PÚBLICA - FIOCRUZ, RIO DE JANEIRO, 2014.

A. KAUFMANN; D. GROUCHKO; R. CRUON; MATHEMATICAL MODELS FOR STUDY OF THE REABILITY OF SYSTEMS; UNIVERSITY OF SOUTHERN CALIFORNIA; VOLUME 124; 1997.

GIOLO, S. R. MODELOS DE RISCOS PROPORCIONAIS. PARANÁ: UNIVERSIDADE FEDERAL DO PARANÁ, 2018.

RUPERT G. MILLER; JR. RUPERT, SURVIVAL ANALYSIS, WILEY-INTERSCIENCE, 1998.

MARILIA ET AL; VALESKA ET AL., ANALISE DE SOBREVIVENCIA: TEORIA E APLICAÇÕES EM SAUDE, FIOCRUZ, 2019.