



**Universidade de Brasília
Departamento de Estatística**

**Determinantes do tempo até a baixa de processos
na Justiça do Trabalho**

Kassyano Kevyn Andrade de Souza

Relatório Final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Kassyano Kevyn Andrade de Souza

**Determinantes do tempo até a baixa de processos
na Justiça do Trabalho**

Orientador(a): Thais Carvalho Valadares Rodrigues

Relatório Final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Dedico este estudo à classe operária, que tem na venda de sua força de trabalho o único meio de garantir a sua sobrevivência, mas cujo trabalho é o que garante a nossa própria existência.

Agradecimentos

Agradeço à minha ilustre, incrível, impecável, incomparável, insuperável e insubstituível orientadora, Thais Rodrigues, por ter me acompanhado e instruído durante todo o projeto.

Aos meus colegas de faculdade e amigos com quem estudei durante os últimos anos, André Vale, Carolina Musso, César Galvão, Davi Guerra, Eduardo Côrtes, Emilly Marques, Hermes Winarski, Juliana Degani, Juliana Rocha, Juliana Rosa, Laura Melo, Matheus Martinez, Maria Luiza Escobar, Micael Papa, Pedro Tepedino, Rafael de Acypreste, a recém-chegada Rita Veloso, Sabrina França, Silvânia Andrade e Tamara Dias, agradeço por fazerem parte da minha vida. A experiência da universidade, tampouco a experiência da própria vida, jamais teria sido a mesma sem vocês.

Aos meus amigos de longa data, Ana Lúcia Amado, Camylla Meirelly, Daniel Badaró, Diego Guedes, Fernanda Gomes, Gabriel Machado, Hudson Pimenta, Letícia Pimenta, Rodrigo de Carvalho, Rodrigo de Freitas, Samuel Barbosa, Silas Bismarck e Tatianne Gomes, agradeço por estarem comigo nos comigo nos piores momentos e terem me incentivado constantemente a seguir os meus sonhos. Ao meu amigo, Lourival Silva, eternizo sua memória com este agradecimento póstumo.

Aos meus colegas do Eixo 4, Ana Rezende, Carlos Becker, Genesis Pereira, Igor Stemler, Isabely Mota, Rayssa Coátio, Ricardo Guidoni e Viviane Fecher, agradeço por terem me incentivado e ajudado a compreender os dados com os quais trabalhei nesse projeto.

Agradeço aos meus webamigos Eduardo Moreira (EliminatorVenom/BLU Spy), Lucas Silva (DeusDoMalDDM), Matheus Rabelo (Takato Blue) e Vinnie Quintana, pelos mais de 10 anos nos quais no conhecemos, e por todas as fases que atravessamos nesse webmundo. A presença de vocês para mim é tão real quanto toda a realidade que me cerca.

Agradeço à minha cachorra, Gabi, e ao meu gato, Thomas, embora o fato de que eles ainda não chegaram à idade de aprender a ler os impeça de tomarem ciência deste agradecimento. Agradeço também à Princesa e à Ursa, as guardiãs noturnas que nos protegem todos os dias.

À minha mãe, Neuza Andrade, e meu pai, Gilmar Souza, agradeço por todo o apoio que me deram para que eu pudesse chegar até aqui.

Por fim, a cada Kevyn dos outros infinitos mundos possíveis (exceto os maus) que seguiram por caminhos diferentes do meu, agradeço pela perseverança. Espero que tenham superado o sofrimento sutil da impermanência e a insatisfação inerente à existência, e encontrado a paz em seus caminhos.

Lista de Tabelas

| | | |
|----|---|----|
| 1 | Quantidade de órgãos julgadores por tribunal | 20 |
| 2 | Teste de normalidade de Shapiro-Wilk para amostras do tempo até a baixa | 22 |
| 3 | Frequência de processos pendentes cruzada por grau e recursos | 27 |
| 4 | Testes para Independência entre Variáveis Categóricas | 27 |
| 5 | Características das varas com menos de 120 processos tramitando e mais de 1.000 dias de tempo até a baixa | 34 |
| 6 | Qualidade do ajuste para os modelos quantílicos para a mediana, sendo o Modelo 1 (sem interações) e Modelo 2 (com interações) | 38 |
| 7 | AIC, BIC, Perda e número de coeficientes para os ajustes do Modelo 1 (sem interações) e do Modelo 2 (com interações) | 39 |
| 8 | Coeficientes significativos para o Modelo 1 (sem interações) e o Modelo 2 (com interações) para os quantis $\tau=0,1$, $\tau=0,25$, $\tau=0,5$, $\tau=0,75$, $\tau=0,9$. . . | 40 |
| 9 | Coeficientes dos modelos selecionados para os quantis $\tau=0,1$, $\tau=0,25$, $\tau=0,5$, $\tau=0,75$, $\tau=0,9$ | 42 |
| 10 | Coeficientes significativos para cada modelo gaussiano | 46 |
| 11 | AIC, BIC, MAE e número de coeficientes para os modelos gaussianos . . . | 46 |

Lista de Figuras

| | | |
|----|---|----|
| 1 | Dispersão de pontos gerados aleatoriamente (elaboração própria) | 9 |
| 2 | Distribuição dos tempos entre o início do processo e a baixa (em dias transcorridos) | 21 |
| 3 | Evolução da frequência relativa de casos novos físicos ao longo do tempo . | 23 |
| 4 | Evolução da frequência relativa de processos pendentes tramitando em formato físico ao longo do tempo | 23 |
| 5 | Distribuição dos tempos de tramitação dos processos para cada formato em processos ainda ativos | 24 |
| 6 | Distribuição dos tempos até a baixa dos processos para o primeiro e o segundo grau de jurisdição | 25 |
| 7 | Distribuição dos tempos até a baixa dos processos originários e recursais . | 26 |
| 8 | Distribuição dos tempos até a baixa dos processos segundo o procedimento | 28 |
| 9 | Quantidade de processos em cada métrica | 30 |
| 10 | Matriz de correlações entre as variáveis explicativas quantitativas | 31 |
| 11 | Gráficos de dispersão entre as variáveis explicativas e o tempo até a baixa com retas de regressão quantílica (mediana, em azul) e regressão gaussiana (média, em verde) | 33 |
| 12 | Gráficos de dispersão entre as variáveis explicativas e o tempo até a baixa com retas de regressão quantílica (mediana, em azul) e regressão gaussiana (média, em verde), com remoção de outliers | 35 |
| 13 | Coefficientes angulares para as principais variáveis | 43 |

Lista de Modelos

| | | |
|---|--|----|
| 1 | Regressão quantílica ($\tau = 0,5$) | 36 |
| 2 | Regressão quantílica ($\tau = 0,5$) com interações e componentes quadráticos . | 37 |
| 3 | Regressão gaussiana | 44 |
| 4 | Regressão gaussiana com interações | 45 |

Sumário

| | |
|---|----|
| 1 Introdução | 8 |
| 2 Referencial Teórico | 9 |
| 2.1 Regressão Gaussiana | 10 |
| 2.2 Regressão Quantílica | 11 |
| 2.2.1 Inferência para os coeficientes do modelo de regressão quantílica . . | 14 |
| 2.3 Seleção de variáveis | 15 |
| 2.3.1 Regressão por <i>stepwise</i> | 16 |
| 3 Metodologia | 18 |
| 3.1 Conjunto de dados | 18 |
| 4 Resultados | 20 |
| 4.1 Análise exploratória | 20 |
| 4.1.1 Quantidade de Órgãos Julgadores | 20 |
| 4.1.2 Distribuição dos tempos | 21 |
| 4.1.3 Formato dos processos | 22 |
| 4.1.4 Graus de jurisdição | 25 |
| 4.1.5 Recursos | 26 |
| 4.1.6 Associação entre as variáveis qualitativas | 26 |
| 4.1.7 Procedimentos | 28 |
| 4.1.8 Indicadores quantitativos | 29 |
| 4.1.9 Tempo até a baixa por indicador | 32 |
| 4.2 Modelos | 36 |
| 4.2.1 Modelos para a mediana | 36 |
| 4.2.2 Modelos de regressão para todos os quantis | 38 |
| 4.2.3 Modelos de regressão gaussiana | 44 |
| 5 Conclusão | 48 |
| Referências | 50 |

1 Introdução

O Poder Judiciário é um dos poderes fundamentais da estrutura tripartite do Estado brasileiro. Suas ações estão diretamente atreladas ao funcionamento das instituições de Estado e à manutenção da justiça, suscitando interesse para a administração pública. O entendimento da dinâmica judicial é o que garante sua transparência e credibilidade frente à sociedade civil.

Nos últimos anos, o emprego de tecnologias para aumento de eficiência dos serviços prestados tem sido uma preocupação do Poder Judiciário, e o Conselho Nacional de Justiça (CNJ) tem proposto medidas para cumprir esse fim, tais como o Processo Judicial Eletrônico (CNJ, 2013) e o Juízo 100% Digital (CNJ, 2020b).

Para este trabalho, a principal variável analisada será o tempo compreendido entre o início do processo e sua baixa, que pertence à classe de indicadores de tempo de tramitação. As baixas, por sua vez, são definidas como quaisquer processos que sejam “[...] a) remetidos para outros órgãos judiciais competentes, desde que vinculados a tribunais diferentes; b) remetidos para as instâncias superiores ou inferiores; c) arquivados definitivamente.” (CNJ, 2023).

A relevância dessa classe de indicadores se dá pelo fato de que uma baixa configura um encerramento da situação de pendente de um processo. O presente trabalho visa avaliar como esse tempo se distribui entre os órgãos julgadores da Justiça do Trabalho, e quais variáveis podem influenciar em sua redução ou aumento. Tal propósito será cumprido com a construção de um modelo de regressão quantílica. O modelo vai permitir observar quais varas estão abaixo ou acima de determinada faixa, analisar tendências e extrair resultados probabilísticos.

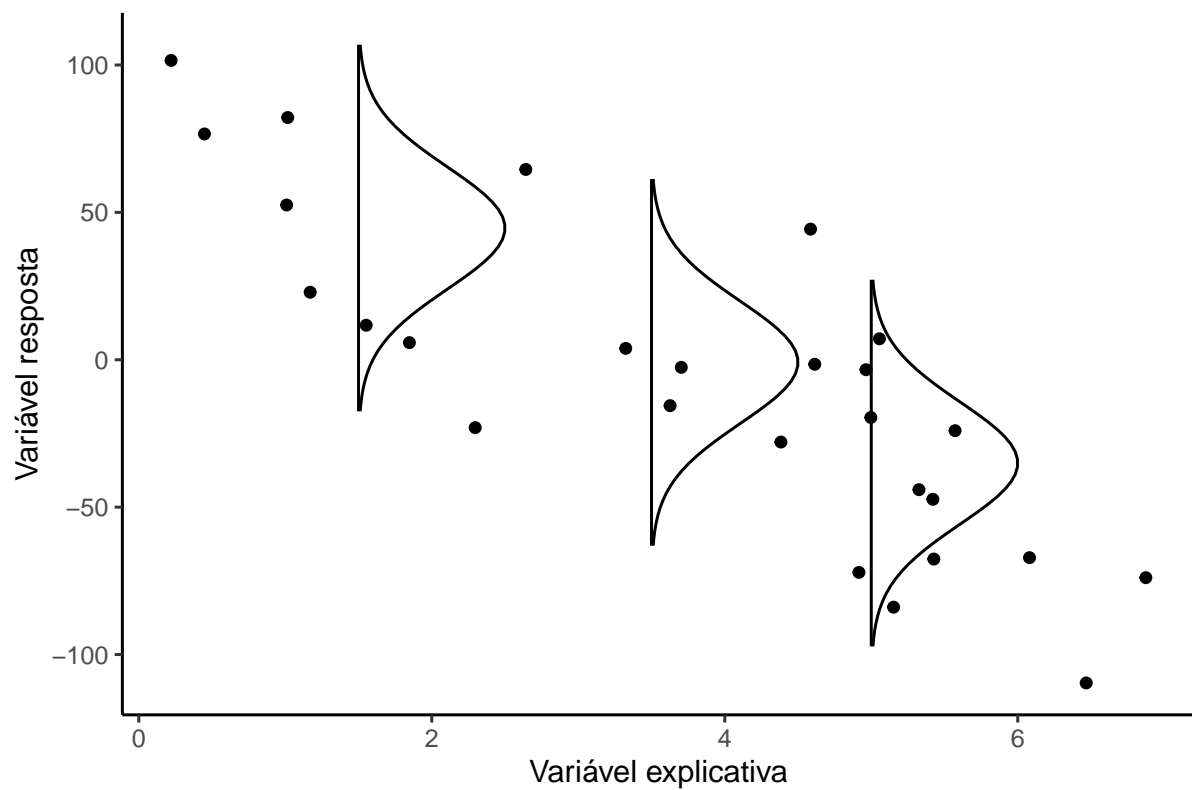
O objetivo deste trabalho é estudar o tempo até a baixa de processos no ramo da Justiça do Trabalho, utilizando dados do Poder Judiciário Brasileiro no retrato referente a janeiro de 2024, e descrever como este se distribui com relação às características da vara, de forma que seja possível prever esse tempo dadas as covariáveis disponíveis.

Na Seção 2, será feita uma discussão a respeito dos métodos utilizados para a construção dos modelos. A Seção 3 irá discutir a respeito dos dados e do contexto que motiva o estudo desse fenômeno. Por fim, a Seção 4 terá foco nos resultados, avaliando descritivamente, construindo os modelos propostos e comparando as aplicações das técnicas da regressão quantílica com as aplicações em regressão gaussiana, com seus impactos no estudo da duração dos processos na Justiça do Trabalho.

2 Referencial Teórico

A análise de regressão é um conjunto de métodos estatísticos que buscam estimar a relação entre uma variável dependente Y (também chamada variável resposta) condicionada a uma ou mais características explicativas X . O caso clássico dos métodos de regressão ajusta uma reta que minimiza o quadrado da diferença entre os valores observados da variável resposta e seus valores esperados segundo o modelo de regressão postulado, em torno da qual se encontra uma distribuição de erros, conforme ilustrado na Figura 1.

Figura 1: Dispersão de pontos gerados aleatoriamente (elaboração própria)



Os dois métodos que serão utilizados neste trabalho são a regressão gaussiana e a regressão quantílica, que estimam, respectivamente, a média condicional $E(Y|X)$ e os quantis condicionais $Q(Y|X)_\tau$.

2.1 Regressão Gaussiana

A regressão gaussiana linear, regressão por mínimos quadrados ou simplesmente regressão linear, é um método que ajusta uma reta para a esperança condicional $E(y_i|\mathbf{x}_i)$, minimizando a diferença quadrática entre a reta dos valores preditos do modelo e o conjunto de valores observados (MONTGOMERY et al., 2012). Sejam $\varepsilon_i = y_i - E(y_i|\mathbf{x}_i)$ os erros aleatórios do modelo, β_0 o coeficiente linear, β_k os coeficientes angulares da regressão e x_{ik} o valor fixado da variável explicativa k na observação i , para $k = 1, \dots, p$. Os pressupostos desse modelo são:

1. $E(y_i|\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$. A esperança $E(y_i|\mathbf{x}_i)$ pode ser descrita como uma combinação linear entre os parâmetros do modelo e as variáveis explicativas.
2. $E(\varepsilon_i) = 0, \forall i$. Os erros têm média zero.
3. $Var(\varepsilon_i) = \sigma^2, \forall i$. A variância dos erros é constante. Essa suposição implica que a variável resposta também tem variância constante.
4. $cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$. Os erros são não-correlacionados entre eles.
5. $\varepsilon_i \sim N(0, \sigma^2), \forall i$. Os erros têm distribuição normal.
6. A matriz de design \mathbf{X} é uma matriz de posto completo, de modo que nenhuma variável pode ser escrita como combinação linear das outras variáveis.

Seja $\mathbf{X}_{n \times (p+1)}$ uma matriz de valores observados das variáveis explicativas com uma coluna de números 1 e \mathbf{y} um vetor n -dimensional de valores observados da variável resposta. É possível demonstrar que, satisfeitos os pressupostos do modelo, o vetor de coeficientes $\hat{\boldsymbol{\beta}}$ é dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y},$$

que constitui um estimador consistente e não viesado de variância uniformemente mínima do vetor de parâmetros $\boldsymbol{\beta}$. Seja

$$MS_E = \frac{1}{n - (p + 1)} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

o estimador não viesado da variância dos erros aleatórios. Segundo Kutner et al. (2008), a matriz de variâncias-covariâncias dos estimadores $\hat{\boldsymbol{\beta}}$ podem ser obtida por

$$s^2(\hat{\beta}) = MS_E(\mathbf{X}^T \mathbf{X})^{-1}. \quad (2.1.1)$$

A partir destes resultados, podemos estabelecer o seguinte teste de hipóteses para avaliação da significância dos coeficientes da regressão:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0,$$

para $k = 0, 1, \dots, p$. Seja $s(\hat{\beta}_k) = \sqrt{s^2(\hat{\beta}_k)}$, conforme definido em 2.1.1. A estatística de teste

$$t^* = \frac{\hat{\beta}_k}{s(\hat{\beta}_k)} \quad (2.1.2)$$

tem distribuição $t(n - p - 1)$ sob H_0 (KUTNER et al., 2008).

Uma vez estabelecido um nível de significância α , hipótese nula H_0 é rejeitada caso $|t^*| > t_{critico}$, onde $t_{critico}$ é o valor da distribuição $t(n - p - 1)$ tal que $F(t_{critico}) < 1 - \frac{\alpha}{2}$. Essa regra de decisão é equivalente a avaliar se o p-valor do teste é menor que o nível de significância α , que será o critério utilizado neste trabalho.

2.2 Regressão Quantílica

Diferente do modelo de regressão gaussiana, que traça a reta para a média de uma variável resposta dependente de variáveis explicativas, a regressão quantílica traça a reta para qualquer quantil condicionado à matriz \mathbf{X} do modelo de regressão, tornando-a mais robusta que o modelo de regressão para a média com base em mínimos quadrados no que se refere à influência de outliers (KOENKER, 2005).

A regressão quantílica depende de menos pressupostos que o modelo de regressão gaussiano que faz uso da inferência de mínimos quadrados, pois não assume normalidade nem variância constante dos erros. A regressão quantílica assume que os erros são não correlacionados entre si e $Q_\tau(\varepsilon_\tau | \mathbf{X}) = 0$, onde $Q_\tau(\varepsilon_\tau | \mathbf{X})$ é o quantil condicional dos erros ε_τ . Embora modelos quantílicos não lineares também sejam possíveis, neste trabalho assumiremos linearidade dos parâmetros dos modelos.

O τ -ésimo quantil de uma distribuição de probabilidades é o ponto y da variável aleatória tal que a probabilidade de um valor ser menor ou igual a y é igual a τ .

Considerando $F(y)$ a função de distribuição acumulada da variável aleatória y , temos que

$$\tau = P(Y \leq y) = F(y).$$

No caso empírico, convencionou-se o uso do valor mínimo do intervalo, de modo que a definição para o quantil amostral assume a forma

$$F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}, \quad (2.2.1)$$

onde \inf é o menor valor do conjunto.

A distribuição empírica permite reduzir a estimativa para o quantil segundo métodos de otimização da função de perda *pinball*, ou perda quantílica. Deseja-se minimizar essa função de perda com relação a uma variável $u \in \mathbb{R}$, podendo ser expressa na forma

$$\rho_\tau(u) = u(\tau - I(u < 0)), \tau \in (0, 1), \quad (2.2.2)$$

onde $I(u < 0)$ é uma função indicadora que assume valor 1 se a condição for satisfeita ($u < 0$) e valor 0, caso contrário.

Considere $u = y - \hat{y}$. Assim sendo, a função ρ assume a forma:

$$\rho_\tau(y - \hat{y}) = (y - \hat{y})(\tau - I((y - \hat{y}) < 0)), \tau \in (0, 1)$$

,

Teremos, portanto, que o valor esperado da função de perda $\rho(y - \hat{y})$ será dado por:

$$E[\rho(y - \hat{y})] = (\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF(y) + \tau \int_{\hat{y}}^{\infty} (y - \hat{y}) dF(y).$$

Como mostra Koenker (2005), é possível minimizar o valor esperado da função de perda igualando sua derivada com relação a \hat{y} a zero:

$$0 = \frac{d}{d\hat{y}} E[\rho(y - \hat{y})] = (1 - \tau) \int_{-\infty}^{\hat{y}} dF(y) - \tau \int_{\hat{y}}^{\infty} dF(y) = F(\hat{y}) - \tau,$$

de modo que a expressão 2.2.2 encontra seu ponto de mínimo para qualquer elemento de

$\{y : F(y) = \tau\}$, ou $\hat{y} = F^{-1}(\tau)$. Em caso de uma solução não única, ou seja, de haver um conjunto de valores que satisfazem a condição, convencionou-se, como já mencionado em 2.2.1, que deve ser selecionado o menor dos valores deste conjunto.

Os métodos podem ser implementados computacionalmente através de algoritmos de otimização. Assim, encontra-se o \hat{y} que minimiza

$$\min_{\hat{y} \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \hat{y}). \quad (2.2.3)$$

Analogamente, podemos reduzir o problema da regressão quantílica a um problema de otimização, minimizando a expressão

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i^{\text{T}} \boldsymbol{\beta}). \quad (2.2.4)$$

Segundo Koenker (2005), é possível solucionar 2.2.3 através de um algoritmo de programação linear. Um problema de programação linear é uma otimização que minimiza ou maximiza uma função (chamada de função objetivo) sujeita a uma restrição. A minimização da função de perda quantílica adiciona $2n$ variáveis artificiais $\{u_i, v_i : 1, \dots, n\}$, onde

$$u_i = (y_i - \hat{y}_i)I(y_i - \hat{y}_i > 0),$$

$$v_i = (y_i - \hat{y}_i)I(y_i - \hat{y}_i < 0),$$

que representam, respectivamente, as partes positivas e negativas do vetor de resíduos, com I sendo uma função indicadora que assume valor 1 se a condição for satisfeita e 0, caso contrário. Note que u_i e v_i não podem ser simultaneamente diferentes de zero. Assim, a solução para 2.2.3 pode ser dada através de

$$\min_{(\hat{y}, \mathbf{u}, \mathbf{v}) \in \mathbb{R} \times \mathbb{R}_+^{2n}} \{\tau \mathbf{1}_n^{\text{T}} \mathbf{u} + (1 - \tau) \mathbf{1}_n^{\text{T}} \mathbf{v} \mid \mathbf{1}_n \hat{y} + \mathbf{u} - \mathbf{v} = \mathbf{y}\},$$

onde o termo $\tau \mathbf{1}_n^{\text{T}} \mathbf{u} + (1 - \tau) \mathbf{1}_n^{\text{T}} \mathbf{v}$ pode ser enunciado como o somatório dos erros ponderados pelo quantil (a função objetivo a ser minimizada) e $\mathbf{1}_n \hat{y} + \mathbf{u} - \mathbf{v} = \mathbf{y}$ é a restrição. Analogamente, 2.2.4 pode ser resolvido com o problema de programação linear expresso em

$$\min_{(\boldsymbol{\beta}, \mathbf{u}, \mathbf{v}) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \{\tau \mathbf{1}_n^T \mathbf{u} + (1 - \tau) \mathbf{1}_n^T \mathbf{v} | \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{u} - \mathbf{v} = \mathbf{y}\},$$

onde $\mathbf{X}_{n \times (p+1)}$ é a matriz de design e $\hat{\boldsymbol{\beta}}$ é o vetor de coeficientes de regressão condicionado a τ . Uma vez resolvido o problema de programação linear, a função quantil condicional $Q_\tau(\mathbf{y} | \mathbf{X})$ pode ser expressa na forma

$$Q_\tau(\mathbf{y} | \mathbf{X}) = \hat{\beta}_0(\tau) + x_1 \hat{\beta}_1(\tau) + \cdots + x_p \hat{\beta}_p(\tau) = \hat{\boldsymbol{\beta}}(\tau) \mathbf{x}$$

onde cada x_i , com $i \in \{1, \dots, p\}$, é uma variável explicativa, e $Q_\tau(\mathbf{y} | \mathbf{X})$ é o quantil condicional τ de \mathbf{y} dado \mathbf{X} (KOENKER, 2005).

Caso a variável resposta tenha distribuição normal e os resíduos sejam homoscedásticos, uma abordagem com a regressão gaussiana também seria capaz de traçar retas quantílicas através da função de distribuição dos resíduos, o que a tornaria preferível, uma vez que é mais eficiente computacionalmente, possui formas fechadas para os estimadores dos coeficientes de regressão e não dependeria de uma estimativa diferente de $\boldsymbol{\beta}$ para cada quantil. Todavia, se a variação de $\boldsymbol{\beta}(\tau)$ for grande em função de τ , significa que os erros não são identicamente distribuídos, tornando a abordagem de regressão quantílica preferível, uma vez que não depende da suposição de homoscedasticidade que não foi satisfeita para o modelo de regressão gaussiana.

2.2.1 Inferência para os coeficientes do modelo de regressão quantílica

A inferência sobre os coeficientes de regressão quantílica pode ser feita de muitas maneiras. Koenker (2005) apresenta três métodos, sendo um baseado em resultados assintóticos (que será o foco neste trabalho), outro em bootstrap e, por fim, um teste baseado em escores ordinais. Koenker e Bassett (1978) demonstraram o resultado que permite a inferência pela distribuição assintótica dos estimadores, como exposto no Teorema 2.1.

Teorema 2.1. *Seja $\{\hat{\beta}_k(\tau_1), \dots, \hat{\beta}_k(\tau_M)\}$, onde $0 < \tau_1 < \cdots < \tau_M < 1$, uma sequência de quantis de regressão únicos estimados para os coeficientes. Seja $\xi(\tau) = F^{-1}(\tau)$ o quantil de nível τ . Assuma que:*

1. F é contínua e tem densidade contínua positiva f em $\xi(\tau_i)$;
2. Uma das colunas da matriz de design $\mathbf{X}_{n \times (p+1)}$ é uma coluna de números 1;
3. $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{X} = \mathbf{Q}$, uma matriz positiva definida.

Então,

$$\sqrt{n}[\hat{\beta}_k(\tau_1) - \beta_k(\tau_1), \dots, \hat{\beta}_k(\tau_M) - \beta_k(\tau_M)] \xrightarrow{D} N_{M \times (p+1)}(\mathbf{0}, \Omega(\tau_1, \dots, \tau_M; F) \otimes \mathbf{Q}^{-1}),$$

onde Ω é a matriz de covariâncias dos M quantis das amostras aleatórias da distribuição F e \otimes é o produto de Kronecker.

Dessa forma, $\hat{\beta}(\tau)$ é um estimador assintoticamente não viesado de $\beta(\tau)$ e possui distribuição assintoticamente normal. Seja $F = \text{diag}(f_1(0), \dots, f_n(0))$. Segundo Kocherginsky, He e Mu (2005), caso os erros ε_i sejam independentes, sem necessariamente serem identicamente distribuídos, o estimador

$$V(\tau) = \tau(1 - \tau)(\mathbf{X}^T \mathbf{F} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{F} \mathbf{X})^{-1} \quad (2.2.5)$$

tende assintoticamente à matriz de covariâncias de $\hat{\beta}(\tau)$. Segundo Santos (2012), é possível estimar $f_i(0)$ pela expressão

$$\frac{2h_n}{\mathbf{x}_i^T \hat{\beta}_{\tau+h_n} - \mathbf{x}_i^T \hat{\beta}_{\tau-h_n}} \quad (2.2.6)$$

onde $\lim_{n \rightarrow \infty} h_n = 0$. Métodos para o cálculo de h_n foram apresentados por Hall e Sheather (1988), citado em Santos (2012) e discutidos por Koenker (2005). A opção de cálculo dos erros padrão `se="nid"`, do método `summary.rq` da biblioteca `quantreg` (em R) implementa a estimação através das expressões 2.2.5 e 2.2.6. Esses resultados assintóticos permitem a construção de intervalos de confiança e testes de hipóteses para os valores de $\beta(\tau)$. As hipóteses

$$H_0 : \beta_k(\tau) = 0$$

$$H_1 : \beta_k(\tau) \neq 0$$

podem ser testadas através dos estimadores assintóticos para as variâncias-covariâncias, com uma estatística de teste análoga à expressão 2.1.2.

2.3 Seleção de variáveis

Ao trabalhar com volumes grandes de dados, métodos para selecionar modelos se tornam relevantes, de forma que evitem tanto a simplificação excessiva quanto o so-

breajuste. Para isso, existem métricas que verificam a qualidade do ajuste, tais como o Critério de Informação de Akaike (AIC), e o Critério de Informação Bayesiano (BIC).

$$AIC_p = 2p - 2 \log \hat{L}$$

$$BIC_p = (\log n)p - 2 \log \hat{L}$$

Em ambas as métricas, quanto menor o valor resultante, maior é a qualidade do ajuste, com uma penalização baseada no número p de variáveis incluídas no modelo. Para $n \geq 8$, temos $\log n \geq 2$, de forma que o BIC, para esses casos, tem uma penalização maior para o número de variáveis que o AIC, favorecendo modelos mais parcimoniosos (KUTNER et al., 2008).

Para a regressão gaussiana, \hat{L} pode ser calculado através do MS_E (MONTGOMERY et al., 2012). Para a regressão quantílica, essa verossimilhança é calculada na forma $\frac{1}{n} \sum_{i=1}^n \rho_{1/2}(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(1/2))$ (KOENKER, 2005). O AIC e o BIC não têm interpretação direta, e sua aplicação é útil para modelos encaixados. Pelo fato de os modelos quantílico e gaussiano possuírem funções de verossimilhança diferentes, o uso dessas métricas não é recomendado para fazer comparações entre os dois tipos de modelo.

2.3.1 Regressão por *stepwise*

Uma comparação completa testaria todos os modelos possíveis e escolheria, dentre eles, o que otimiza o critério de seleção. Porém, conforme o número de variáveis aumenta, essa comparação se torna computacionalmente inviável, e se faz necessário o uso de algum procedimento de seleção automática. Os principais métodos são:

- **Seleção *forward*:** O modelo inicia sem variáveis preditivas, adicionando uma a cada passo através do critério inicialmente estabelecido, omitindo e removendo casos onde a variável adicionada não foi significativa.
- **Eliminação *backward*:** O modelo inicia com todas as potenciais variáveis preditoras e, a cada passo, remove a variável menos relevante segundo o critério estabelecido.

Dessa forma, nem todos os modelos possíveis são testados, mas o modelo resultante do procedimento é um modelo que satisfaz os critérios (KUTNER et al., 2008). Os dois procedimentos podem ser combinados, resultando no *stepwise* bidirecional, onde, a

cada passo, é possível adicionar e remover variáveis. No software R, a biblioteca MASS disponibiliza as três opções de seleção através do método `stepAIC`, bastando definir o argumento `direction` como "forward", "backward" ou "both".

3 Metodologia

3.1 Conjunto de dados

A Base Nacional de Dados do Poder Judiciário (DataJud) é uma fonte de dados primária que centraliza todos os dados referentes a processos judiciais, englobando tanto processos públicos quanto processos sigilosos, bem como físicos e eletrônicos (CNJ, 2020a). Os dados utilizados neste trabalho são todos oriundos do DataJud, dentre os quais a principal fonte será a Tabela Fato, que alimenta o Painel de Estatísticas do Poder Judiciário (CNJ, 2023).

Embora o DataJud contenha dados de processos sigilosos, a Tabela Fato agrega esses dados, removendo o nível de processo, de modo que nenhum processo possa ser identificado de forma individual. Esses dados são atualizados constantemente, e disponibilizados publicamente pelo Conselho Nacional de Justiça.

A Tabela Fato utilizada compreende desde o início de 2021 até janeiro de 2024. Todas as análises de evolução de indicadores ao longo do tempo considerarão este período. Todavia, quaisquer avaliações que não envolvam evolução ao longo do tempo considerarão apenas o mês mais recente disponível, que representa as condições do Poder Judiciário no momento em que os dados foram obtidos.

Com o período completo, a tabela para a Justiça do Trabalho possui 566.196 linhas. Filtrando o período para a data mais recente, esse número cai para 14.045 linhas.

Os níveis de agregação da Tabela Fato são:

- Grau de Jurisdição: Grau em que o processo está ou esteve em tramitação;
- Formato: Se físico ou eletrônico;
- Procedimento:
 - Conhecimento (criminal ou não criminal), onde se coleta e fornece provas ao juiz responsável;
 - Execução (judicial, fiscal, extrajudicial não fiscal e penal), onde se cumpre uma decisão judicial;
- Se originário ou recursal;

- Data: Ano e mês em que a situação do processo referente ao indicador em análise estava em tramitação.

São disponíveis 26 indicadores para cada um desses níveis de agregação, que mostram:

- o quantitativo de processos em determinada situação naquele período;
- as datas referentes a processos (por exemplo, a menor e a maior data dentre os 5% de processos mais antigos ainda em tramitação);
- o tempo compreendido entre uma situação processual e outra.

Algumas dessas métricas serão usadas para a construção do modelo preditivo. As variáveis explicativas serão as características da vara de justiça e dos processos, tais como grau de jurisdição, quantidade de processos eletrônicos e a quantidade de processos pendentes. A variável resposta, por sua vez, será o indicador agregado de número 16, que representa o tempo médio de todos os processos de cada vara na Justiça do Trabalho, registrado entre o início do processo e a baixa, em dias. Esse tempo é representado como uma média devido à natureza agregada dos dados. Ao longo do trabalho, essa variável será referenciada como "tempo até a baixa".

4 Resultados

4.1 Análise exploratória

Nesta seção, serão feitas investigações de associações unidimensionais entre as covariáveis disponíveis e o tempo até a baixa, bem como as possíveis associações entre as covariáveis. Para isso, serão utilizadas ferramentas gráficas, tabelas e testes de associação.

4.1.1 Quantidade de Órgãos Julgadores

A Tabela 1 exibe a quantidade de órgãos julgadores por tribunal.

Tabela 1: Quantidade de órgãos julgadores por tribunal

| Tribunal | Estado | Frequência | |
|--------------|-----------------|------------|----------|
| | | Absoluta | Relativa |
| TRT2 | SP | 319 | 14,23% |
| TRT1 | RJ | 222 | 9,91% |
| TRT15 | SP | 218 | 9,73% |
| TRT3 | MG | 210 | 9,37% |
| TRT4 | RS | 190 | 8,48% |
| TRT9 | PR | 140 | 6,25% |
| TRT5 | BA | 111 | 4,95% |
| TRT6 | PE | 97 | 4,33% |
| TRT12 | SC | 86 | 3,84% |
| TRT8 | AP e PA | 77 | 3,44% |
| TRT18 | GO | 64 | 2,86% |
| TRT7 | CE | 56 | 2,5% |
| TRT10 | DF e TO | 55 | 2,45% |
| TRT11 | AM e RR | 48 | 2,14% |
| TRT23 | MT | 45 | 2,01% |
| TRT13 | PB | 43 | 1,92% |
| TRT14 | AC e RO | 41 | 1,83% |
| TRT17 | ES | 41 | 1,83% |
| TRT24 | MS | 39 | 1,74% |
| TRT21 | RN | 35 | 1,56% |
| TRT19 | AL | 31 | 1,38% |
| TRT16 | MA | 26 | 1,16% |
| TRT22 | PI | 24 | 1,07% |
| TRT20 | SE | 23 | 1,03% |
| Total | Nacional | 2.241 | 100% |

Dos tribunais trabalhistas, o que conta com maior representação em quantidade de órgãos julgadores é o TRT2, com pouco mais de 13% do total nacional. Dois dos

tribunais trabalhistas com mais órgãos julgadores são pertencentes ao estado de São Paulo, correspondendo, juntos, a quase 25% do total nacional. O estado de São Paulo é, também, a única unidade federativa com dois tribunais da Justiça do Trabalho.

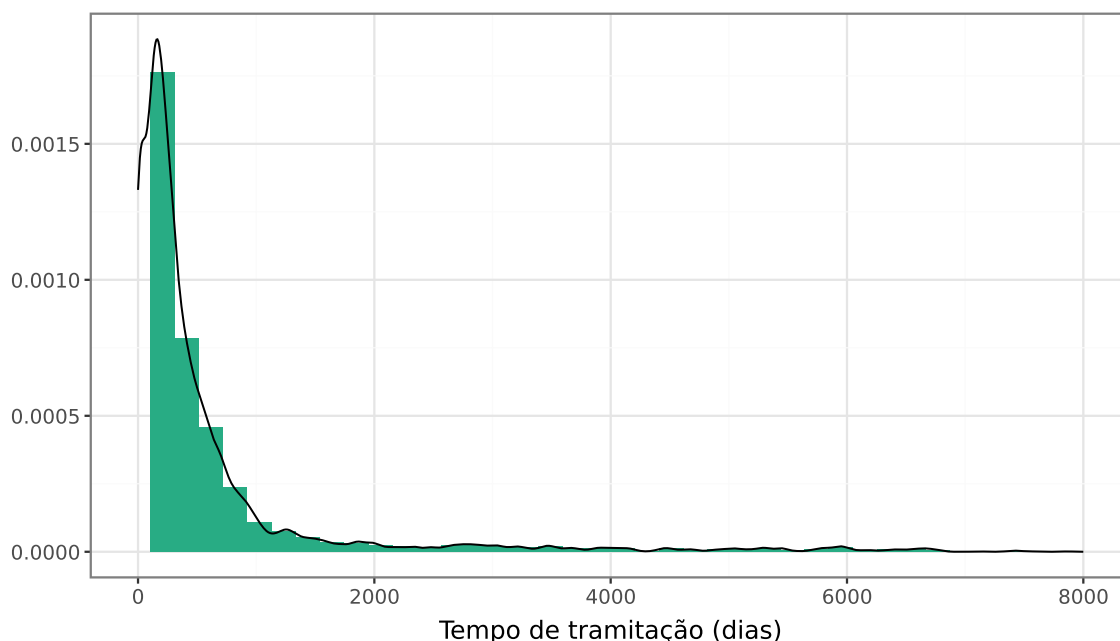
Há quatro tribunais da Justiça do Trabalho que possuem jurisdição em mais de uma Unidade da Federação (UF): TRT8, TRT10, TRT11 e TRT14. Todos esses se restringem a operar em duas UFs cada.

A diferença entre os tribunais para cada ranking decresce de forma suave, sem nenhum grande salto entre um colocado e seu posterior ou anterior. O ramo trabalhista conta com um total de 2.241 órgãos julgadores.

4.1.2 Distribuição dos tempos

A Figura 2 mostra a distribuição dos tempos até a baixa em dias transcorridos.

Figura 2: Distribuição dos tempos entre o início do processo e a baixa (em dias transcorridos)



Todos os valores observados são não-negativos, conforme esperado pela natureza da variável. A maior parte dos processos leva até 2.000 dias para receber baixa, e a distribuição dos tempos é assimétrica à direita. Esse comportamento, por sua vez, não se assemelha ao comportamento de uma distribuição normal, que é simétrica e possui valores negativos com probabilidade não-nula.

Devido à quantidade excessiva de observações, foram feitas 10 amostras aleatórias

simples de tamanho 500, sem reposição, e aplicado, para cada, o teste de normalidade de Shapiro-Wilk sob nível de significância de 5%. As hipóteses foram:

$$\begin{cases} H_0 : \text{O tempo até a baixa segue distribuição Normal.} \\ H_1 : \text{O tempo até a baixa não segue distribuição Normal.} \end{cases}$$

Tabela 2: Teste de normalidade de Shapiro-Wilk para amostras do tempo até a baixa

| Teste | Variáveis | P-valor máximo | Decisão do teste |
|--------------|-------------------|----------------|------------------|
| Shapiro-Wilk | Tempo até a baixa | ≈ 0 | Rejeita H_0 |

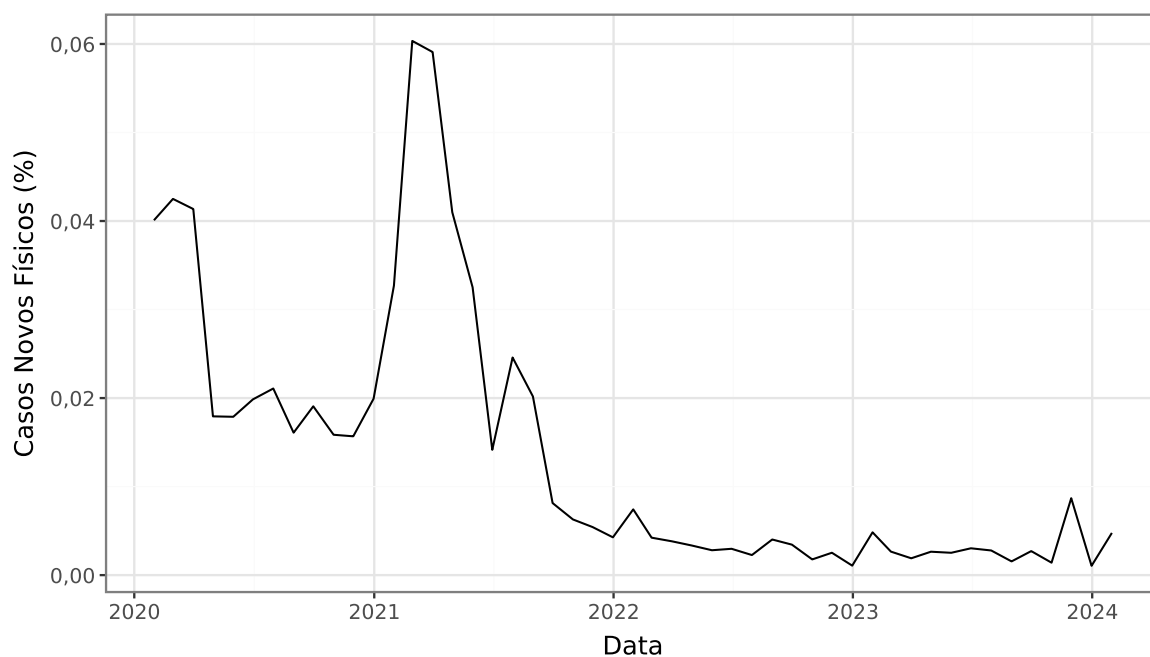
A Tabela 2 mostra que mesmo o maior p-valor obtido ainda foi próximo de zero. Para todos os testes, a hipótese de normalidade do tempo até a baixa de um processo foi rejeitada, apresentando, assim, evidências de que sua distribuição não é normal.

4.1.3 Formato dos processos

O Conselho Nacional de Justiça instituiu o Processo Judicial Eletrônico (PJe) como forma de tornar o Poder Judiciário mais célere (CNJ, 2013). A comparação entre os formatos de processos é relevante, não só pela predição do tempo até a baixa, mas também para observar se os resultados da instituição do PJe correspondem aos esperados pelas instituições de justiça.

Espera-se que a quantidade de casos novos físicos diminua progressivamente ao longo do tempo, até se aproximar de zero. A Figura 3 mostra essa evolução de janeiro 2021 até janeiro de 2024.

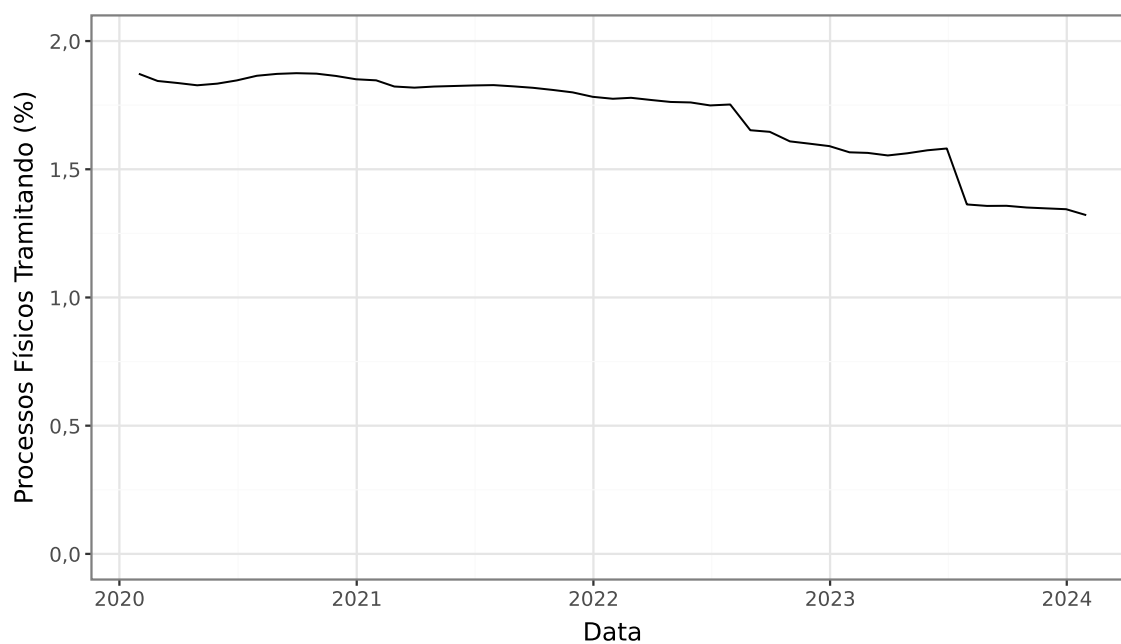
Figura 3: Evolução da frequência relativa de casos novos físicos ao longo do tempo



A Figura 3 mostra uma tendência geral de queda. A frequência relativa de casos novos físicos já era de apenas 0,06% em 2021, e seguiu em queda, chegando a valores muito próximos de zero no começo de 2024.

A Figura 4 mostra a evolução da frequência de processos físicos pendentes tramitando ao longo dos últimos anos.

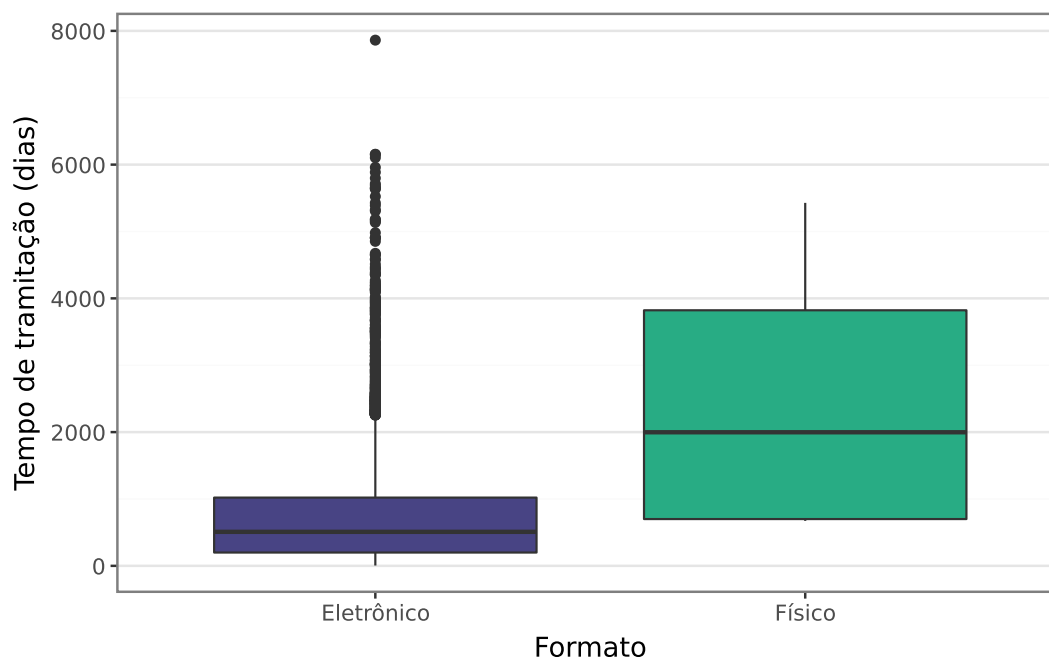
Figura 4: Evolução da frequência relativa de processos pendentes tramitando em formato físico ao longo do tempo



A frequência de processos físicos tramitando permanece relativamente estável e acima de 1%, apesar de o número de casos novos físicos cair em ritmo acelerado. A discrepância mostra que a maior parte destes processos físicos pendentes não foi iniciada durante o período, sugerindo que os processos com formato físico podem chegar a durar vários anos. A baixa frequência de processos físicos em tramitação também indica que essa variável se trata de um fenômeno raro, em especial durante o período.

A Figura 5 mostra a distribuição dos tempos entre o início do processo e a data de referência para cada um dos formatos, onde há processos ainda ativos. Foram omitidos quaisquer processos para os quais o formato era indisponível.

Figura 5: Distribuição dos tempos de tramitação dos processos para cada formato em processos ainda ativos



A Figura 5 mostra uma discrepância grande de magnitude entre os tempos de processos físicos e eletrônicos. Cerca de 50% dos processos físicos tramitando possuem tempo de tramitação próxima ou maior que seis anos.

Além da clara diferença no tempo entre os formatos, há uma variabilidade muito menor dos processos eletrônicos em relação aos processos físicos, que indica que os processos eletrônicos podem ter predições mais precisas.

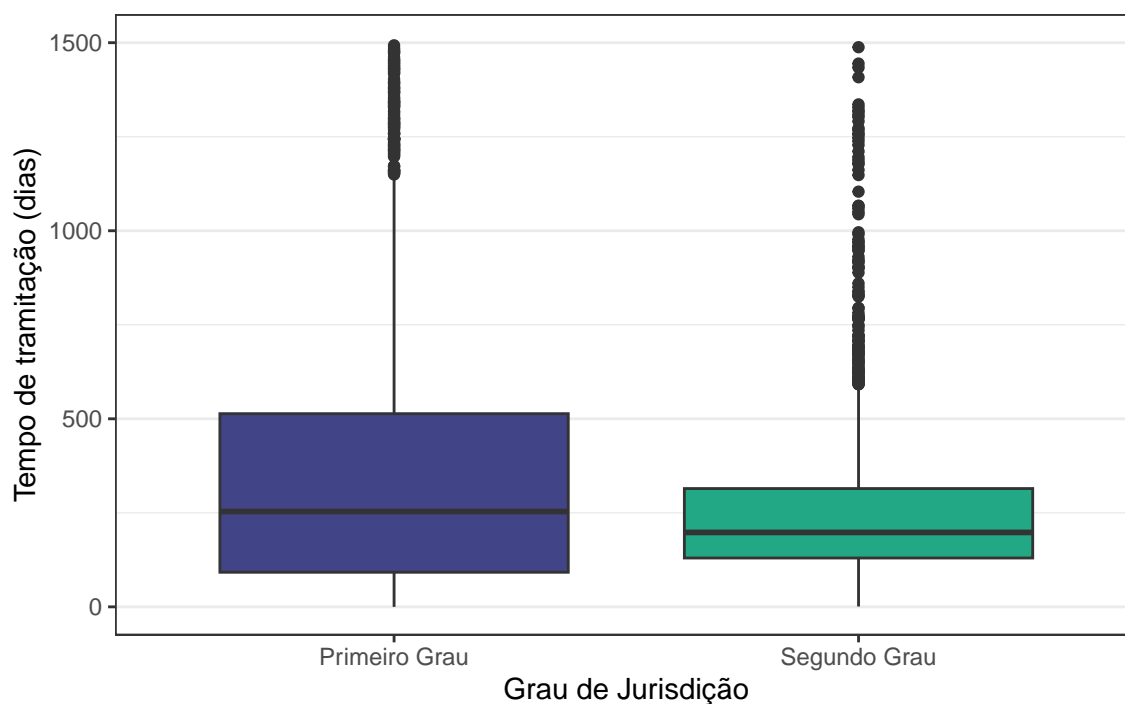
A diferença sugere que o PJe, de fato, tornou os processos mais céleres, conforme as expectativas do Conselho Nacional de Justiça.

4.1.4 Graus de jurisdição

O Poder Judiciário é hierarquizado por três graus de jurisdição: o Primeiro Grau, o Segundo Grau e os Tribunais Superiores. O Primeiro Grau, por sua vez, é também o grau com maior carga processual, fato que motivou a instauração da Política Nacional de Priorização do Primeiro Grau, equilibrando orçamento e pessoal entre os graus segundo suas demandas (CNJ, 2022).

A Figura 6 mostra a distribuição dos tempos até a baixa para o primeiro e segundo graus de jurisdição. Foi feito um corte no eixo Y para evitar que os outliers prejudiquem a visualização.

Figura 6: Distribuição dos tempos até a baixa dos processos para o primeiro e o segundo graus de jurisdição

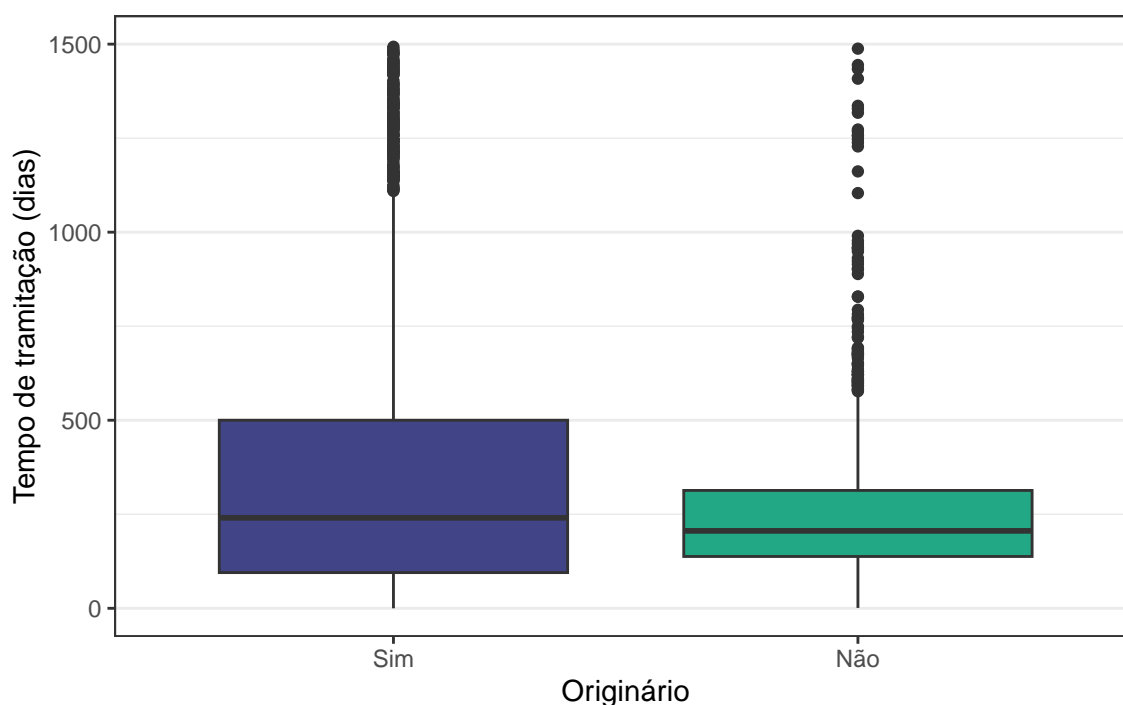


Os tempos até a baixa de um processo são consistentemente menores para o segundo grau, quando comparados com o primeiro. De fato, a discrepância que o Relatório Justiça em Números aponta entre o Primeiro Grau e os outros graus de jurisdição ainda é refletida nos indicadores, tornando o grau uma variável candidata relevante para a predição dos tempos.

4.1.5 Recursos

A primeira decisão judicial sobre um processo é tomada em sua competência originária. Em caso de impugnação de alguma das partes envolvidas no processo, é aberto um recurso para revisar a decisão. A Figura 7 mostra a distribuição dos tempos para processos originários e recursais. O eixo Y foi cortado para evitar prejuízo à visualização devido aos outliers.

Figura 7: Distribuição dos tempos até a baixa dos processos originários e recursais



A Figura 7 mostra uma diferença de dispersão e de magnitude, onde não só os recursos são mais céleres, como também são concentrados em um intervalo menor.

Ainda, o comportamento dos processos originários e recursais é muito parecido com o comportamento dos processos de primeiro e segundo grau, o que pode indicar associação entre as variáveis. Essa possível associação será discutida na Seção 4.1.6.

4.1.6 Associação entre as variáveis qualitativas

Foi observada uma semelhança entre os tempos até a baixa nos gráficos da Seção 4.1.5 e a Seção 4.1.4. A Tabela 3 investiga as frequências cruzadas de ocorrências de cada grau e nível de recurso.

Tabela 3: Frequência de processos pendentes cruzada por grau e recursos

| Grau | Originário | | Total |
|-------|------------|---------|-----------|
| | Sim | Não | |
| 1º | 5.284.459 | 0 | 5.284.459 |
| 2º | 113.080 | 670.461 | 783.541 |
| Total | 5.397.539 | 670.461 | 6.068.000 |

Olhando através dos níveis de recurso, é perceptível que a maior parte das ocorrências originários está compreendida em primeiro grau, bem como a maior parte das ocorrências não-originárias está compreendida em segundo grau. Dentre os processos de segundo grau, há cerca de seis vezes mais processos recursais do que processos originários, e nenhum dos processos recursais observados no período pertence ao primeiro grau.

Visando investigar possíveis outras associações, serão aplicados testes de associação entre as variáveis qualitativas. A Tabela 4 exibe o resultado dos testes bivariados de independência para procedimento, grau, originário e formato.

O teste exato de Fisher foi aplicado em todas as tabelas 2x2. Para as tabelas com dimensão maior, foi utilizado o teste Qui-Quadrado de independência.

Tabela 4: Testes para Independência entre Variáveis Categóricas

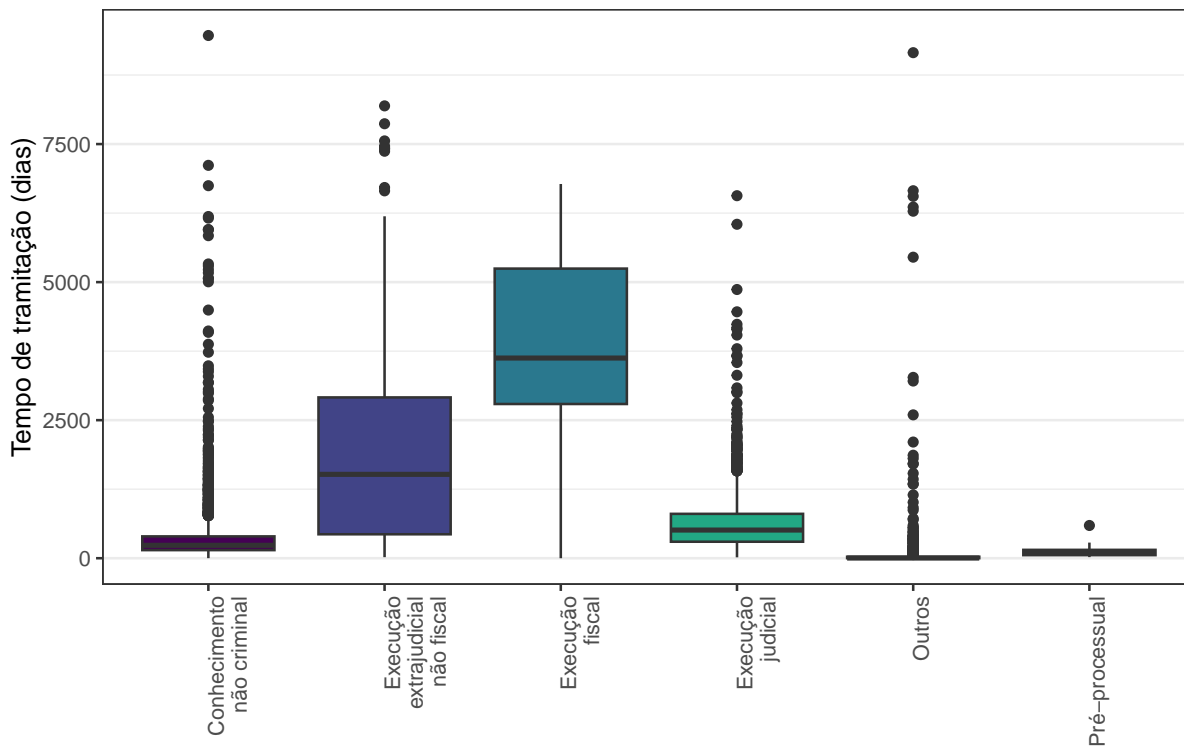
| Teste | Variáveis | P-valor | Decisão do teste |
|-----------------|---------------------------|-------------|-------------------|
| Qui-Quadrado | Grau x Procedimento | ≈ 0 | Rejeita H_0 |
| Qui-Quadrado | Procedimento x Originário | ≈ 0 | Rejeita H_0 |
| Qui-Quadrado | Procedimento x Formato | 0,628 | Não rejeita H_0 |
| Exato de Fisher | Formato x Grau | 0,307 | Não rejeita H_0 |
| Exato de Fisher | Formato x Originário | 0,585 | Não rejeita H_0 |
| Exato de Fisher | Grau x Originário | ≈ 0 | Rejeita H_0 |

Não há evidências de que a informação sobre o procedimento esteja associada ao processo ser físico ou eletrônico. Apesar disso, a hipótese de independência foi rejeitada entre procedimento e grau, bem como em procedimento e originário. O teste também confirma o que era apontado pela Tabela 3, mostrando evidências de associação entre grau e originário. Para as outras variáveis, não foram encontradas evidências significativas de associação.

4.1.7 Procedimentos

A Figura 8 mostra a distribuição dos tempos médios até a baixa para cada tipo de procedimento.

Figura 8: Distribuição dos tempos até a baixa dos processos segundo o procedimento



É notável que os procedimentos de conhecimento em geral apresentam tempos drasticamente menores que os procedimentos de execução. Os pré-processuais e outros apresentaram as menores tendências de tempo entre todos os procedimentos observados.

Além da diferença quantitativa entre as execuções, cada execução apresentou um padrão diferente de dispersão. A execução judicial foi a que menos variou entre os três procedimentos de execução. A execução fiscal não apresentou nenhum outlier e teve o maior valor mediano dentre todos os procedimentos, além de apresentar comportamento assimétrico.

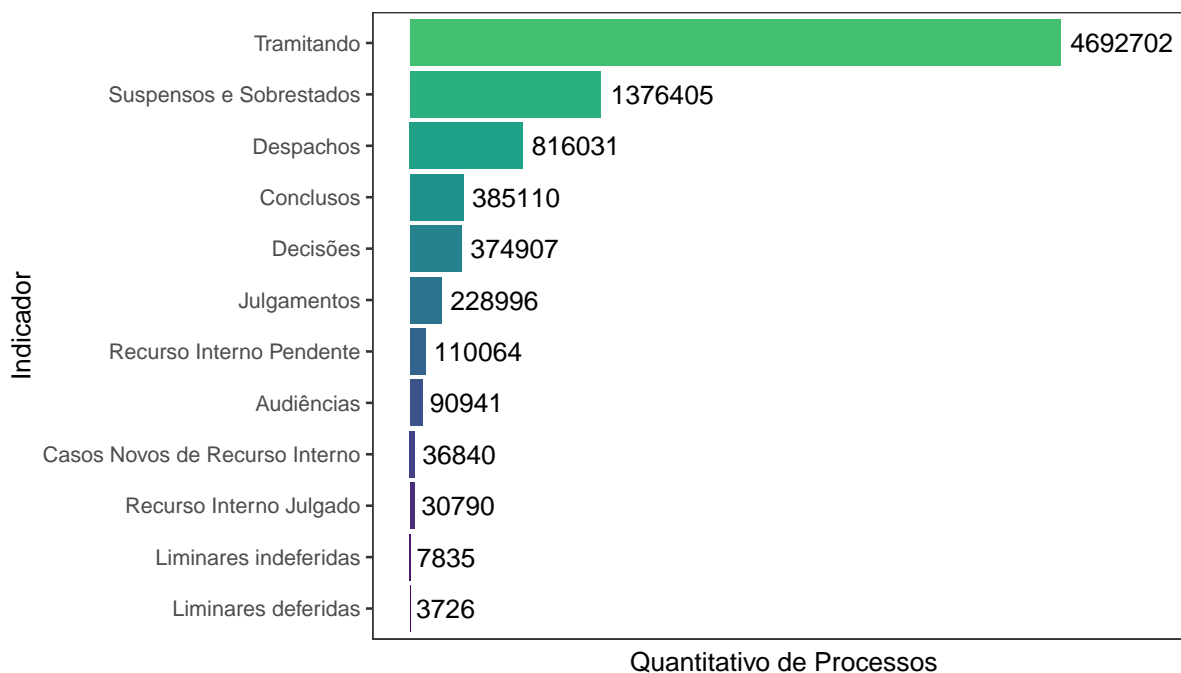
4.1.8 Indicadores quantitativos

Dos 26 indicadores presentes na Tabela Fato, alguns são correlacionados com outras métricas de forma direta (CNJ, 2023). Estão entre eles:

- Total de processos Concluídos para o Magistrado a mais de 50 dias: Está correlacionado com o indicador de Concluídos para o Magistrado.
- Processos sem tramitação há mais de 50 dias: Está correlacionado ao indicador de Tramitando.
- 5% mais antigos em tramitação: Corresponde a 5% dos processos no indicador de Tramitando.
- Tramitando (ou Pendentes líquidos): Agrega todos os processos que não receberam Baixa, Suspensão ou Sobrestamento.
- Pendentes: Agregam todos os processos que não receberam Baixa, incluindo Tramitando, Suspensos e Sobrestados.
- Concluídos: Agrega Concluídos para Julgamento, para Despacho, para Admissibilidade recursal, para Decisão, sem especificação e outros.
- Audiências: Agrega Audiências Conciliatórias e Não Conciliatórias.

Em decorrência das correlações, serão escolhidas as métricas que, sozinhas, agregam a maior quantidade de outros indicadores. A exceção a essa regra será o indicador da situação Tramitando, bem como Suspensos e Sobrestados (que, juntos, constituiriam os Pendentes). Assim sendo, os indicadores escolhidos são: Tramitando, Suspensos e Sobrestados, Concluídos, Julgamentos, Despachos, Decisões, Audiências, Total de liminares deferidas, Total de liminares indeferidas, Casos Novos de Recurso Interno, Recurso Interno Pendente e Recurso Interno Julgado. A Figura 9 mostra a quantidade de ocorrências de cada um dos indicadores selecionados. Apesar de o indicador Pendentes não ter sido incluído, a informação completa dele está incluída, uma vez que os Pendentes são a soma de Tramitando, Suspensos e Sobrestados. Os processos Suspensos e Sobrestados, por simplicidade, serão referidos no documento apenas como “Suspensos”.

Figura 9: Quantidade de processos em cada métrica

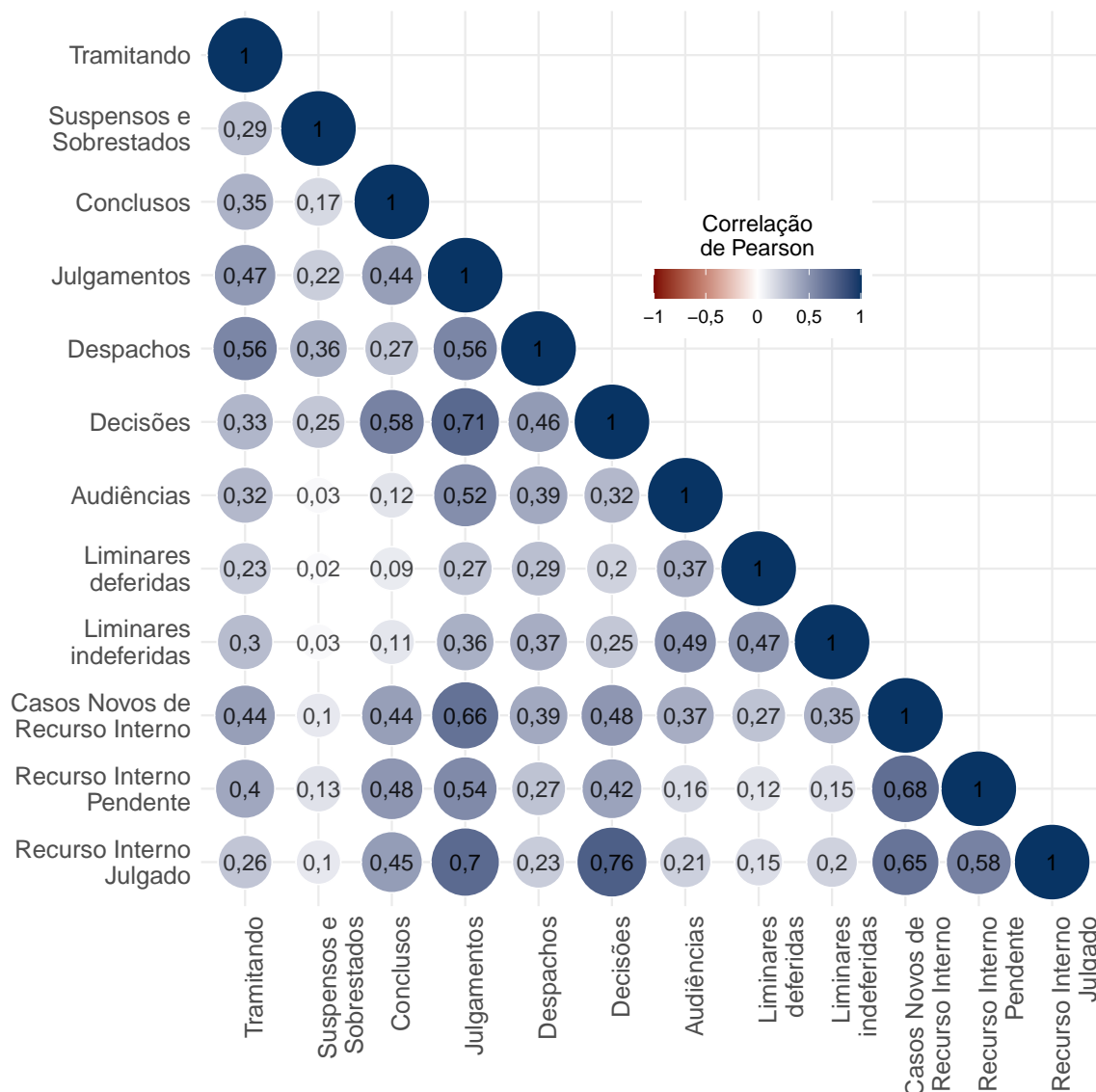


A quantidade de processos Tramitando é a predominante dentre todos os indicadores exibidos. Nos dois indicadores mais frequentes, há um salto entre o primeiro e o segundo, onde o quantitativo de processos Tramitando é mais de três vezes superior ao número de Suspensos.

Os demais indicadores decaem com poucos saltos, onde os indicadores com menor representação são os de liminares indeferidas e liminares deferidas. Estes, por sua vez, têm uma quantidade de processos pouco maior do que a quantidade de órgãos julgadores. Nota-se que a quantidade de liminares deferidas tem uma média de pouco mais de 1,5 liminares para cada órgão julgador.

A Figura 10 investiga possíveis correlações entre variáveis explicativas quantitativas.

Figura 10: Matriz de correlações entre as variáveis explicativas quantitativas



Não houve nenhuma correlação negativa dentre todas as observadas na matriz, embora haja diversas correlações muito próximas de zero.

Os Casos Novos de Recurso Interno, Recurso Interno Pendente e Recurso Interno Julgado, apesar de não conterem um ao outro (como é o caso de outras variáveis supracitadas), não só se correlacionam pelo fato de tratarem de fenômenos relacionados aos recursos internos, como apresentam uma correlação de moderada a forte (aproximadamente de 0,6), com intensidade próxima para todos eles. O indicador Recurso Interno Julgado apresenta uma correlação ainda mais forte com as variáveis Decisões e Julgamentos, que configuraram, também, as maiores correlações observadas na matriz.

As Decisões e Julgamentos, por sua vez, também estão correlacionadas entre elas com intensidade próxima à de Recursos Internos Julgados. Houve uma intensidade

moderada entre os dois indicadores e o quantitativo de Conclusos.

Os dois indicadores de Liminares (deferidas e indeferidas), Audiências, Despachos, Conclusos e Suspensos, não mostraram correlações fortes com nenhum indicador. Os maiores valores registrados para a correlação das liminares indeferidas foram 0,49 e 0,47, com, respectivamente, as variáveis de Audiências e Liminares deferidas. Os outros indicadores citados, eventualmente, apresentaram correlações moderadas entre eles, onde os Conclusos tiveram a maior de suas correlações com as Decisões e os Despachos tiveram suas maiores correlações com Tramitando e Julgamentos.

O indicador Tramitando, mesmo com maior representatividade dentre todos os indicadores, não apresentou correlação forte com nenhum deles, tendo variado entre fraca e moderada.

Os Suspensos foram os casos que apresentaram menores correlações com todos os indicadores, com valores maiores que 0,3 apenas quando correlacionados com os Despachos.

4.1.9 Tempo até a baixa por indicador

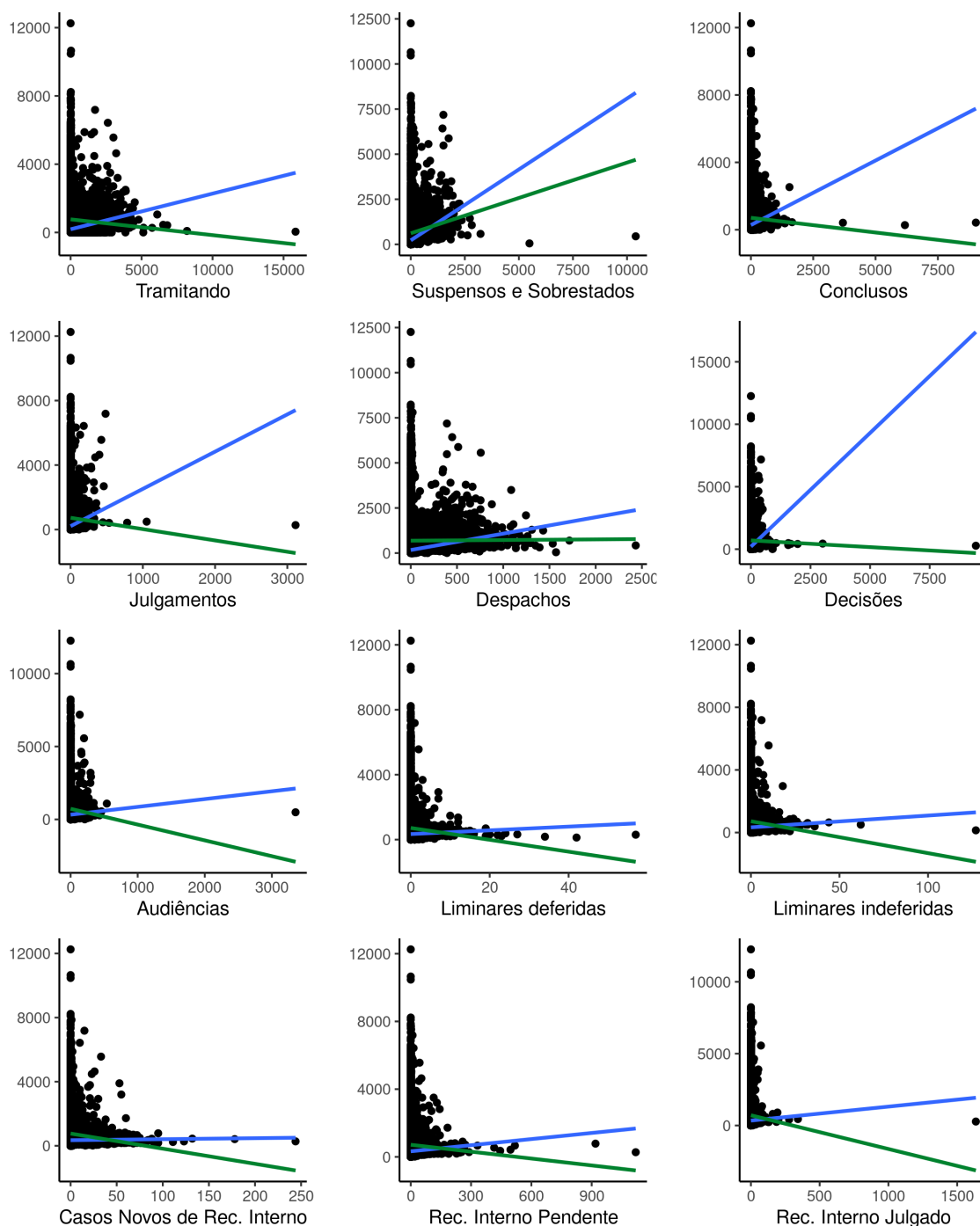
A Figura 11 mostra gráficos de dispersão de cada variável explicativa quantitativa contra o tempo até a baixa dos processos. A reta azul indica a reta de regressão quantílica para a mediana, enquanto a reta verde indica a reta de regressão gaussiana para a média.

Na visualização bidimensional, não há, segundo os modelos quantílicos, qualquer indicativo de tendência de decrescimento para o tempo mediano de tramitação dos processos conforme o crescimento de um indicador, de modo que esperam-se coeficientes positivos em quaisquer lugares onde eles sejam não-nulos. Apesar disso, vários dos modelos de regressão por mínimos quadrados apresentaram retas decrescentes, eventualmente podendo levar a valores inferiores a zero, sendo eles: Conclusos, Julgamentos, Decisões, Audiências, Liminares deferidas, Liminares indeferidas, Casos Novos de Rec. Interno, Rec. Interno Pendente e Rec. Interno Julgado, de modo que não só a reta mediana e a reta média não coincidem, como se contradizem. Nota-se que a regressão gaussiana é mais suscetível à influência de outliers do que a regressão quantílica.

Apesar de, graficamente, os Casos Novos de Recurso Interno aparentarem ter uma menor inclinação, as inclinações estão prejudicadas não só pela escala do eixo X e Y serem diferentes, como também pela presença de outliers nas variáveis explicativas.

Em todos os indicadores, existe uma concentração grande das observações em

Figura 11: Gráficos de dispersão entre as variáveis explicativas e o tempo até a baixa com retas de regressão quantílica (mediana, em azul) e regressão gaussiana (média, em verde)



torno do valores onde a variável explicativa é próxima de zero. Fica notável a presença de vários órgãos julgadores que, apesar de relatarem poucos processos tramitando, levam um tempo significativamente grande para dar baixa nos processos que existem. A Tabela 5 exhibe características das varas com menos de 120 processos pendentes e tempo até a baixa maior que 1.000 dias.

A Tabela 5 mostra que, das varas com menos de 120 processos tramitando e

Tabela 5: Características das varas com menos de 120 processos tramitando e mais de 1.000 dias de tempo até a baixa

| Formato | Grau | Originário | Procedimento | Frequência | Tempo |
|------------|------|------------|--------------------------------|------------|-------|
| Eletrônico | 1º | Sim | Exec. fiscal | 267 | 4299 |
| Eletrônico | 1º | Sim | Exec. extrajudicial não fiscal | 69 | 3978 |
| Eletrônico | 2º | Não | Conhec. não criminal | 25 | 1830 |
| Eletrônico | 2º | Sim | Conhec. não criminal | 22 | 1411 |
| Eletrônico | 1º | Sim | Exec. judicial | 3 | 3220 |
| Eletrônico | 1º | Sim | Conhec. não criminal | 1 | 3295 |
| Físico | 1º | Sim | Conhec. não criminal | 1 | 9466 |
| Físico | 1º | Sim | Exec. judicial | 1 | 3085 |

com tempo até a baixa acima de 1.000 dias, 336 (86,37%) delas estão entre os processos eletrônicos de primeiro grau, originários, com procedimento de execução (fiscal e extrajudicial não fiscal). Esse comportamento está dentro do esperado, uma vez que, como já observado na Figura 8, os dois procedimentos citados possuem tanto maiores tempos até a baixa quanto maior dispersão nesses tempos, bem como os processos originários e de primeiro grau.

Poucas ocorrências relatavam processos físicos, mas, das que relatavam, há uma vara com processos físicos cujo tempo até a baixa foi de quase 10.000 dias, que foi o maior valor registrado na tabela. Essa vara tinha procedimento de conhecimento não criminal, que foi o único procedimento de conhecimento relatado nos grupos observados, e também o procedimento de conhecimento com maiores tempos até a baixa (Figura 8).

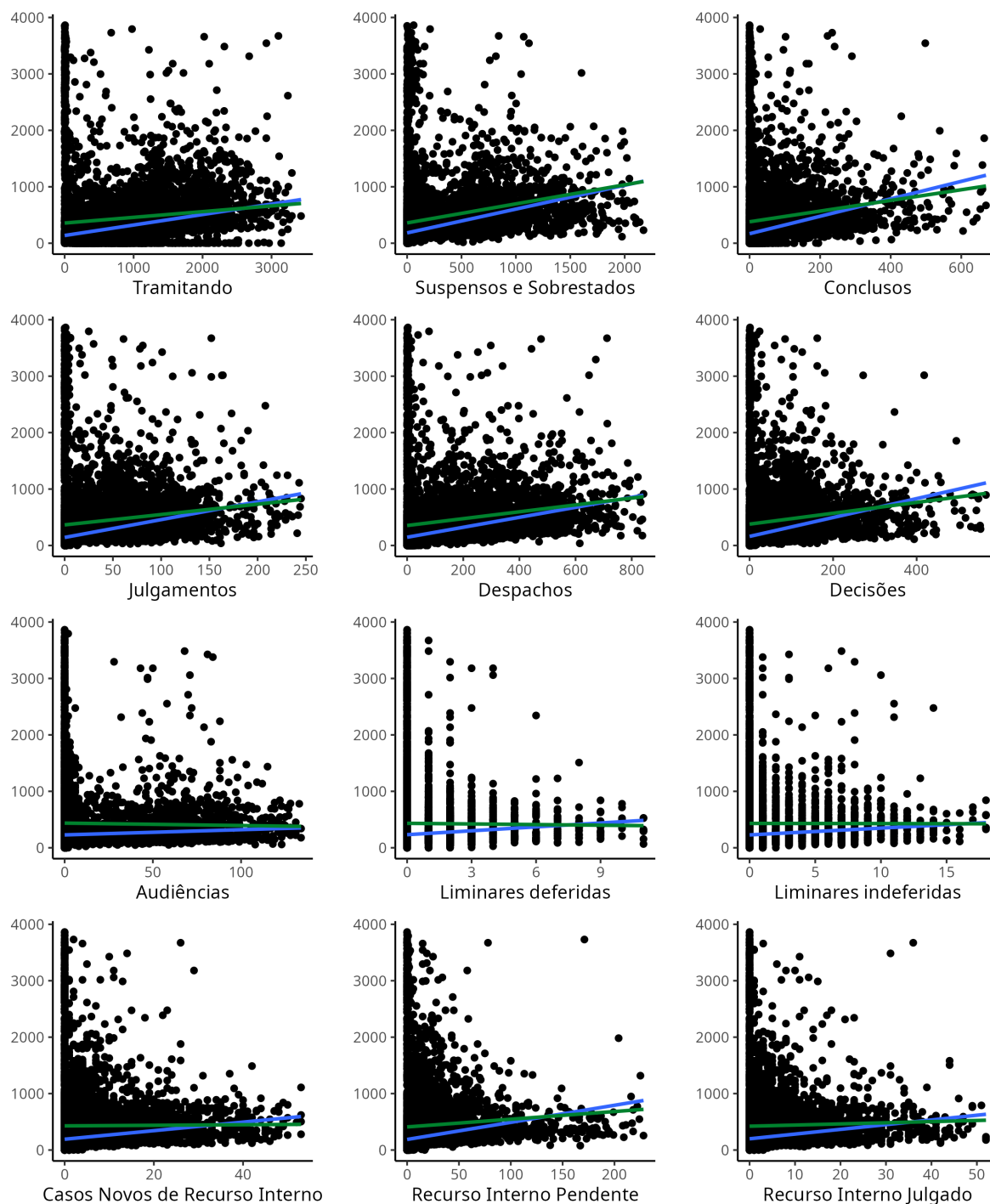
A Figura 12 exibe os mesmos diagramas de dispersão, com a remoção dos outliers das variáveis explicativas e remoção das quatro categorias da Tabela 5 que apresentaram os maiores tempos até a baixa. A reta azul indica a reta de regressão quantílica para a mediana, enquanto a reta verde indica a reta de regressão gaussiana para a média.

Com a remoção das categorias supracitadas, quase todas as contradições de tendência entre as médias e as medianas sumiram para todas as variáveis explicativas observadas. Ainda, até mesmo a proximidade das retas aumentou, fornecendo valores, visualmente, muito próximos entre elas.

As retas medianas permaneceram estritamente crescentes. Já as retas médias ainda apresentaram um comportamento que aparenta estar entre o decrescente e o constante para as Audiências, Liminares deferidas, Liminares indeferidas e Casos Novos de Rec. Interno.

As audiências, liminares e recursos internos ainda apresentam dispersão maior dos tempos até a baixa para valores pequenos das variáveis explicativas, bem como tempos

Figura 12: Gráficos de dispersão entre as variáveis explicativas e o tempo até a baixa com retas de regressão quantílica (mediana, em azul) e regressão gaussiana (média, em verde), com remoção de outliers



propriamente ditos consideravelmente maiores onde os valores das variáveis explicativas eram próximos de zero. O comportamento delas é não-linear, e elas possuem uma amplitude muito pequena quando comparadas às outras seis variáveis.

4.2 Modelos

Para a construção dos modelos, os dados foram filtrados para conter apenas o retrato do mês de janeiro. Foram utilizadas todas as observações do filtro. A seleção de variáveis foi feita utilizando *stepwise* bidirecional e o AIC como critério de seleção. Uma vez selecionado um modelo com o menor AIC, será feita uma outra seleção de variáveis utilizando *stepwise* pelo teste de significância dos parâmetros, removendo a variável com o maior p-valor e ajustando um modelo sem ela, repetindo o processo até que todas as variáveis tenham um p-valor menor que 0,1.

4.2.1 Modelos para a mediana

O Modelo 1 mostra o resultado do ajuste considerando todas as covariáveis de forma independente, sem quaisquer interações ou componentes não-lineares.

Modelo 1: Regressão quantílica ($\tau = 0,5$)

| Variável | Coefficiente | Erro Padrão | P-valor |
|-------------------------------|--------------|-------------|-------------|
| Intercepto | 7,1 meses | 18 | ≈ 0 |
| Procedimento - Pré-processual | -3,6 meses | 36 | 0,002 |
| Procedimento - Outros | -1 mês | 12 | 0,012 |
| Grau | -2,1 meses | 18 | ≈ 0 |
| Tramitando | 2h29min | 0,007 | ≈ 0 |
| Suspensos | 3h53min | 0,024 | ≈ 0 |
| Despachos | 5h12min | 0,041 | ≈ 0 |
| Decisões | -7h56min | 0,097 | 0,009 |
| Conclusos | 4h4min | 0,076 | 0,026 |

No que se refere às variáveis qualitativas, a variável “Formato” não se mostrou relevante, contrariando as expectativas geradas pela Seção 4.1.3, na qual a variável foi avaliada incondicionalmente e de forma exploratória. Apesar disso, tanto as expectativas da Seção 4.1.4 quanto da Seção 4.1.7 foram confirmadas, dado que o Grau e alguns procedimentos foram considerados relevantes. É notável que os procedimentos resultantes estão entre os procedimentos que apresentaram o menor tempo, e apresentaram coeficientes negativos, portanto.

Todas as variáveis quantitativas apresentaram evidências estatisticamente significativas para rejeição da hipótese de ausência de regressão, conforme o método de seleção

induz. A variável “Decisões” foi a única variável quantitativa cujo coeficiente foi negativo.

Quase todas as covariáveis resultantes apresentam uma correlação fraca entre elas, como se pode observar retornando à Figura 10. Em alguns casos, como a correlação de “Conclusos” com “Suspensos”, ela é quase indistinguível de uma correlação nula. Das 10 duplas de variáveis, 7 apresentaram correlação entre 0,32 e 0,63, com a correlação de “Despachos” com “Suspensos” sendo a maior registrada.

Em função das correlações ainda existentes, será ajustado um modelo onde constam as covariáveis do Modelo 1, suas interações e os componentes quadráticos, utilizando seleção via *stepwise* através do menor AIC e do teste de significância das covariáveis. O resultado dessa seleção consta no Modelo 2.

Modelo 2: Regressão quantílica ($\tau = 0,5$) com interações e componentes quadráticos

| Variável | Coefficiente | Erro Padrão | P-valor |
|-------------------------------|--------------|-------------|-------------|
| Intercepto | 7,4 meses | 15 | ≈ 0 |
| Procedimento - Pré-processual | -4 meses | 4 | ≈ 0 |
| Procedimento - Outros | -1,1 mês | 12 | 0,005 |
| Grau | -2,4 meses | 15 | ≈ 0 |
| Tramitando | 2h30min | 0,006 | ≈ 0 |
| Suspensos | 2h39min | 0,032 | ≈ 0 |
| Despachos | 2h57min | 0,013 | ≈ 0 |
| Suspensos:Decisões | -20 segundos | 2,3e-05 | ≈ 0 |
| Suspensos:Conclusos | 47 segundos | 1,7e-04 | 0,001 |
| Suspensos:Despachos | 11 segundos | 5,0e-05 | 0,009 |

Nenhuma componente quadrática das covariáveis restou no Modelo 2. Todas as três interações restantes envolveram a variável “Suspensos”, sendo a interação com “Decisões” a única de todas as variáveis quantitativas que exibiu um coeficiente negativo, comportamento semelhante ao obtido no Modelo 1. As covariáveis “Conclusos” e “Decisões”, que apareceram individualmente no Modelo 1, não aparecem no Modelo 2 de forma individual, e só permanecem nele através das interações. Os dois modelos possuem coeficientes muito próximos nas variáveis que têm em comum, e possuem um número de coeficientes diferentes de zero igualmente próximo, onde o Modelo 2 apresentou apenas um coeficiente a mais. Todas as variáveis qualitativas são as mesmas em ambos os modelos.

A Tabela 6 exibe algumas medidas da qualidade do ajuste para cada um dos modelos.

Tabela 6: Qualidade do ajuste para os modelos quantílicos para a mediana, sendo o Modelo 1 (sem interações) e Modelo 2 (com interações)

| Modelo | AIC | BIC | MAE | Perda (ρ) |
|---------------|------------|------------|------------|----------------------------------|
| Modelo 1 | 48.143 | 48.198 | 228 | 385.422 |
| Modelo 2 | 48.104 | 48.165 | 226 | 383.093 |

Apesar de possuir um coeficiente a mais, o Modelo 2 é o que minimiza tanto o AIC quanto o BIC, de modo que a penalização por uma variável a mais não o tornou um modelo pior segundo esse critério. O erro médio absoluto, por sua vez, apresentou uma diferença muito pequena entre os dois modelos.

Apesar da diferença, o fato de os dois modelos serem muito parecidos em quase todos os aspectos torna o Modelo 1 preferível, tanto por parcimônia, quanto pelo fato de não possuir interações, que torna a interpretação mais factível.

O Modelo 2 pode ser enunciado na forma: Para cada processo tramitando a mais, mantendo constantes as demais covariáveis, se espera que o tempo mediano até a baixa no órgão julgador cresça em 2h29min. Em se tratando do tempo mediano, os processos em segundo grau levam 2,1 meses a menos para receber baixa que os processos em primeiro grau, e os processos em procedimento Pré-processual ou Outros levam de 1 a 3,6 meses a menos para receber baixa que todos os outros tipos de procedimento. Um processo suspenso influencia mais no tempo mediano até a baixa que um processo tramitando, e quanto mais decisões forem proferidas, menor o tempo mediano até a baixa, de modo que, a cada decisão, o tempo mediano cai em cerca de 8h.

4.2.2 Modelos de regressão para todos os quantis

Para a construção dos modelos, foram considerados os quantis de nível de 10%, 25%, 50%, 75% e 90%. O mesmo processo de seleção feito na Seção 4.2.1 foi aplicado para os quantis restantes. A Tabela 7 mostra o AIC, o BIC e a perda pinball de cada modelo, seguido do número de coeficientes significativos e a quantidade de coeficientes comuns nos dois modelos. Os modelos marcados com 1 representam as seleções lineares, enquanto os modelos marcados com 2 representam as seleções com interações e termos quadráticos.

Tabela 7: AIC, BIC, Perda e número de coeficientes para os ajustes do Modelo 1 (sem interações) e do Modelo 2 (com interações)

| Quantil | AIC | | BIC | | Perda (ρ) | | Coeficientes | | |
|---------|--------|--------|--------|--------|------------------|---------|--------------|----|--------|
| | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | Comuns |
| 0,10 | 46.744 | 46.674 | 46.824 | 46.772 | 112.668 | 111.410 | 13 | 16 | 10 |
| 0,25 | 46.947 | 46.876 | 47.014 | 46.962 | 242.032 | 239.293 | 11 | 14 | 9 |
| 0,50 | 48.143 | 48.104 | 48.198 | 48.165 | 385.422 | 383.093 | 9 | 10 | 7 |
| 0,75 | 50.584 | 50.524 | 50.639 | 50.573 | 414.883 | 411.335 | 9 | 8 | 6 |
| 0,90 | 53.972 | 53.856 | 54.015 | 53.899 | 329.020 | 323.413 | 7 | 7 | 5 |

O ajuste do terceiro quantil mostrou um comportamento diferente dos outros quantis: O modelo com interações e transformações quadráticas apresentou menos coeficientes significativos que o modelo sem essas transformações, assim como um AIC, BIC e perda menores. Todos os outros modelos tiveram um número menor ou igual de coeficientes significativos nos modelos sem transformações.

Os modelos para o quantil de 0,9 tiveram menos coeficientes que todos os outros (7 nos dois casos), e compartilham muitas variáveis em comum, embora não sejam todas. De modo geral, todos os modelos apresentaram um número relativamente grande de coeficientes significativos comuns.

A Tabela 8 mostra quais coeficientes estão presentes em cada modelo candidato, sendo os marcados com 1 os candidatos lineares e os marcados com 2 os candidatos com termos quadráticos e interações.

O resultado mostra que a regressão quantílica não chega sempre no mesmo modelo, de forma que as variáveis relevantes podem variar de quantil para quantil. Há três variáveis dos dados originais que foram consistentemente presentes em todos os modelos: Procedimento - Pré-processual, Grau, Tramitando. Nenhuma transformação quadrática das variáveis permaneceu em qualquer dos modelos, não sendo estatisticamente significativa em nenhum quantil.

Apesar de o Formato não ter aparecido na Seção 4.2.1 (regressão mediana), ele aparece na maior parte dos modelos, destacando a presença dele nas extremidades, dos quantis 0,1, 0,75 e 0,9. Uma influência maior do Formato nos quantis mais altos está dentro das expectativas para essa variável, uma vez que, como mostra a Figura 5, o formato “Físico” concentra tempos consistentemente maiores que o formato “Eletrônico”.

Os modelos para o quantil 0,9 apresentaram uma diferença sutil entre eles. Enquanto o modelo sem interações considera as variáveis Conclusos e Suspensos individualmente, o modelo com interações não considera nenhuma delas de forma individual, mas a interação entre as duas (interação essa que se mostrou significativa em quatro dos cinco modelos). Um comportamento parecido foi observado no quantil de 0,75.

Todos os modelos são muito parecidos com suas contrapartes de mesmo quantil, mas o fato de possuírem muitas interações a mais prejudica a interpretabilidade dos modelos, e métodos como o AIC não penalizam a interpretabilidade.

Devido ao fato de possuírem um número igual ou menor de coeficientes em relação aos modelos lineares e valores menores mais expressivos para as medidas de qualidade do ajuste (AIC, BIC e perda pinball), os modelos com transformações serão escolhidos para os quantis 0,7 e 0,9, apesar de comprometerem ligeiramente a interpretação.

Em relação aos modelos de quantis 0,1 e 0,25, os modelos com interações perdem tanto a interpretabilidade quanto a simplicidade, sendo os modelos que apresentaram maiores diferenças no número de coeficientes e menores ganhos de AIC, BIC e perda pinball. Neste caso, os modelos sem interações serão os preferidos.

A Tabela 9 mostra os coeficientes destes modelos.

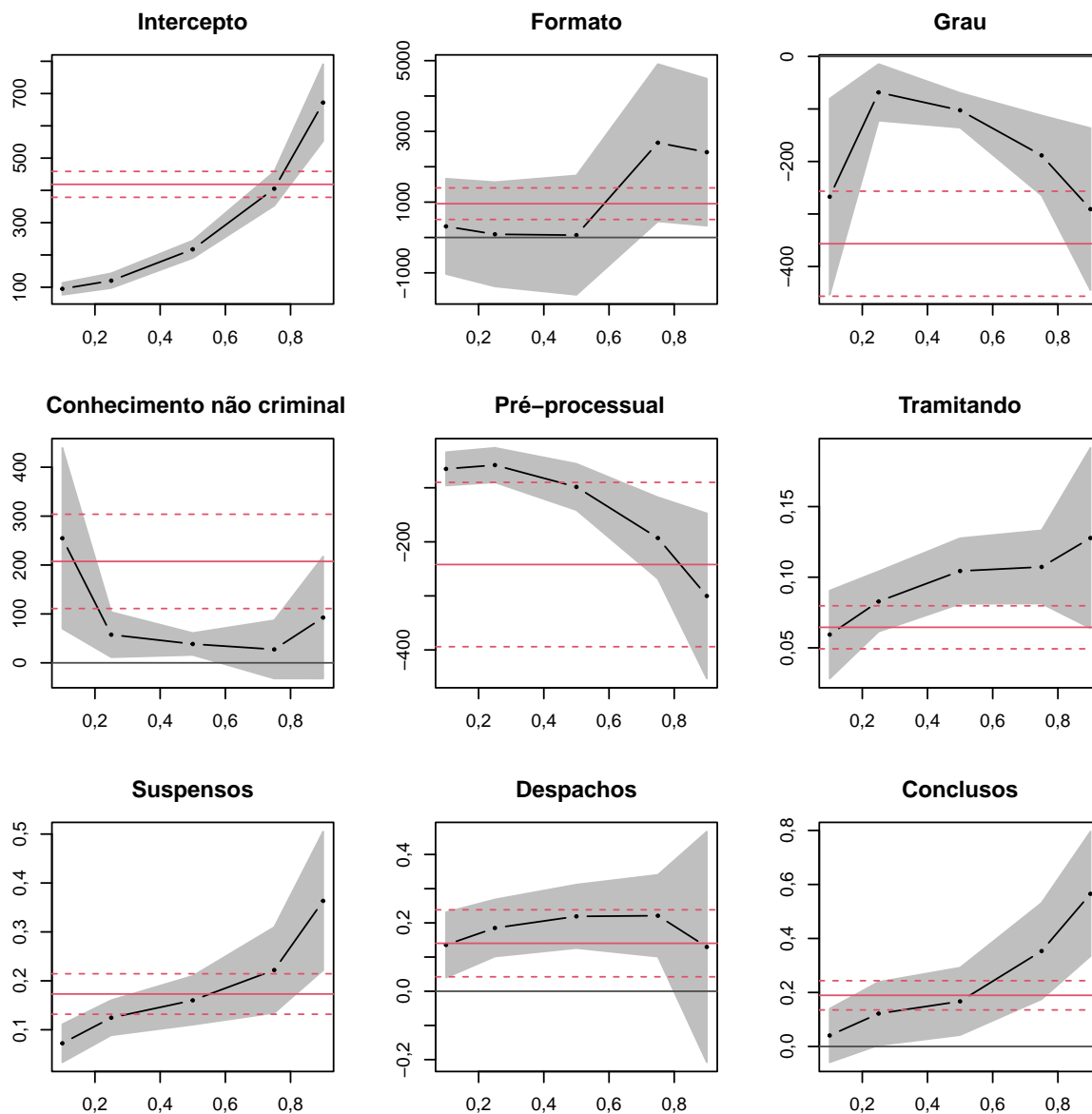
Tabela 9: Coeficientes dos modelos selecionados para os quantis $\tau=0,1$, $\tau=0,25$, $\tau=0,5$, $\tau=0,75$, $\tau=0,9$

| Variável | Quantil | | | | |
|-------------------------------------|---------|--------|--------|--------------------|--------------------|
| | 0,1 | 0,25 | 0,5 | 0,75 | 0,9 |
| Intercepto | 99 | 125 | 219 | 437 | 769 |
| Grau | -98 | -66 | -65 | -179 | -261 |
| Formato | 141 | | | 2648 | 2315 |
| Procedimento - Conhec. não criminal | 78 | 46 | | | |
| Procedimento - Pré-processual | -69 | -63 | -113 | -228 | -424 |
| Procedimento - Outros | | | -31 | | |
| Tramitando | 0,031 | 0,067 | 0,104 | 0,106 | 0,058 |
| Suspensos | 0,07 | 0,132 | 0,162 | 0,135 | |
| Despachos | 0,158 | 0,183 | 0,217 | | |
| Decisões | -0,124 | -0,242 | -0,253 | | |
| Conclusos | | 0,113 | 0,169 | | |
| Rec. Interno Julgado | 0,996 | 0,62 | | | |
| Liminares indeferidas | -11 | | | | |
| Julgamentos | 0,085 | | | | |
| Rec. Interno Pendente | 0,473 | 0,224 | | | |
| Suspensos:Conclusos | | | | 9×10^{-4} | 0,001 |
| Tramitando:Despachos | | | | 10^{-4} | |
| Suspensos:Tramitando | | | | | 2×10^{-4} |

Nas extremidades, a variável Formato se mostrou a que mais aumenta o tempo dentre todas, corroborando com as expectativas da Seção 4.1.3 para esses casos (embora não tenha sido o caso para o quantil de 0,25 e 0,5). Para o quantil 0,9 (os seja, para os processos mais demorados), o formato físico aumenta o tempo até a baixa em 2.315 dias (6,3 anos) em relação ao formato eletrônico, mantendo as demais covariáveis constantes.

A Figura 13 mostra a variação dos coeficientes angulares para cada variável, com seus respectivos intervalos de confiança.

Figura 13: Coeficientes angulares para as principais variáveis



Os coeficientes das variáveis quantitativas se mostraram diferentes em função do quantil ajustado, mostrando que a mudança dos hiperplanos não se trata de um mero deslocamento dos interceptos, mas de uma mudança da inclinação.

A linha sólida em vermelho representa o valor do coeficiente para o modelo linear ajustado com as mesmas covariáveis, e as linhas tracejadas representam os limites dos intervalos de confiança para o valor dos coeficientes. É notável que, no caso das variáveis "Suspensos" e "Conclusos", os coeficientes para a mediana e para o modelo linear são muito próximos, o que não ocorre com as outras covariáveis.

4.2.3 Modelos de regressão gaussiana

Para comparações posteriores com o modelo para a mediana, serão testados alguns modelos de regressão gaussiana.

O Modelo 3 faz a seleção pelo stepwise através do menor AIC, seguida do stepwise do p-valor, considerando todas as covariáveis de forma linear, sem interações.

Modelo 3: Regressão gaussiana

| Variável | Coefficiente | Erro Padrão | P-valor |
|-------------------------------|--------------|-------------|-------------|
| Intercepto | 1,1 ano | 24 | ≈ 0 |
| Grau | -4,7 meses | 26 | ≈ 0 |
| Formato | 2,7 anos | 271 | ≈ 0 |
| Procedimento - Outros | -7,2 meses | 60 | ≈ 0 |
| Procedimento - Pré-processual | -8,8 meses | 91 | 0,003 |
| Tramitando | 1h35min | 0,009 | ≈ 0 |
| Decisões | -9h43min | 0,06 | ≈ 0 |
| Conclusos | 4h32min | 0,033 | ≈ 0 |
| Suspensos | 4h10min | 0,025 | ≈ 0 |
| Despachos | 3h23min | 0,059 | 0,017 |
| Audiências | 7 dias | 2 | 0,012 |

O modelo selecionado é muito parecido com o Modelo 1, com a adição das variáveis Formato e Audiências.

A variável Audiências não apareceu em nenhum dos modelos de regressão quantílica. No modelo gaussiano, por sua vez, ela apresentou o maior coeficiente entre as variáveis quantitativas, o que é esperado, dada a baixa amplitude da variável em relação às outras (Figura 12).

O Modelo 3 apresenta o resultado da seleção de variáveis considerando as covariáveis presentes no Modelo 3, mais as interações entre elas e os termos quadráticos.

Modelo 4: Regressão gaussiana com interações

| Variável | Coefficiente | Erro Padrão | P-valor |
|-----------------------------------|--------------|----------------------|-------------|
| Intercepto | 1,3 ano | 21 | ≈ 0 |
| Grau | -7,4 meses | 23 | ≈ 0 |
| Procedimento - Pré-processual | -10,8 meses | 90 | ≈ 0 |
| Procedimento - Outros | -5,6 meses | 60 | 0,004 |
| Formato | 2,5 anos | 268 | ≈ 0 |
| Tramitando | 2h18min | 0,012 | ≈ 0 |
| Decisões | -6h3min | 0,093 | 0,008 |
| Despachos | -6h18min | 0,084 | 0,005 |
| Conclusos | 3h10min | 0,039 | ≈ 0 |
| Rec. Interno Julgado ² | 0h28min | 0,008 | 0,011 |
| Suspensos:Despachos | 36 segundos | $6,8 \times 10^{-5}$ | ≈ 0 |
| Despachos:Audiências | 1h39min | 0,013 | ≈ 0 |
| Suspensos:Conclusos | 36 segundos | $8,5 \times 10^{-5}$ | ≈ 0 |
| Suspensos:Audiências | -1h26min | 0,006 | ≈ 0 |
| Tramitando:Decisões | -5 segundos | $1,4 \times 10^{-5}$ | ≈ 0 |

Diferente dos modelos quantílicos, a seleção com transformações não-lineares escolheu uma transformação quadrática para a variável Recurso Interno Julgado, com coeficiente menor do que todas as variáveis sem transformações e interações. O resultado conta com cinco coeficientes estimados a mais do que o Modelo 2, que é o análogo para a mediana.

Muitas interações também foram consideradas significativas. Porém, como mostra o Modelo 4, três delas apresentaram influência de poucos segundos, e tiveram o menor erro padrão entre os observados.

A Tabela 10 mostra os coeficientes presentes em cada um dos modelos gaussianos.

Tabela 10: Coeficientes significativos para cada modelo gaussiano

| Variável | Coeficiente | |
|-----------------------------------|-------------|----------|
| | Modelo 3 | Modelo 4 |
| Intercepto | ✓ | ✓ |
| Formato | ✓ | ✓ |
| Procedimento - Pré-processual | ✓ | ✓ |
| Procedimento - Outros | ✓ | ✓ |
| Grau | ✓ | ✓ |
| Tramitando | ✓ | ✓ |
| Decisões | ✓ | ✓ |
| Conclusos | ✓ | ✓ |
| Despachos | ✓ | ✓ |
| Suspensos | ✓ | |
| Audiências | ✓ | |
| Rec. Interno Julgado ² | | ✓ |
| Suspensos:Despachos | | ✓ |
| Despachos:Audiências | | ✓ |
| Suspensos:Conclusos | | ✓ |
| Suspensos:Audiências | | ✓ |
| Tramitando:Decisões | | ✓ |

A variável de "Recurso Interno Julgado" foi a única das variáveis do Modelo 4 que não apareceu de nenhuma forma no Modelo 3, e só foi considerada significativa em uma transformação quadrática. As outras variáveis são todas comuns entre os modelos, embora não se apresentem na mesma configuração.

As variáveis Suspensos e Audiências, que estão presentes no Modelo 3 de forma independente, só são significativas no Modelo 4 com interações, tanto entre as duas, quanto com as outras variáveis independentes do modelo.

A Tabela 11 mostra o AIC e o BIC para cada modelo.

Tabela 11: AIC, BIC, MAE e número de coeficientes para os modelos gaussianos

| Modelo | AIC | BIC | MAE | Coeficientes |
|----------|--------|--------|-----|--------------|
| Modelo 3 | 51.089 | 51.163 | 255 | 11 |
| Modelo 4 | 51.023 | 51.121 | 252 | 15 |

De forma análoga ao Modelo 1 e ao Modelo 2, os modelos gaussianos sem transfor-

mações não são os que minimizam o AIC e o BIC. Os dois critérios consideram o Modelo 4 como o melhor, apesar de ter mais coeficientes estimados.

Apesar da superioridade no que se refere ao BIC e AIC, o Modelo 4 é muito menos interpretável que o Modelo 3, característica que não é considerada pelo BIC e pelo AIC, e que será priorizada para os modelos de tendência central, tal qual como foi com as regressões para a mediana.

Enunciando o Modelo 3, tem-se que uma audiência adicional aumenta o tempo médio até a baixa dos processos da vara em 7 dias, em média (mantendo as demais covariáveis constantes). O processo ser físico ao invés de eletrônico causa um aumento de mais de dois anos e meio no tempo, em média. Um processo em segundo grau, em média, leva cinco meses a menos para receber baixa do que um processo em primeiro grau. Quanto mais decisões o órgão julgador tomar, menor será o tempo médio até a baixa. Quanto maior o número de processos tramitando, maior é o tempo médio, mas um único processo a mais tramitando aumenta o tempo médio em quase duas horas, mantendo fixos os valores das outras covariáveis.

5 Conclusão

O interesse principal deste estudo foi investigar algumas das fontes de variabilidade da duração dos processos na Justiça do Trabalho, determinando quais dessas fontes são as mais influentes a nível de órgão julgador. Para isso, foram utilizados dados agregados oriundos da Base Nacional de Dados do Poder Judiciário - DataJud, disponibilizados publicamente pelo Conselho Nacional de Justiça. A variável de interesse escolhida foi o tempo que cada órgão julgador leva, em média, para dar baixa em um processo, uma vez recebido. As técnicas estatísticas utilizadas para avaliar essas influências foram a regressão gaussiana e a regressão quantílica.

A variável “Formato”, que determina se o processo tramita em ambiente físico ou eletrônico, se mostrou relevante para explicar a variabilidade dos tempos, especialmente dentre os que registraram durações mais longas. Essa variável já era observada pelo Conselho Nacional de Justiça desde a instituição do PJe (CNJ, 2013), cujo propósito era tornar os processos mais céleres. O estudo mostra que o PJe foi bem-sucedido em reduzir o tempo até a baixa nos processos, e a Justiça do Trabalho, atualmente, já conta com quase todo o seu acervo em ambiente eletrônico.

A variável “Grau”, que incentivou políticas de equilíbrio de orçamento devido à disparidade do número de processos (CNJ, 2022), também se mostrou relevante para as estimativas do tempo até a baixa, aparecendo em todos os modelos, tanto quantílicos quanto gaussianos. Já a variável “Originário”, apesar de apresentar diferenças nas durações de processos originários e recursais, não entrou em nenhum modelo final, pois não trazia novas informações que já não constavam na variável “Grau”, dadas as fortes associações.

Os procedimentos “Pré-processual”, “Conhecimento não-criminal” e “Outros”, que estão entre os procedimentos com menores tempos até a baixa, se mostraram relevantes em muitos dos modelos. O “Pré-processual” foi o único procedimento que apareceu em todos eles, mostrando que, no geral, processos em procedimento pré-processual podem levar vários meses a menos para receber baixa quando comparado com os outros procedimentos.

O número de processos na situação “Tramitando” se mostrou a mais relevante das variáveis quantitativas. Apesar de, individualmente, aumentar poucas horas na duração média dos processos, ela foi a única dentre as variáveis quantitativas que apareceu consistentemente em todos os modelos, tanto quantílicos quanto gaussianos. As variá-

veis “Suspensos”, “Despachos”, “Decisões”, “Conclusos”, “Recurso Interno Pendente” e “Recurso Interno Julgado” também apareceram em mais de um modelo. A variável “Decisões” foi a única dentre as variáveis quantitativas citadas que apresentou influência negativa no tempo em todos eles, de modo que, quanto mais decisões forem proferidas, menor é, tanto o tempo médio, quanto o tempo para os quantis até a baixa dos processos. Os indicadores quantitativos mais determinantes para a variabilidade dos tempos médios até a baixa foram: Processos Tramitando, Decisões, Processos Conclusos, Suspensões e Despachos. No que se refere às variáveis qualitativas, as variáveis Grau, Formato e Procedimento - Pré-processual foram as mais relevantes.

O estudo demonstrou que a regressão quantílica apresenta resultados promissores no contexto da estimação de métricas temporais do Poder Judiciário, mostrando não só que essa predição é possível, como que as variáveis disponíveis publicamente são úteis para predizê-las. O método se torna mais relevante pelo fato de que os efeitos de cada variável são diferentes em cada um dos níveis de tempo.

Apesar de o estudo ser restrito à Justiça do Trabalho, modelagens semelhantes podem ser feitas para os outros ramos de justiça, sendo possível investigar a distribuição dos tempos até a baixa para cada ramo. De forma análoga, existem outras variáveis temporais oriundas do DataJud, como tempo médio do pendente líquido, tempo médio do pendente e tempo médio até o julgamento, que poderiam ser estudadas através de metodologias semelhantes. Outras variáveis que não estavam disponíveis no mesmo nível de agregação dos dados desse trabalho, como número de servidores, juízes e magistrados, permitem construir novas hipóteses e abrem espaço para estudos futuros.

Referências

- CNJ. *Resolução N^o 185, de 18 de dezembro de 2013*: Institui o sistema processo judicial eletrônico - pje como sistema de processamento de informações e prática de atos processuais e estabelece os parâmetros para sua implementação e funcionamento. 2013. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3512>>.
- CNJ. *Resolução N^o 331, de 20 de agosto de 2020*: Institui a base nacional de dados do poder judiciário – datajud como fonte primária de dados do sistema de estatística do poder judiciário – siespj para os tribunais indicados nos incisos ii a vii do art. 92 da constituição federal. 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3428>>.
- CNJ. *Resolução N^o 345, de 9 de outubro de 2020*: Dispõe sobre o “juízo 100% digital” e dá outras providências. 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3512>>.
- CNJ. *Justiça em Números*. 2022. Disponível em: <<https://www.cnj.jus.br/wp-content/uploads/2022/09/justica-em-numeros-2022-1.pdf>>.
- CNJ. *Painel de Estatísticas do Poder Judiciário*. 2023. Disponível em: <<https://painel-estatistica.stg.cloud.cnj.jus.br/estatisticas.html>>.
- HALL, P.; SHEATHER, S. J. On the distribution of a studentized quantile. *Journal of the Royal Statistical Society*, 1988. Disponível em: <<https://www.jstor.org/stable/2345702>>.
- KOCHERGINSKY, M.; HE, X.; MU, Y. Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, 2005. Disponível em: <<https://www.jstor.org/stable/27594096>>.
- KOENKER, R. *Quantile Regression*. [S.l.]: Cambridge University Press, 2005.
- KOENKER, R.; BASSETT, G. Regression quantiles. *The Econometric Society*, Janeiro 1978. Disponível em: <<http://www.jstor.org/stable/1913643>>.
- KUTNER, M. H. et al. *Applied Linear Statistical Models*. 5th. ed. [S.l.]: Cambridge University Press, 2008.
- MONTGOMERY, D. et al. *Introduction to Linear Regression Analysis*. 5th. ed. [S.l.]: Wiley, 2012.
- SANTOS, B. R. dos. *Modelos de Regressão Quantílica*. 2012. Disponível em: <<https://teses.usp.br/teses/disponiveis/45/45133/tde-12042012-154333/publico/DissertacaoBrunoRamosdosSantos.pdf>>.