



Universidade de Brasília
Departamento de Estatística

Estudo da adesão de políticas públicas da área de infraestrutura por
municípios brasileiros via análise de sobrevivência

Natália Oliveira Vieira

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

Brasília
2025

Natália Oliveira Vieira

**Estudo da adesão de políticas públicas da área de infraestrutura por
municípios brasileiros via análise de sobrevivência**

Orientadora: Prof^a Juliana Betini Fachini Gomes

Coorientador: Mateus Felipe Santos Araújo

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

**Brasília
2025**

Agradecimentos

Aos meus pais e meu irmão que me apoiaram, incentivaram e sempre me proporcionaram todas as oportunidades possíveis para meu crescimento pessoal e profissional. Mesmo em momentos que eu não acreditava em mim mesma, eles sempre acreditavam e me davam forças para continuar.

Ao Vítor, meu melhor amigo e parceiro. Durante essa trajetória seu companheirismo, suporte, conselhos, incentivo e principalmente amor.

A Prof^a. Juliana por toda a confiança, paciência e por exalar uma paz que sempre me acalmou durante o desenvolvimento desse trabalho.

A minha amiga Gabriela pela companhia e suporte diário durante os últimos anos, você deixou o curso mais leve e divertido.

Aos meus amigos Eduardo, Larissa e Isabela pelo apoio, amizade e por estarem sempre presentes em momentos de alegria ou de dificuldade.

Por fim, gostaria de agradecer a todos que me acompanharam na trajetória acadêmica e os que me auxiliaram na minha formação profissional.

Resumo

O presente estudo buscou analisar fatores que pudessem influenciar a adoção de políticas públicas da área de infraestrutura por 318 municípios brasileiros. Foram utilizadas técnicas estatísticas de análise de sobrevivência para estudar o tempo até a adesão de duas políticas dessa área, sendo elas o Programa de Aceleração do Crescimento - PAC (criado em 2007) e o Programa de Parcerias e Investimentos - PPI (criado em 2016). As variáveis explicativas foram selecionadas considerando a possível relação entre a adoção dessas políticas por um mesmo município. Durante o estudo foram ajustados os modelos de fragilidade compartilhada, de regressão paramétrica bivariada e univariada. Considerando o ajuste do modelo de fragilidade compartilhada, fatores como PIB, percentual de orçamento discricionário empenhado, ideologia partidária e anos eleitorais nacionais e municipais reduzem a probabilidade de adesão aos programas de infraestrutura, enquanto o percentual de conflito em medidas provisórias tem efeito positivo sobre a adoção das políticas. Os resultados do modelo de regressão bivariado evidenciaram a ausência de associação entre os tempos das políticas. O modelo univariado ajustado para a política PAC mostrou que o percentual de orçamento discricionário empenhado e os anos eleitorais reduzem a probabilidade de adesão da política, e o modelo univariado para a política PPI apontou que o percentual de conflito em medidas provisórias aumenta as chances de adoção de um município ao programa enquanto o tempo de mandato do ministro e o ano eleitoral nacional reduzem as chances de adoção. Por fim, a avaliação do ajuste dos modelos por meio da análise de resíduos de Cox-Snell indica que a adesão das políticas de infraestrutura e os fatores que a influenciam podem ser identificados e modelados por técnicas que não consideram dependência entre os tempos de um mesmo indivíduo, isto é, dentro os quatro modelos testados, os modelos univariados se adequaram de forma mais satisfatória aos dados.

Palavras-chaves: difusão de políticas públicas; municípios brasileiros; infraestrutura; dados censurados; modelos de regressão bivariado via cópulas; modelo de fragilidade compartilhada.

Lista de Tabelas

1	Variáveis e hipóteses usadas no estudo	28
2	Resultado do modelo de Fragilidade Compartilhada	45
3	Resultado do modelo de regressão bivariado através da cópula de Clayton para as políticas de Infraestrutura	49
4	Resultados dos modelos de regressão univariado logístico para o Programa de Aceleração do Crescimento	51
5	Resultados dos modelos de regressão univariado logístico para o Programa de Parcerias e Investimento	52

Lista de Figuras

1	Função de Sobrevivência estimada via Kaplan-Meier para as políticas de Infraestrutura	31
2	Função de Sobrevivência estimada via Kaplan-Meier por política pública .	32
3	Gráfico de barras da quantidade de municípios adotantes	32
4	Boxplot do PIB per capita pela indicador de adesão das políticas	33
5	Boxplots do PIB per capita pela situação de adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI	33
6	Boxplot da Taxa de orçamento discricionário empenhado pela situação da adesão das políticas	34
7	Boxplots da Taxa de orçamento discricionário empenhado pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI	35
8	Boxplot do Percentual de conflitos em MPs pela situação da adesão das políticas	35
9	Boxplots do Percentual de conflitos em MPs pela situação de adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI	36
10	Boxplot da Ideologia do prefeito pela situação da adesão das políticas . . .	36
11	Boxplots da Ideologia do prefeito pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI	37
12	Boxplot do Tempo de mandato dos ministros pela situação da adesão das políticas	37
13	Boxplots do Tempo de mandato dos ministros pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI	38
14	Curva de Kaplan-Meier por Ano Eleitoral Nacional	38
15	Gráfico de colunas do ano eleitoral nacional pela situação de adoção das políticas	39
16	Curva de Kaplan-Meier por Ano Eleitoral Nacional para política PACA . .	39

17	Gráfico de colunas do ano eleitoral nacional pela situação de adoção da política PAC	40
18	Curva de Kaplan-Meier por Ano Eleitoral Nacional para a política PPI . .	40
19	Gráfico de colunas do ano eleitoral nacional pela situação de adoção da política PPI	41
20	Curva de Kaplan-Meier por ano eleitoral municipal	41
21	Gráfico de colunas do ano eleitoral municipal pela situação de adoção das políticas	42
22	Curva de Kaplan-Meier por ano eleitoral municipal para a política PAC . .	42
23	Gráfico de colunas do ano eleitoral municipal pela situação de adoção da política PAC	43
24	Curva de Kaplan-Meier por ano eleitoral municipal para a política PPI . .	43
25	Gráfico de colunas do ano eleitoral municipal pela situação de adoção da política PPI	44
26	Gráfico da fragilidade compartilhada por município	47
27	Resíduo de Cox-Snell do modelo de fragilidade compartilhada	48
28	Curva de Kaplan-Meier estimada marginalmente para o Programa de Aceleração do Crescimento - PAC	50
29	Curva de Kaplan-Meier estimada marginalmente para o Programa de Parcerias e Investimentos - PPI. (a.1) Curva estimada completa e (a.2) Curva estimada com visualização em $0.9 \leq S(y_1) \leq 1$	50
30	Resíduos de Cox-Snell para os modelos de regressão univariados da política PAC. (a) Resíduos da política PAC com a primeira seleção de variáveis (b) Resíduos da política PAC com a segunda seleção de variáveis	52
31	Resíduos de Cox-Snell para os modelos de regressão univariados da política PAC. (a) Resíduos da política PPI com a primeira seleção de variáveis (b) Resíduos da política PPI com a segunda seleção de variáveis	53

Sumário

1 Introdução	9
2 Revisão Bibliográfica	11
2.1 Conceitos Básicos de Análise de Sobrevida	11
2.2 Especificação do Tempo de Falha	12
2.2.1 Função de Densidade de Probabilidade	12
2.2.2 Função de Sobrevida	12
2.2.3 Função Taxa de Falha	13
2.2.4 Função de Taxa de Falha Acumulada	13
2.2.5 Relação entre as funções	13
2.3 Estimador de Kaplan-Meier	14
2.4 Modelo de Regressão de Cox	15
2.4.1 Definição do modelo	15
2.4.2 Estimação	16
2.4.3 Avaliação do Ajuste do Modelo de Cox	18
2.5 Modelos de Eventos Múltiplos	19
2.5.1 Definição Geral	19
2.5.2 Modelo de Fragilidade Compartilhado	20
2.5.3 Estimação via Função de Verossimilhança Penalizada	21
2.6 Modelos de regressão paramétricos	22
2.6.1 Distribuição logística	22
2.6.2 Modelo de regressão logística	23
2.6.3 Cópula	23
2.6.4 Modelo de regressão bivariado	25
3 Metodologia	27
3.1 Material	27
3.2 Método	29

4 Resultados	31
4.1 Análise Descritiva	31
4.2 Modelagem	44
4.2.1 Modelo de Fragilidade Compartilhada	44
4.2.2 Modelos paramétricos	48
5 Conclusão	54
Referências	57

1 Introdução

A difusão de políticas públicas, como definida por Rogers (1983), é o processo em que uma inovação é comunicada ao longo do tempo entre membros de um sistema social. Ele também a caracteriza como sendo cumulativa no tempo, uma vez que a replicação de uma nova política depende de um conjunto de variáveis.

Os autores Berry e Berry (1990) fizeram o seguinte questionamento "O que leva um governo a adotar um novo programa ou política?". Como solução, eles ofereceram duas respostas que consideraram fundamentais, que podem ser classificadas em duas categorias: o modelo de determinantes internos e o modelo de determinantes externos.

Assim, o modelo de determinantes internos tem os fatores políticos, econômicos e sociais explicando a motivação de um governo por inovar. Já o modelo de determinantes externos enfatiza a influência dos governos vizinhos como fator explicativo (BERRY, 1987; BERRY; BERRY, 1990).

No contexto de Ciências Políticas, o Event History Analysis (EHA) ou análise de sobrevivência permite mensurar no tempo o efeito dos determinantes internos e externos sobre a adoção das políticas (COELHO; CAVALCANTE; TURGEON, 2016). A principal vantagem desse método é a utilização de informações como quantidade de casos, o momento, a sequência e a duração dos eventos (BOX-STEFFENSMEIER; JONES, 1997).

No Brasil, o estudo do processo de difusão de políticas públicas é emergente, sendo o caso brasileiro considerado um cenário propício para representar uma unidade de análise comparativa. Devido ao seu federalismo descentralizado, com esferas subnacionais de governo autônomas, os municípios brasileiros podem ser vistos como um grande laboratório para o estudo de políticas públicas (COELHO; CAVALCANTE; TURGEON, 2016; MELLO, 2021).

Segundo Pereira (2018), as políticas públicas de infraestrutura brasileira têm recebido um aumento de investimentos nos últimos anos, embora ela apresente desafios fiscais, administrativos e políticos. Os programas dessa área focam em construções, manutenções e aprimoramentos nos campos de comunicação, energia, saneamento e transporte buscando alavancar o desenvolvimento socioeconômico do país.

Por esse motivo, este trabalho foca no estudo das políticas de infraestrutura. O principal interesse é entender quais fatores influenciam na adoção de um programa público em um município, considerando que ele pode aderir mais de uma política. Dessa forma, por meio de modelos de eventos múltiplos de análise de sobrevivência, é possível analisar

as motivações que podem acelerar o processo de difusão das políticas de infraestrutura nos municípios.

Além disso, esse trabalho agrega novas análises no estudo da difusão de políticas públicas no Brasil, trazendo estudo dos modelos de regressão logísticos bivariados e univariados e uma comparação entre métodos para identificar e avaliar padrões nos fatores que influenciam a adesão dos programas de infraestrutura.

2 Revisão Bibliográfica

2.1 Conceitos Básicos de Análise de Sobrevivência

A Análise de Sobrevivência é a área da estatística que engloba uma série de técnicas e procedimentos para analisar dados em que a variável resposta é o tempo até a ocorrência de um evento. Esse tempo é denominado tempo de falha e é composto pelo tempo de início, momento em que deve ser possível comparar todos os indivíduos do estudo; da escala de medida, que costuma ser o tempo real ou "de relógio"; e o evento de interesse ou falha, que pode ocorrer mais de uma vez e deve ser definido de forma clara e precisa.

Os modelos que caracterizam a estrutura dos modelos de sobrevivência além de apresentar o tempo de falha como uma das suas características também é constituído pela presença das censuras. Elas são a observação parcial ou incompleta da resposta e podem ocorrer por diversos motivos, como a perda do acompanhamento do indivíduo ou a não ocorrência da falha até o final do estudo.

É importante ressaltar que mesmo com a presença de censuras, todos os dados do estudo de sobrevivência devem ser usados na análise. Segundo Colosimo e Giolo (2006), existem duas razões para a não exclusão das observações censuradas do estudo, a primeira diz que mesmo incompletas elas fornecem informação sobre o tempo de vida dos indivíduos e a segunda é que a omissão das censuras pode acarretar em conclusões viciadas

Desse modo, a variável resposta é dada pelo par (t_i, δ_i) , em que t_i é o tempo observado do indivíduo i e δ_i é a variável indicadora de censura, que é dada por:

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é o tempo de falha;} \\ 0, & \text{se } t_i \text{ é o tempo de censura.} \end{cases}$$

As censuras podem ser classificadas em três categorias: censura à direita, censura à esquerda ou censura intervalar. A censura à esquerda ocorre quando o tempo registrado é maior que o tempo de falha, isto é, o evento de interesse já aconteceu quando o indivíduo foi observado. O mecanismo de censura intervalar acontece quando a única informação que se tem é que o evento de interesse decorreu em um certo intervalo de tempo. Por fim, a censura à direita ocorre quando o tempo de falha é maior que o tempo registrado.

Além disso, a censura à direita pode ser classificada em:

- Censura do tipo I: é aquela em que o estudo será finalizado após um período preestabelecido de tempo e indivíduos que não tiverem sofrido a falha serão censurados;
- Censura do tipo II: é aquela em que o estudo terá término após a ocorrência do evento de interesse em um número pré-definido de indivíduos e os demais são censurados;
- Censura aleatória: é aquela em que o indivíduo é retirado do estudo por um motivo qualquer sem ter ocorrido a falha.

2.2 Especificação do Tempo de Falha

O tempo de falha, denotado por T , é uma variável aleatória não negativa e usualmente contínua. Em análise de sobrevivência, a variável T é especificada pela função de densidade, função de sobrevivência e função de risco. As informações dessa subseção foram baseadas em Colosimo e Giolo (2006).

2.2.1 Função de Densidade de Probabilidade

A função de densidade de probabilidade é definida como a probabilidade de um indivíduo sofrer ao evento de interesse em um intervalo de tempo Δt . Ela é dada por:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}, \quad (2.2.1)$$

com $f(t) \geq 0, \forall t$ e $\int_0^\infty f(t)dt = 1$.

2.2.2 Função de Sobrevivência

A função de sobrevivência é definida como a probabilidade de um indivíduo não falhar até o tempo t , isto é, a probabilidade desse indivíduo sobreviver ao tempo t , ela pode ser representada como:

$$S(t) = P(T > t), \quad (2.2.2)$$

com $t \geq 0$ (COX; OAKES, 2018).

Além de ser uma das principais funções probabilísticas usada para descrever o tempo de sobrevivência, ela é uma função não-crescente e contínua, tal que $\lim_{t \rightarrow 0} S(t) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

Como consequência da função de sobrevivência, tem-se que a função de distribuição acumulada é a probabilidade de um indivíduo não sobreviver ao tempo t . A função de distribuição acumulada é expressa por:

$$F(t) = 1 - S(t). \quad (2.2.3)$$

2.2.3 Função Taxa de Falha

A taxa de falha é definida como a probabilidade de que a falha ocorra em um intervalo, dado que sobreviveu ao tempo t . Segundo Lawless (2003), ela pode ser definida por :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (2.2.4)$$

A taxa de falha é uma função muito utilizada para descrever a distribuição do tempo de vida de um indivíduo.

2.2.4 Função de Taxa de Falha Acumulada

A função de taxa de falha acumulada fornece o risco acumulado do indivíduo no tempo t e ela é definida por:

$$H(t) = \int_0^t h(u) du, t \geq 0. \quad (2.2.5)$$

$H(t)$ não possui uma interpretação direta, mas ajuda na avaliação da função de risco $h(t)$.

2.2.5 Relação entre as funções

As funções especificadas para o tempo de falha possuem algumas relações matemáticas, de forma que o conhecimento de uma delas implica no conhecimento das demais. De acordo com Colosimo e Giolo (2006), as relações entre elas são expressas por:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t)), \quad (2.2.6)$$

$$S(t) = \exp(-H(t)) = \exp\left\{-\int_0^t h(u)du\right\}, \quad (2.2.7)$$

o que implica em

$$H(t) = -\int_0^t h(u)du = -\log S(t). \quad (2.2.8)$$

2.3 Estimador de Kaplan-Meier

A análise descritiva estatística consiste em encontrar medidas de tendência central e variabilidade, entretanto a presença da censura invalida esse método tradicional. Para a análise de sobrevivência, a análise descritiva consiste no estudo da função de sobrevivência, a qual é estimada por meio do estimador de Kaplan-Meier, proposto por Kaplan e Meier (1958).

O estimador de Kaplan-Meier é um estimador não-paramétrico e também é chamado de estimador limite-produto. Quando não há a presença de censura, ele consiste a função de sobrevivência empírica, que é dada por:

$$\hat{S}(t) = \frac{\text{nº de observações que não falharam até o tempo } t}{\text{nº total de observações no estudo}}, \quad (2.3.1)$$

sendo que $\hat{S}(t)$ é uma função escada, na qual os degraus são os tempos registrados de falha de tamanho $\frac{1}{n}$, em que n é o tamanho da amostra.

Além disso, em sua construção, o estimador de Kaplan-Meier considera tantos intervalos de tempo quanto forem as quantidades de falhas distintas, o que implica nos tempos de falhas sendo os limites dos intervalos de tempos.

A definição do estimador de Kaplan-Meier é:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j}\right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right), \quad (2.3.2)$$

em que:

- $t_1 < t_2 < \dots < t_k$ são os k tempos distintos e ordenados de falha;
- d_j é o número de falhas no tempo t_j , para $j = 1, 2, \dots, k$;
- n_j é a quantidade de indivíduos sob risco em t_j , isto é, os indivíduos que não falharam

e não foram censurados no momento imediatamente anterior a t_j .

Por fim, Colosimo e Giolo (2006) definem que as principais características do estimador de Kaplan-Meier são:

- não-viciado para amostras grandes;
- fracamente consistente;
- converge assintoticamente para um processo gaussiano;
- é o estimador de máxima verossimilhança de $S(t)$.

2.4 Modelo de Regressão de Cox

Embora as técnicas não-paramétricas sejam de suma importância para a descrição dos dados, elas não permitem a inclusão direta de covariáveis no estudo. Para a inclusão dessas é necessário a utilização de modelos de regressão.

No caso da análise de sobrevivência, existem duas categorias de modelos de regressão: os modelos paramétricos e os semiparamétricos. Os modelos paramétricos, também conhecidos como modelos de tempo de vida acelerado, costumam ser mais eficientes que os modelos semiparamétricos, mas são menos flexíveis. Os modelos semiparamétricos, também denominados de modelos de regressão de Cox, também permitem a incorporação de covariáveis dependentes do tempo.

2.4.1 Definição do modelo

Proposto por Cox (1972), o modelo de regressão de Cox é formulado pela função:

$$h(t_i|\mathbf{x}_i) = h_0(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad (2.4.1)$$

em que $h_0(t)$ é a função de risco base, ou seja, o risco do indivíduo com vetor de covariáveis nulo. Ele é o componente não-paramétrico do modelo. O componente paramétrico é utilizado de forma multiplicativa por meio de uma função $g(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}$, sendo $\mathbf{x}$ o vetor de covariáveis observadas para o indivíduo i e $\boldsymbol{\beta}$ o vetor de parâmetros associados às covariáveis, ambos vetores com dimensão p e $i = 1, \dots, n$.

O modelo de regressão de Cox possui a suposição de riscos proporcionais, pois

a razão dos riscos de dois indivíduos diferentes não depende do tempo (COLOSIMO; GIOLO, 2006). Essa razão é dada por:

$$\frac{h(t|\mathbf{x}_i)}{h(t|\mathbf{x}_j)} = \frac{h_0(t) \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{h_0(t) \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} - \mathbf{x}_j^T \boldsymbol{\beta}\}. \quad (2.4.2)$$

Por conta disso, o modelo de regressão de Cox também é denominado modelo de riscos proporcionais.

Dessa forma, a função de sobrevivência do modelo de Cox para um indivíduo com vetor de covariáveis $\mathbf{x}$ é dada por:

$$S(t|\mathbf{x}) = (S_0(t))^{\exp\{\mathbf{x}^T \boldsymbol{\beta}\}}, \quad (2.4.3)$$

na qual, $S_0(t) = \exp\{-\int_0^t h_0(u)du\}$ é a função de sobrevivência e $\boldsymbol{\beta}$ é o vetor de parâmetros das covariáveis.

2.4.2 Estimação

Na literatura de estatística, existem vários métodos para estimação dos parâmetros de um modelo. No caso da análise de sobrevivência, utiliza-se o método de máxima verossimilhança, pois ele consegue incorporar as censuras, sendo simples de ser entendido e tendo ótimas propriedades para grandes amostras.

O método de máxima verossimilhança baseia-se nos resultados obtidos da amostra para identificar qual é a distribuição com maior possibilidade de ter gerado a amostra obtida dentre todas possíveis, considerando os valores dos parâmetros.

Assim, dado uma amostra de tamanho n com os tempos observados, $t_1, t_2, \dots, t_n$, a função de verossimilhança para um vetor de parâmetros $\boldsymbol{\theta}$, é expressa por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i; \boldsymbol{\theta}). \quad (2.4.4)$$

Na função de verossimilhança, os tempos de falhas contribuem com a sua função de densidade. Entretanto, as observações censuradas contribuem com sua função de sobrevivência, informando que o tempo de falha é maior que o de censura observado. Logo, tem-se a seguinte expressão:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i}. \quad (2.4.5)$$

Os estimadores de máxima verossimilhança são os valores de $\boldsymbol{\theta}$ que maximizam $L(\boldsymbol{\theta})$ ou equivalentemente $\log L(\boldsymbol{\theta})$ (COLOSIMO; GIOLO, 2006). Eles são encontrados pelo desenvolvimento do sistema de equações:

$$U(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \quad (2.4.6)$$

Logo, o método de máxima verossimilhança é utilizado para encontrar as estimativas pontuais dos parâmetros do modelo. A partir das propriedades para grandes amostras, é possível construir intervalos de confiança para essas estimativas dos parâmetros.

Entretanto, para o modelo de Cox, a presença do componente não-paramétrico $h_0(t)$ na função da verossimilhança torna o método da máxima verossimilhança inapropriado. Por isso, Cox (1975) propôs o método da máxima verossimilhança parcial.

Considere uma amostra de n indivíduos com $t_1 < t_2 < \dots < t_k$ tempos e $k \leq n$ falhas distintas. Conhecendo as observações que estão sob risco em t_i , a probabilidade condicional da i -ésima observação vir a falhar no tempo t_i é:

$$\frac{h_i(t|\mathbf{x}_i)}{\sum_{j \in R(t_i)} h_j(t|\mathbf{x}_j)} = \frac{h_0(t) \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{\sum_{j \in R(t_i)} h_0(t) \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}} = \frac{\exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}}, \quad (2.4.7)$$

em que $R(t_i)$ é o conjunto do índices das observações sob risco no tempo t_i .

A função de verossimilhança a ser utilizada para se fazer inferências acerca dos parâmetros do modelo é formada pelo produto de todos os termos apresentados em 2.4.7 associados aos tempos distintos de falhas, isto é:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{\prod_{j \in R(t_i)} \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}} = \prod_{i=1}^n \left(\frac{\exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{\prod_{j \in R(t_i)} \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}} \right)^{\delta_i}, \quad (2.4.8)$$

em que δ_i é o indicador de censura.

Ainda segundo Colosimo e Giolo (2006), os valores de $\boldsymbol{\beta}$ que maximizam a função de verossimilhança parcial $L(\boldsymbol{\beta})$ são obtidos resolvendo o sistema de equações $U(\boldsymbol{\beta}) = 0$, ou seja,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left[x_i - \frac{\sum_{j \in R(t_i)} x_j \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}}{\exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}} \right] = 0, \quad (2.4.9)$$

em que $U(\boldsymbol{\beta})$ é vetor escore de derivadas de primeira ordem da função $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$.

A função de verossimilhança parcial assume que os tempos de sobrevivência são contínuos e não pressupõe a possibilidade de empates nos valores observados (COLOSIMO; GIOLO, 2006). Para isso, Breslow (1972) e Peto (1972) propuseram uma aproximação para 2.4.8, em que é possível incorporar as observações empatadas. Ela é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp\{\mathbf{s}_i^T \boldsymbol{\beta}\}}{\left[\sum_{j \in R(t_i)} \exp\{\mathbf{x}_j^T \boldsymbol{\beta}\} \right]^{d_i}}. \quad (2.4.10)$$

Na expressão 2.4.10, considera-se s_i como o vetor formado pela soma das p covariáveis para os indivíduos que falharam no mesmo tempo $t_i (i = 1, \dots, k)$ e d_i o número de falhas nestes tempos. Esta aproximação só é adequada quando o número de observações empatadas em qualquer tempo não é grande. (COLOSIMO; GIOLO, 2006)

2.4.3 Avaliação do Ajuste do Modelo de Cox

Embora o modelo de regressão de Cox tenha como característica ser mais flexível devido à presença do componente não-paramétrico, assim como nos modelos paramétricos, é necessário realizar técnicas estatísticas para avaliar o ajuste do modelo. Particularmente, uma das suposições do modelo de Cox são as taxas de falha proporcionais, que uma vez violadas acarretam em sérios vícios na estimação dos coeficientes.

Para a avaliação da proporcionalidade, existem métodos gráficos e testes estatísticos. Uma das técnicas gráficas consiste na divisão dos dados em m estratos, segundo uma variável, e estima-se $\hat{H}_0(t)$ para cada estrato usando a expressão:

$$\hat{H}_0(t) = \sum_{j: t_j < t} \frac{d_j}{\sum_{l \in R(t_j)} \exp\{\mathbf{x}_l^T \boldsymbol{\beta}\}}, \quad (2.4.11)$$

sendo que d_i é o número de falhas no tempo t_i .

Em seguida, com a estimativa de $\hat{H}_0(t)$, constrói-se o gráfico das curvas do $\log \hat{H}_0(t)$ versus t ou $\log(t)$, se a suposição de taxa de falha proporcionais for válida,

as curvas devem apresentar diferenças aproximadamente constantes no tempo.

Outro método utilizado são os resíduos de Cox-Snell, que também auxiliam na avaliação do ajuste dos modelos paramétricos. Esses resíduos são definidos por:

$$\hat{e}_i = \hat{H}_0(t_i) \exp \left\{ \sum_{k=1}^p \mathbf{x}_{ik} \beta_k \right\}, i = 1, \dots, n. \quad (2.4.12)$$

Se o modelo estiver bem ajustado, os resíduos de Cox-Snell devem seguir uma distribuição exponencial padrão e o gráfico $\hat{H}(\hat{e}_i)$ versus $\hat{e}_i$ deve ser aproximadamente uma reta.

Outro resíduo usado é o resíduo martingal, que são uma modificação do resíduo de Cox-Snell. Os resíduos martingal para o modelo de Cox quando os dados apresentarem censuras à direita e todas as variáveis não forem dependentes do tempo, pode ser definido por:

$$\hat{m}_i = \delta_i - \hat{H}_0(t_i) \exp \left\{ \sum_{k=1}^p \mathbf{x}_{ik} \hat{\beta}_k \right\} = \delta_i - \hat{e}_i, i = 1, \dots, n. \quad (2.4.13)$$

Além disso, os resíduos deviance são usados para detecção de pontos atípicos e no modelo de taxas de falhas proporcionais são definidos por:

$$\hat{d}_i = \text{sin}(\hat{m}_i) \left[-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i)) \right]^{1/2}. \quad (2.4.14)$$

Os resíduos deviance não são tão assimétricos quanto os resíduos martingal. Dessa forma, se o modelo for apropriado, os resíduos deviance apresentam um comportamento aleatório em torno de zero.

2.5 Modelos de Eventos Múltiplos

2.5.1 Definição Geral

Em muitos estudos, existe a suposição de que os tempos de sobrevivência distintos são independentes. Entretanto, em estudos que os tempos de sobrevivência são observados em grupos ou conglomerados de indivíduos, não se pode assumir independência entre os tempos de um mesmo grupo. Nesses casos, espera-se que o comportamento dos tempos de indivíduos de um mesmo grupo apresente certas semelhanças que não seriam observadas

entre indivíduos de grupos distintos.

Outra situação em que a suposição de tempos independentes não é válida ocorre quando cada indivíduo de um estudo está sujeito a múltiplos eventos do mesmo tipo, chamados de eventos recorrentes. Nesses casos, é possível observar mais de um tempo de sobrevivência para um indivíduo, assim pode-se supor que exista associação entre os tempos de um mesmo indivíduo.

Os eventos múltiplos podem ser do mesmo tipo, como a recorrência de um câncer ou ataques cardíacos; ou eles podem ser de tipos diferentes como acidentes cardiovasculares ou complicações pós-operatórias.

Nas situações citadas anteriormente é possível supor a existência de associação entre os tempos de sobrevivência e modelos que supõe a existência dessa associação devem ser utilizados para analisar dados dessa natureza.

2.5.2 Modelo de Fragilidade Compartilhado

No caso da análise de sobrevivência, um modelo que tem sido usado frequentemente é o modelo de fragilidade. Nesse modelo, um efeito aleatório, denominado fragilidade, é introduzido na função de taxa de falha para descrever a possível associação entre os tempos de sobrevivência.

De acordo com Colosimo e Giolo (2006), o nome de modelo de fragilidade decorre do fato de que quanto maior for o valor do efeito aleatório maior será o risco de uma falha ocorrer, isto é, mais frágeis os indivíduos do grupos estarão para falhar.

Os modelos de fragilidade possuem uma abordagem condicional, uma vez que assumem que os tempos que apresentam uma possível associação são independentes condicionalmente às variáveis de fragilidade. Nesse cenário, o modelo de fragilidade compartilhada vem sendo utilizado com frequência em estudos da análise de sobrevivência em que múltiplos eventos são observados para os indivíduos.

No modelo de fragilidade compartilhado, a fragilidade representa um efeito aleatório que descreve um risco comum, ou seja, uma fragilidade compartilhada entre os indivíduos de um mesmo grupo. Segundo Colosimo e Giolo (2006), a ideia do modelo é que os grupos apresentem fragilidades diferentes, sendo que grupos com fragilidades de valores grandes experimentarão o evento de interesse em tempos menores que aqueles grupos com fragilidades em valores menores.

Assim, para o modelo de fragilidade compartilhado, formalmente, considere que

$T_j = (T_{1j}, T_{2j}, \dots, T_{n_jj})^T$ os n_j tempos de sobrevivência do j -ésimo grupo e Z_j a variável de fragilidade não observada associada a este grupo. Para $Z_j = z_j$, é assumido, condicionalmente a z_j , que os componentes de T_j são independentes com as distribuições dos T_{ij} modeladas por:

$$h_{ij} = z_j h_0(t) \exp\{\mathbf{x}_{ij}^T \boldsymbol{\beta}\}, \quad (2.5.1)$$

para $i = 1, \dots, n_j, j = 1, \dots, m$ e com $h_{ij}(t)$ a função de risco para T_{ij} condicionalmente ao valor não observado z_j e um vetor $\mathbf{x}_{ij}$ de dimensão p de covariáveis, $h_0(t)$ uma função de risco base desconhecida, $\boldsymbol{\beta}$ um vetor de dimensão p de coeficientes de regressão desconhecidos e $z_j, (j = 1, \dots, m)$ os valores das fragilidades, assumidas serem uma amostra independente de variáveis aleatórias Z_j com distribuição de probabilidade conhecida de média 1 e variância desconhecida. Além disso, Z_j é assumida constante no tempo.

Segundo Colosimo e Giolo (2006), diversas distribuições têm sido proposta na literatura para a variável de fragilidade, sendo a mais comum a distribuição gama por sua conveniência algébrica. Nesse trabalho, será utilizada a distribuição gaussiana, pois ela apresentou um ajuste mais satisfatório aos dados.

2.5.3 Estimação via Função de Verossimilhança Penalizada

Diversos processos de estimação têm sido usados para a estimação e otimização da função de verossimilhança. O algoritmo EM permite a estimação de dados perdidos (*missing*), sendo assim utilizado nas variáveis de fragilidade. Entretanto, esse método apresenta alguns problemas como: ser relativamente lento em algumas situações, em grandes amostras não convergir para um resultado, e em alguns pacotes estatísticos não possuir uma implementação disponível, sendo necessário criar um código ou função para sua utilização.

Uma alternativa proposta para o modelo de fragilidade compartilhada é utilizar a função de verossimilhança parcial penalizada no processo de estimação (HOUGAARD, 2000; THERNEAU, 2000). Esse método possui algumas similaridades com o algoritmo EM e é baseado em uma modificação da função de verossimilhança parcial de Cox.

Formalmente, a função de verossimilhança é descrita como um produto em que o primeiro termo é a função de verossimilhança parcial, incluindo as fragilidades como parâmetros, e o segundo termo é uma penalidade introduzida para evitar diferenças grandes entre as fragilidades para os diferentes grupos. Dessa forma, o logaritmo da função

de verossimilhança parcial penalizada é expresso por:

$$PPL(\boldsymbol{\beta}, w, \theta) = \log(L(\boldsymbol{\beta}, w)) - g(w, \theta), \quad (2.5.2)$$

sendo

$$\log(L(\boldsymbol{\beta}, w)) = \sum_{i=1}^n \delta_i \left[(\mathbf{x}'_{ij} \boldsymbol{\beta} + w_j) - \log \left(\sum_{k \in R(t_i)} \exp\{\mathbf{x}'_{ik} \boldsymbol{\beta} + w_k\} \right) \right], \quad (2.5.3)$$

e $g(w, \theta)$ a função de penalidade. O uso do logaritmo de uma densidade como função de penalidade é bastante frequente.

2.6 Modelos de regressão paramétricos

2.6.1 Distribuição logística

Em análise de sobrevivência, uma distribuição que tem sido bastante utilizada devido a sua aplicabilidade é a distribuição log-logística. Considerando T como uma variável aleatória com essa distribuição, as funções de densidade, sobrevivência e taxa de falha podem ser expressas por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} (1 + (t/\alpha)^\gamma)^{-2}, \quad (2.6.1)$$

$$S(t) = \frac{1}{1 + (t/\alpha)^\gamma}, \quad (2.6.2)$$

e

$$\lambda(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^\gamma]}, \quad (2.6.3)$$

sendo $\alpha > 0$ o parâmetro de escala, e $\gamma > 0$ o parâmetro de forma.

Assim, se T segue uma distribuição log-logística, a variável transformada $Y = \log(T)$ segue a distribuição logística com as funções de densidade de probabilidade, sobrevivência e taxa de falha dadas por:

$$f(t) = \frac{1}{\sigma} \exp \left\{ \frac{y - \mu}{\sigma} \right\} \left(1 + \exp \left\{ \frac{y - \mu}{\sigma} \right\} \right)^{-2}, \quad (2.6.4)$$

$$S(t) = \frac{1}{1 + \exp \left\{ \frac{y - \mu}{\sigma} \right\}}, \quad (2.6.5)$$

e

$$\lambda(t) = \frac{1}{\sigma} \exp \left\{ \frac{y - \mu}{\sigma} \right\} \left(1 + \exp \left\{ \frac{y - \mu}{\sigma} \right\} \right)^{-1}, \quad (2.6.6)$$

com $-\infty < \mu < \infty$ e $\sigma > 0$, sendo os parâmetros de locação e escala, respectivamente. Ainda, os parâmetros da distribuição log-logística e logística se relacionam por meio das funções $\gamma = \frac{1}{\sigma}$ e $\alpha = \exp(\mu)$ (COLOSIMO; GIOLO, 2006).

2.6.2 Modelo de regressão logística

Os modelos de regressão permitem a inclusão de covariáveis que podem influenciar a variável resposta T . Segundo Gomes (2007), a classe de modelos de locação-escala caracteriza-se pela distribuição da variável transformada $Y = \log(T)$ ter parâmetros de locação dependente das covariáveis e um parâmetro de escala, $\mu(x)$ e σ . Utilizando $\mu(x) = \mathbf{x}^T \boldsymbol{\beta}$, com $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$ sendo o vetor de parâmetros desconhecidos, é possível construir um modelo de locação-escala para Y da seguinte maneira.

$$Y = \log(T) = \mathbf{x}^T \boldsymbol{\beta} + \sigma v. \quad (2.6.7)$$

Considerando que Y segue uma distribuição logística e que ele é um modelo linear, pode-se escrever a sobrevivência para Y condicional a $\mathbf{x}$ da forma mostrada abaixo:

$$S(y|\mathbf{x}) = \frac{1}{1 + \exp \left\{ \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right\}}. \quad (2.6.8)$$

2.6.3 Cópula

Segundo Nelsen (1986), as cópulas são funções que permitem o relacionamento de funções de distribuições multivariadas com funções de distribuições marginais. O Teorema

de Sklar, (SKLAR, 1959), apresenta a definição de cópulas a seguir.

Teorema 1 *Seja F uma função de distribuição k -dimensional com marginais $F_1, \dots, F_k$. Então existe uma cópula C_α tal que para todo $(y_1, \dots, y_k)$ em $\mathbb{R}^k$,*

$$F(y_1, \dots, y_k) = C_\alpha(F_1(y_1), \dots, F_k(y_k)). \quad (2.6.9)$$

Se $F_1, \dots, F_k$ são todas contínuas então C_α é única, em que α é o parâmetro de associação. Caso contrário, C_α é determinada unicamente na imagem $F_1 \times \dots \times F_k$. Inversamente, se C_α é uma cópula e $F_1, \dots, F_k$ são funções de distribuição, então a função F definida acima é uma função de distribuição k -dimensional com marginais $F_1, \dots, F_k$ (NÚÑEZ, 2005).

Além disso, algumas propriedades importantes são:

- $C_\alpha(F_1(y_1), \dots, F_k(y_k))$ tem como condição que $F_j(Y_j) \sim U(0, 1) \forall j$;
- As variáveis aleatórias $Y_1, \dots, Y_k$ são independentes se, e somente se, $C_\alpha(F_1(y_1), \dots, F_k(y_k)) = \prod_{j=1}^k F_j(Y_j)$.

Existe uma grande variedade de famílias de cópulas. Nesse trabalho, aplicou-se a cópula de Clayton com duas variáveis respostas, $k = 2$. Essa cópula pertence a família de cópulas Arquimedianas em que há possibilidade de obter facilmente a medida de associação τ de Kendall (FACHINI, 2011).

A cópula de Clayton, também conhecida como modelo de fragilidade gama, possui a seguinte função de sobrevivência conjunta:

$$S(y_1, y_2) = [S(y_1)^{-\alpha} + S(y_2)^{-\alpha} - 1]^{-\frac{1}{\alpha}}. \quad (2.6.10)$$

Note que, $S(y_1)$ e $S(y_2)$ são as funções de sobrevivência marginais para as variáveis aleatórias Y_1 e Y_2 , e, quando $\alpha \rightarrow 0$ tem-se que $S(y_1, y_2) = S(y_1)S(y_2)$, indicando que as variáveis são independentes. O coeficiente τ de Kendall para essa família se relaciona com o parâmetro α da seguinte forma:

$$\tau_\alpha(Y_1, Y_2) = \frac{\alpha}{\alpha + 2}. \quad (2.6.11)$$

2.6.4 Modelo de regressão bivariado

Os modelos no formato da equação 2.6.7 permitem o estudo do tempo até a ocorrência de apenas um evento de interesse, mas nesse estudo são considerados dois eventos de interesse. Assim, para modelos bivariados é possível reescrever a equação 2.6.7 como:

$$Y_k = \mathbf{x}^T \boldsymbol{\beta}_k + \sigma_k Z_k, \quad (2.6.12)$$

sendo que $Y_k = \log(T_k)$ é o logaritmo do tempo de falha do k -ésimo evento de interesse e possui distribuição logística com parâmetros de localização e escala $\mu_k(x) = \mathbf{x}^T \boldsymbol{\beta}_k$ e $\sigma_k > 0$, respectivamente. Nesse contexto, $\mathbf{x}$ é o vetor de covariáveis, $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \dots, \beta_{pk})$ é o vetor de coeficientes desconhecidos e Z_k é o erro aleatório (FACHINI, 2011).

O modelo definido na equação 2.6.12 caracteriza-se como um modelo de regressão linear para Y_k , sendo que a função de sobrevivência é representada por $S(y_k|\mathbf{x})$ e a função de densidade é dada por $f(y_k|\mathbf{x})$. Assim, a função de sobrevivência para o modelo de regressão bivariado dada pela cópula de Clayton é:

$$S(y_1, y_2|\mathbf{x}) = [S(y_1|\mathbf{x})^{-\alpha} + S(y_2|\mathbf{x})^{-\alpha} - 1]^{-\frac{1}{\alpha}}. \quad (2.6.13)$$

Note que quando $\alpha \rightarrow 0$, as variáveis Y_1 e Y_2 serão independentes e o modelo volta a ser um modelo de regressão linear univariado.

Na equação 2.6.13, as funções de sobrevivência marginais $S(y_1|\mathbf{x})$ e $S(y_2|\mathbf{x})$ podem ser quaisquer função de sobrevivência, não necessariamente iguais. Neste trabalho, ambas serão consideradas funções de sobrevivência da distribuição logística.

Inferência do modelo bivariado

Assim como nos demais modelos de regressão citados anteriormente, para o modelo bivariado também é utilizado o método de máxima verossimilhança, com uma pequena modificação do método usual. Por conta das duas variáveis respostas, a função de verossimilhança é composta pelas seguintes combinações de observações censuradas:

- Falha nos dois eventos de interesse;
- Falha no evento 1 e censura no evento 2;

- Censura no evento 1 e falha no evento 2;
- Censura em ambos eventos.

Considerando uma amostra composta por $(y_{1k}, \delta_{1k}, \mathbf{x}_{1k}), \dots, (y_{nk}, \delta_{nk}, \mathbf{x}_{nk})$, em que y_{ik} é o logaritmo do tempo, δ_{ik} é o indicador de censura e $\mathbf{x}_{ik}$ representa o vetor de covariáveis do k -ésimo evento de interesse do i -ésimo indivíduo, sendo $k = 1, 2$ e $i = 1, \dots, n$. Assim a função de verossimilhança é dada por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)^{\delta_{i1}\delta_{i2}} \times \left[\frac{-\partial S(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)}{\partial y_{i1}} \right]^{\delta_{i1}(1-\delta_{i2})} \times \left[\frac{-\partial S(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)}{\partial y_{i2}} \right]^{(1-\delta_{i1})\delta_{i2}} \times S(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)^{(1-\delta_{i1})(1-\delta_{i2})}, \quad (2.6.14)$$

em que $S(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)$ é a função de sobrevivência bivariada obtida por meio de uma cópula, $f(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2) = \frac{(-1)^2 \partial^2 S(y_{i1}, y_{i2} | \mathbf{x}_1, \mathbf{x}_2)}{\partial y_{i1} \partial y_{i2}}$ é a função de densidade conjunta de (y_{i1}, y_{i2}) , $\boldsymbol{\theta} = (\alpha; \boldsymbol{\theta}_1^T; \boldsymbol{\theta}_2^T)^T$ é o vetor de parâmetros desconhecidos, sendo que $\boldsymbol{\theta}_i^T = (\boldsymbol{\beta}_k^T, \sigma_k)^T$ e $\boldsymbol{\beta}_k^T = (\beta_{0k}, \beta_{1k}, \dots, \beta_{pk})$ para $k = 1, 2$ (FACHINI, 2011).

Os estimadores de máxima verossimilhança para o vetor de parâmetros $\boldsymbol{\theta} = (\alpha; \boldsymbol{\theta}_1^T; \boldsymbol{\theta}_2^T)^T$ são os valores de $\boldsymbol{\theta}$ que maximizam a equação 2.6.14, ou equivalentemente, o logaritmo dela ($l(\boldsymbol{\theta})$). Dessa forma, os estimadores são encontrados ao se resolver as equações abaixo:

$$U(\boldsymbol{\theta}) = \frac{\partial \log[L(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \quad (2.6.15)$$

3 Metodologia

3.1 Material

Os dados utilizados neste estudo foram obtidos por meio da raspagem de dados (web scraping) de diários oficiais municipais, com a metodologia desenvolvida pelo Grupo de Pesquisa do LAPCIPP/IPOL/UnB, em uma parceria do Departamento de Estatística (EST) e do Instituto de Ciência Política (IPOL) da Universidade de Brasília (UnB) com a Escola Nacional de Administração Pública (ENAP) (GONÇALVES, 2025).

A base de dados possui 11.130 observações e 49 variáveis, inclusas as de análise de sobrevivência (indicador de censura e tempo de falha). As variáveis explicativas, e as hipóteses iniciais para elas, foram obtidas através da tese de doutorado da Celina Pereira, "Da Política às Políticas: o que faz com que os programas federais cheguem à ponta?", da Universidade de Brasília apresentada em 2025.

As políticas presentes na base foram coletadas do banco do Instituto de Pesquisa Econômica Aplicada (IPEA) com nos seguintes critérios:

1. Políticas passíveis de adesão por qualquer município;
2. Criação da política entre 2000 e 2022;
3. Municípios com população a partir de 100 mil habitantes;
4. Taxa de adesão de pelo menos 5%;
5. Políticas com intervalos maiores de tempo de adoção.

O primeiro critério foi considerado, pois um dos interesses do estudo é utilizar a técnica estatística de análise de sobrevivência, para isso é necessário que todos os municípios tenham a possibilidade de aderir às políticas. O recorte temporal é feito para garantir uma base mais homogênea, pois os dados obtidos pela raspagem de dados para períodos anteriores possuíam muitas lacunas nas variáveis e muitas diferenças político-institucionais (PEREIRA, 2025). A seleção de municípios com pelo 100 mil habitantes foi feita também para manter todos os municípios observados mais homogêneos.

A taxa de adesão de 5%, no mínimo, por cada política vem da literatura de difusão de políticas públicas. (BOEHMKE, 2019). Por fim, o recorte de políticas com tempos de adoção maiores também foi feito com base na literatura difusionista, de forma

a evitar políticas adotadas em poucos anos que não trariam informação sobre a difusão, mas sobre a rápida disseminação das políticas.

Após todos os recortes feitos acima, realizou-se um último recorte referente às políticas da área de infraestrutura, a fim de compreender os fatores que influenciam na adoção das políticas dessa área.

Dessa forma, a base de dados apresentou 636 observações com a mesma quantia inicial de variáveis. As políticas que satisfazem os critérios acima e fazem parte da área de infraestrutura são: o Programa de Aceleração do Crescimento (PAC) criado em 2007 e o Programa de Parcerias e Investimento (PPI) criado em 2016.

As variáveis explicativas podem ser divididas em 4 dimensões analíticas: governabilidade, capacidades institucionais, inovação ministerial e complexidade das políticas. Segundo Pereira (2025), embora essas dimensões possam explicar o comportamento de um município para a adoção de uma política, também era importante uma dimensão que ajude a explicar a dinâmica eleitoral entre municípios e União.

Assim, a dimensão de governabilidade possui 11 variáveis, a capacidade institucional possui 10, a inovação ministerial possui apenas uma, a complexidade das políticas possui 6, a eleitoral possui 9 e existem 3 variáveis de controle. A Tabela 1 apresenta essas variáveis, assim como as hipóteses da literatura difusionista sobre a influência de cada uma das variáveis no risco de adesão às políticas públicas.

Tabela 1: Variáveis e hipóteses usadas no estudo

Dimensão	Variável	Hipótese
Governabilidade	Taxa de vetos	Diminui o risco
	Taxa de conflito em MPs	Diminui o risco
	Taxa de participação	Aumenta o risco
	Taxa de sucesso legislativo	Aumenta o risco
	Taxa de dominância	Aumenta o risco
	Número efetivo de partidos	Diminui o risco
	Tamanho proporcional da base	Aumenta o risco
	Taxa de disciplina da base	Aumenta o risco
	Taxa de participação ministerial	Aumenta o risco
	Taxa de aprovação do PR	Aumenta o risco
	Tempo de mandato do ministro	Aumenta o risco

Dimensão	Variável	Hipótese
Capacidade Institucional	Total de servidores por órgão	Aumenta o risco
	Total de servidores efetivos por órgão	Aumenta o risco
	Orçamento discricionário autorizado	Aumenta o risco
	Orçamento discricionário executado	Aumenta o risco
	Orçamento total autorizado	Aumenta o risco
	Orçamento total executado	Aumenta o risco
	Taxa de orçamento discricionário empenhado	Aumenta o risco
	Taxa de orçamento total empenhado	Aumenta o risco
	Orçamento de TI previsto	Aumenta o risco
	Orçamento de TI executado	Aumenta o risco
Índice de inovação	Índice de inovação ministerial	Aumenta o risco
Complexidade das políticas	Número de dispositivos de controle	Diminui o risco
	Taxa de dispositivos de controle	Diminui o risco
	Número de dispositivos de contrapartida	Diminui o risco
	Taxa de dispositivos de contrapartida	Diminui o risco
	Número de palavras	Aumenta o risco
	Número de dispositivos	Aumenta o risco
Eleitoral	Reeleição do prefeito	Diminui o risco
	Ano eleitoral municipal	Aumenta o risco
	Ano eleitoral nacional	Aumenta o risco
	Margem de vitória do prefeito	Aumenta o risco
	Ideologia do presidente	Diminui o risco
	Ideologia do governador	Diminui o risco
	Ideologia do prefeito	Diminui o risco
	Turno de eleição do prefeito	Aumenta o risco
	Alinhamento partidário do presidente e do governador	Aumenta o risco
Controle	PIB	Diminui o risco
	Hierarquia urbana do município	Aumenta o risco
	Porte populacional do município	Aumenta o risco

3.2 Método

Este estudo busca identificar e explicar fatores que influenciam a adoção de uma política pública por um determinado município. Para isso, são utilizadas técnicas de

análise de sobrevivência, na qual a variável resposta é o tempo até a ocorrência do evento de interesse, sendo esse tempo chamado de tempo de falha (COLOSIMO; GIOLO, 2006).

No contexto desse estudo, o evento de interesse é a adesão de uma política pública de infraestrutura por um município. Entretanto, como um ente federativo pode escolher aderir ou não aderir o programa, os dados tem a presença da censura, que ocorre quando o município escolhe pela não adesão da política proposta.

Como particularidade do evento de interesse desse trabalho, existe a possibilidade de um indivíduo experimentar o evento de interesse mais de uma vez, ou seja, um município pode aderir mais de uma política. Esse acontecimento é chamado de eventos múltiplos. Por conta disso, inicialmente, foi usado um modelo que considerasse a possibilidade de associação entre esses tempos.

Posteriormente, com as variáveis selecionadas no modelo de eventos múltiplos, para entender o comportamento dos tempos de cada uma das políticas, foram ajustados os modelos de regressão logístico bivariado e o univariado. Isso auxiliou no estudo da suposição de dependência entre os tempos de falha das políticas dentro de um mesmo município.

Dessa forma, inicialmente, será realizada a organização e o tratamento dos dados, assim como uma análise exploratória desses, por meio de metodologias de análise descritiva tradicionais. Dada a natureza dos dados, também será realizada uma análise descritiva das curvas de sobrevivência e das funções de taxa de falha, fazendo uso do estimador de Kaplan-Meier.

Após o estudo descritivo dos dados, ocorrerá a seleção de variáveis e os ajustes dos modelos. Por fim, o ajuste dos modelos serão avaliados e os resultados obtidos serão discutidos.

Para realização da análise descritiva, ajuste dos modelos e análise de resíduos, utilizou-se o software R na versão 4.3.1.

4 Resultados

4.1 Análise Descritiva

Nessa seção, são apresentadas as análises descritivas da variável resposta e das variáveis explicativas dos modelos ajustados.

Visão geral

Os dados para as políticas de Infraestrutura apresentam 329 falhas e 307 censuras, ou seja, aproximadamente 53% dos municípios adotaram uma das políticas públicas e os 47% restantes possuem observações censuradas.

Na Figura 1, é apresentada a curva de sobrevivência das políticas calculada por meio do estimador de Kaplan-Meier. Nela é possível notar que após o primeiro ano (365 dias), há uma queda brusca na probabilidade de sobrevivência, em que a probabilidade reduz de 97% para 64%, aproximadamente. Em seguida, existem algumas baixas na sobrevivência, sendo que duas se destacam. A primeira próxima do terceiro ano (1190 dias), em que a probabilidade baixa para cerca de 54%, a segunda por volta de 6 anos e meio (2300 dias), quando a probabilidade de adesão de uma política pública reduz para 48,3% e permanece assim até o último tempo observado, devido aos 307 municípios com observações censuradas.

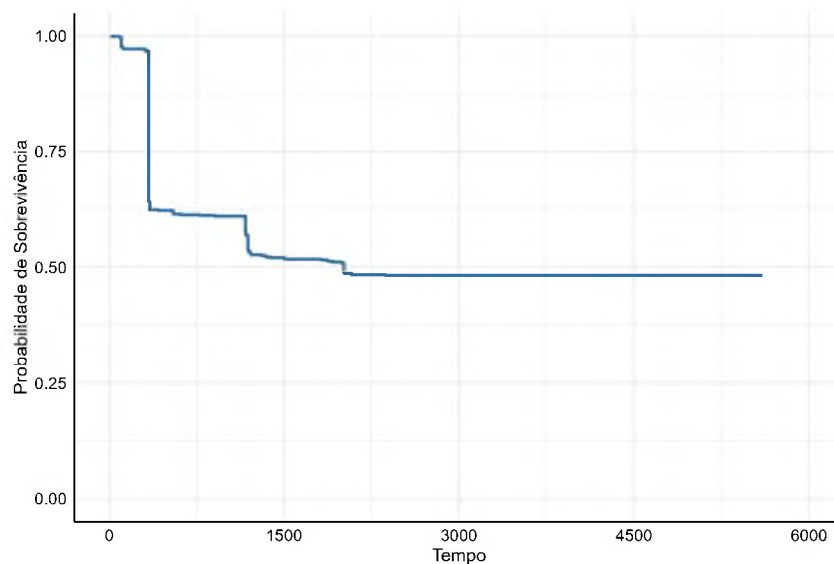


Figura 1: Função de Sobrevivência estimada via Kaplan-Meier para as políticas de Infraestrutura

Observando as políticas de forma individual, 94,97% dos municípios adotaram o Programa de Aceleração do Crescimento (PAC), enquanto o Programa de Parcerias e Investimento teve uma taxa de adesão de apenas 8,49% dos municípios. Isso se torna evidente na Figura 2, em que no último tempo observado a probabilidade de um ente federativo aderir a política PPI é 91,5%, resultado dos 291 municípios com observações censuradas. Além disso, para o política PAC, a probabilidade de um município adotar a política após 1 ano reduz drasticamente de 93% para 25%, e após 3 anos (1095 dias) essa probabilidade baixa para 7,86%.

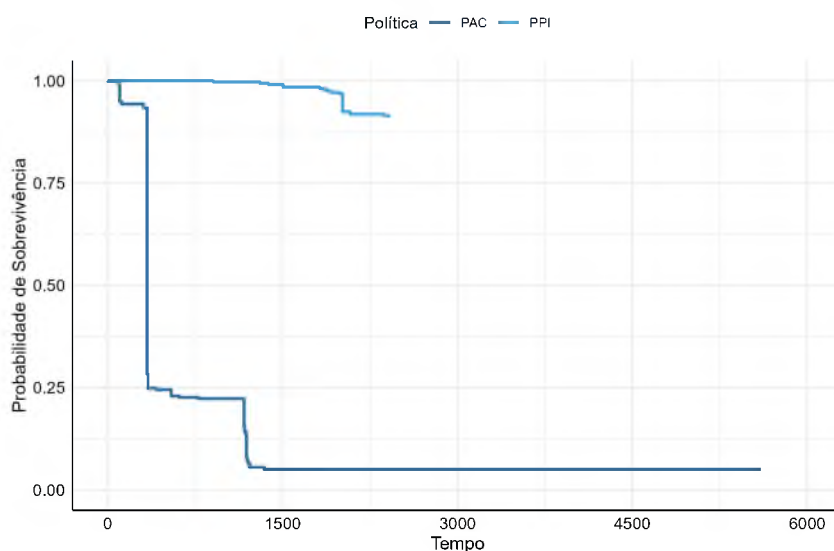


Figura 2: Função de Sobrevivência estimada via Kaplan-Meier por política pública

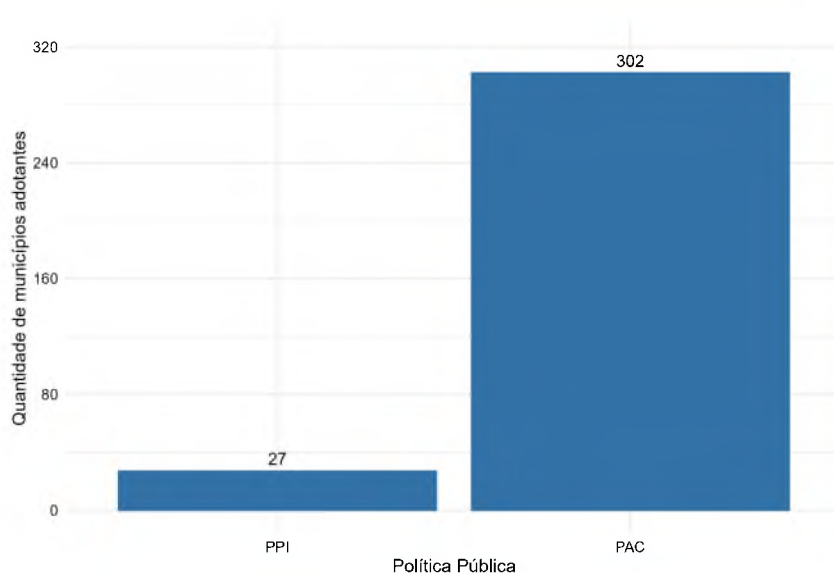


Figura 3: Gráfico de barras da quantidade de municípios adotantes

PIB per capita

O PIB per capita municipal foi uma variável de controle essencial durante a modelagem, principalmente por ele estar relacionado diretamente com a população e com a economia do município, trazendo informações sobre o desenvolvimento econômico e qualidade de vida deste.

Na Figura 4, percebe-se que os municípios com PIB mais baixo possuem mais adesão que municípios com maior renda.

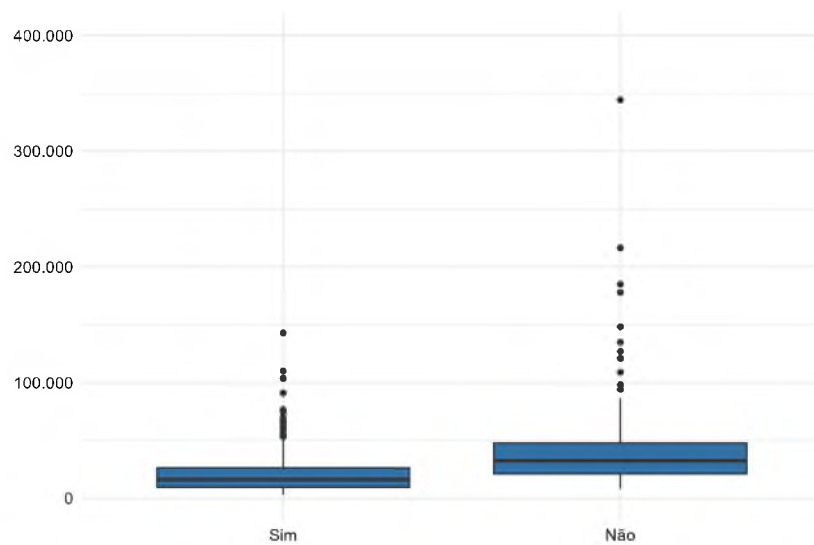


Figura 4: Boxplot do PIB per capita pela indicador de adesão das políticas

Explorando o comportamento dessa variável em cada uma das políticas, na Figura 5, observa-se que a política PAC aparenta ter mais adesão entre municípios de menor renda. Os *boxplots* da política PPI mostram que o PIB não parece influenciar na adesão dessa política, mesmo com a presença de alguns municípios não adotantes com renda elevada.

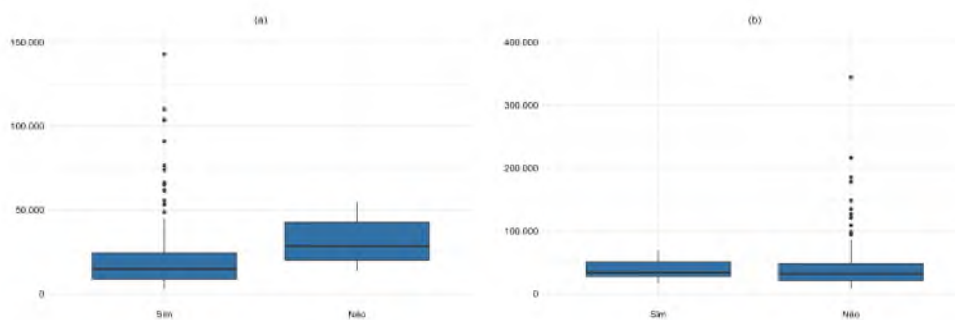


Figura 5: Boxplots do PIB per capita pela situação de adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI

Percentual de orçamento discricionário empenhado

O orçamento discricionário trata da parte do orçamento público em que a gestão do município tem liberdade para decidir onde e como aplicar. Alguns exemplos de despesas que compõem esse orçamento são investimentos em políticas públicas, serviços e infraestrutura. O percentual do orçamento discricionário empenhado diz quanto do orçamento autorizado para esses gastos está comprometido.

O *boxplot* da Figura 6 mostra que municípios que não aderiram às políticas de Infraestrutura comprometeram cerca de 94% de seu orçamento discricionário autorizado.

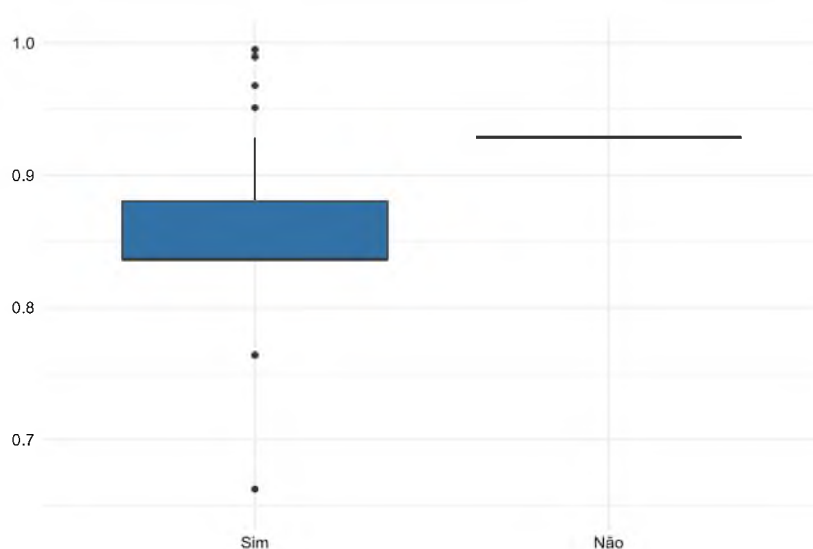


Figura 6: Boxplot da Taxa de orçamento discricionário empenhado pela situação da adesão das políticas

Na Figura 7, o gráfico (a) demonstra que os municípios que adotaram o programa PAC têm menor capacidade de execução orçamentária discricionária, enquanto o gráfico (b) mostra que os municípios adotantes do programa PPI possuem maior empenho de seus orçamentos.

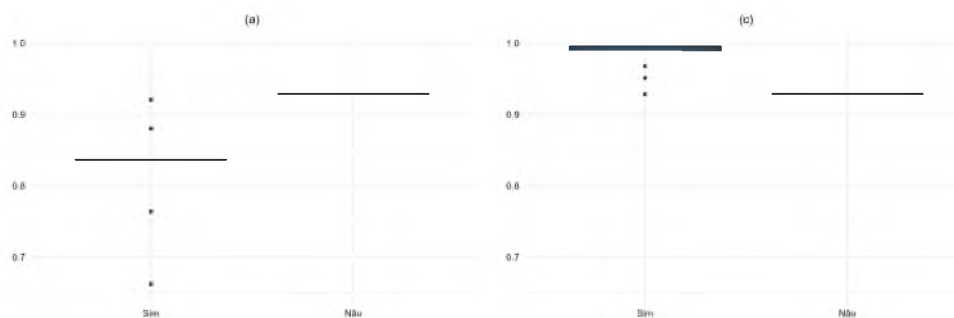


Figura 7: Boxplots da Taxa de orçamento discricionário empenhado pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI

Percentual de conflitos em Medidas Provisórias

As medidas provisórias (MPs) propostas pelo Executivo podem enfrentar três tipos de conflitos durante sua tramitação no congresso: i) rejeição, ii) perda do prazo de votação constitucional, implicando na perda de sua eficácia e iii) veto presidencial. Assim, o percentual de conflitos em Medidas Provisórias é o quantitativo de MPs que se enquadram nas situações descritas dentre o total de medidas enviadas.

A Figura 8 apresenta a visão geral dessa variável em relação as falhas e censuras dos dados. Nela é possível perceber que municípios que não aderiram aos programas de infraestrutura tem um percentual de conflito nas medidas provisórias menor do que os municípios adotantes.

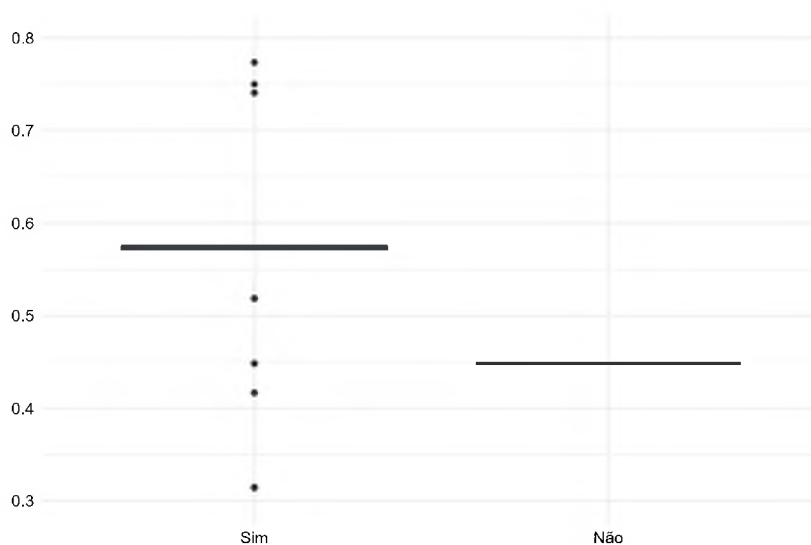


Figura 8: Boxplot do Percentual de conflitos em MPs pela situação da adesão das políticas

Analisando as políticas individualmente pela Figura 9, ambos os programas apre-

sentam o mesmo comportamento do cenário geral, mesmo com a política PAC apresentando alguns valores discrepantes. Dessa forma, é razoável dizer que os municípios adotantes das políticas de infraestrutura apresentam maior taxa de conflitos em MPs.

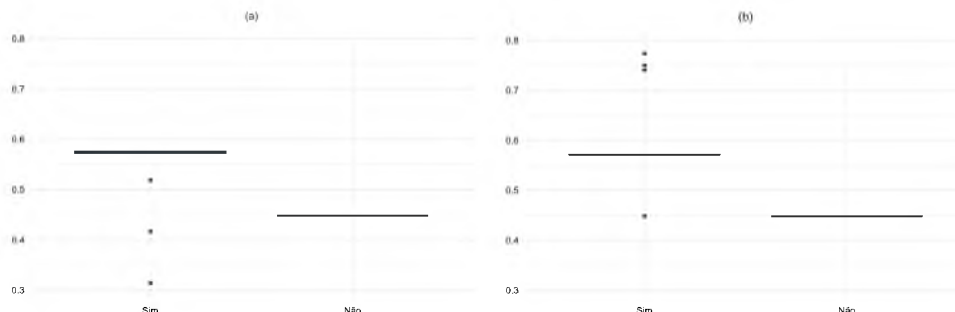


Figura 9: Boxplots do Percentual de conflitos em MPs pela situação de adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI

Ideologia do Prefeito

A ideologia partidária do prefeito pode variar de -1 (extrema esquerda) a 1 (extrema direita). No *boxplot* da Figura 10, percebe-se que a adesão de políticas em municípios com prefeitos que tendem para a direita não são comuns, já em municípios com prefeitos de centro-esquerda a adoção de políticas de infraestrutura é maior.

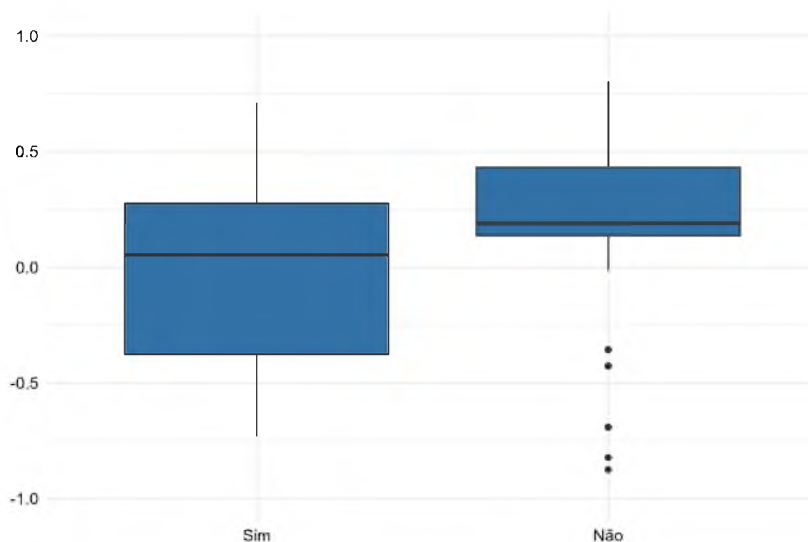


Figura 10: Boxplot da Ideologia do prefeito pela situação da adesão das políticas

Na Figura 11, observa-se para a política PAC o mesmo desfecho da visão geral, menos municípios com prefeitos de direita adotaram o programa. Já para a política PPI, não é possível notar uma grande diferença entre a distribuição dos municípios adotantes e

não adotantes, mostrando que ela pode ser uma política mais neutra entre os municípios.

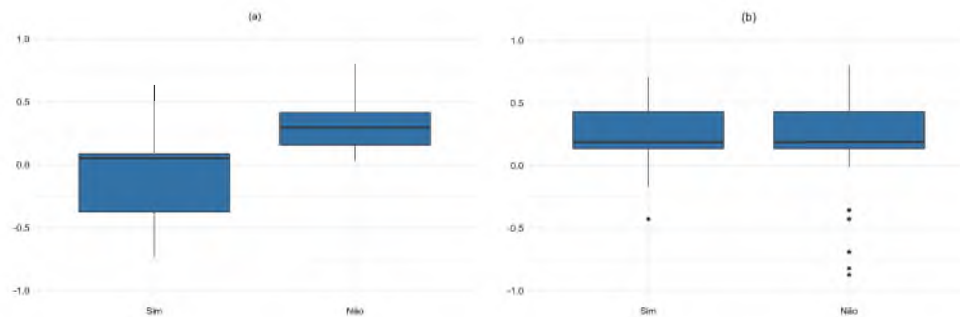


Figura 11: Boxplots da Ideologia do prefeito pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI

Tempo de Mandato dos Ministros

O tempo de mandato dos ministros é o tempo médio em meses que os ministros permaneceram no cargo em cada gestão. Na Figura 12, observa-se que em mandatos de ministros mais longos a adesão dos programas de infraestrutura é menor que em mandatos menores.

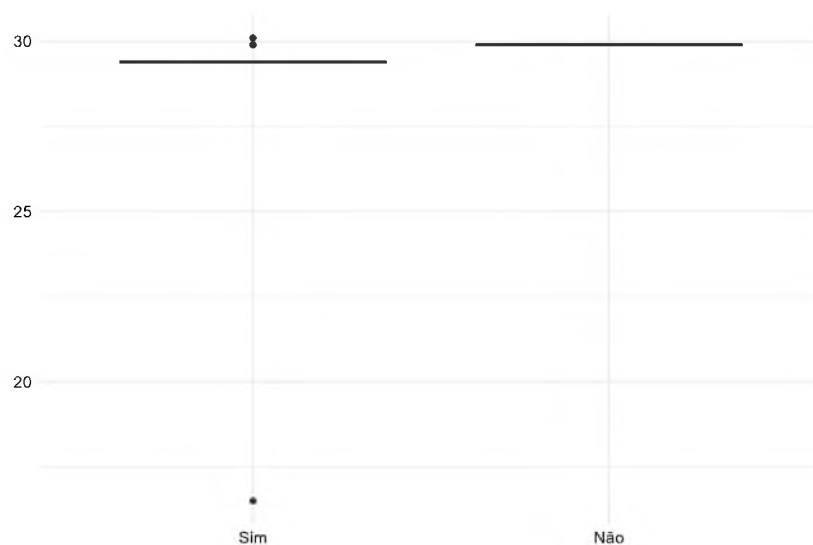


Figura 12: Boxplot do Tempo de mandato dos ministros pela situação da adesão das políticas

A Figura 13 apresenta a distribuição dessa variável para cada uma das políticas. Nela é possível notar o mesmo comportamento do cenário geral para o Programa de Aceleração do Crescimento. Por outro lado, municípios adotantes e não adotantes do Programa de Parcerias e Investimentos possuem o mesmo comportamento.

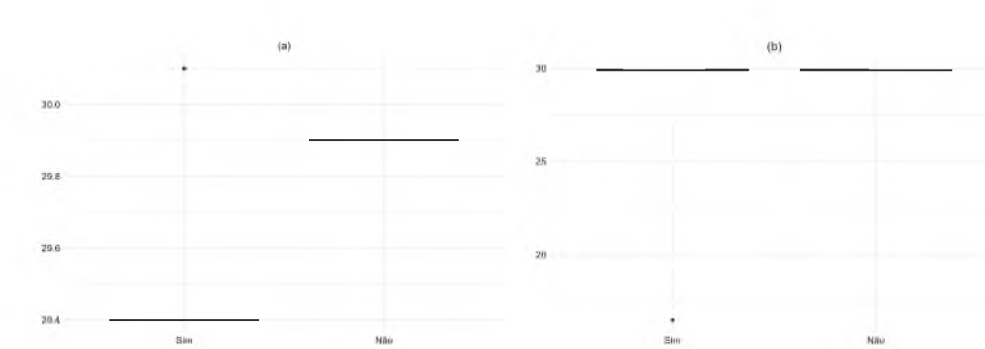


Figura 13: Boxplots do Tempo de mandato dos ministros pela situação da adoção das políticas PAC e PPI. (a) Boxplot da política PAC e (b) Boxplot da política PPI

Ano Eleitoral Nacional

A variável ano eleitoral nacional é categórica, sendo que 1 indica ano eleitoral e 0 ano não eleitoral. Ela apresenta um comportamento curioso, uma vez que a expectativa era que em anos eleitorais o risco de falha fosse maior. Entretanto, as Figuras 14 e 15 mostram que todas as observações para os anos não eleitorais falharam, enquanto em anos eleitorais existem muitas observações censuradas.

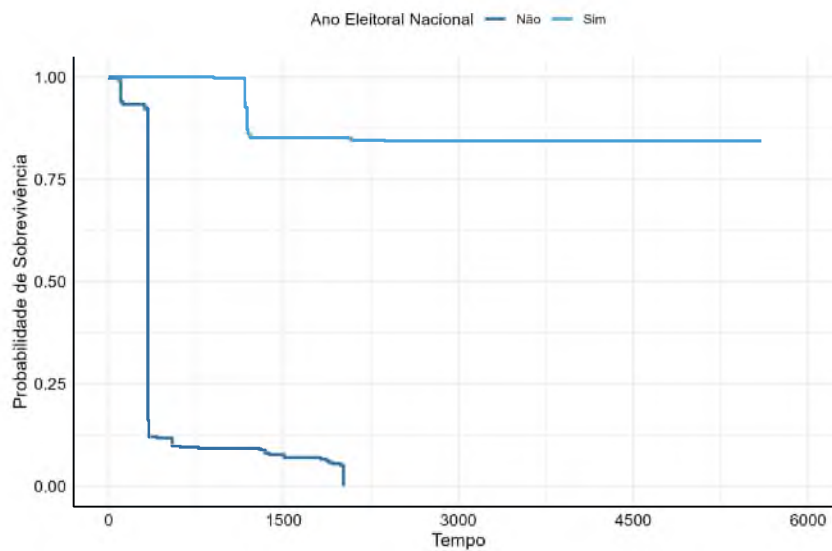


Figura 14: Curva de Kaplan-Meier por Ano Eleitoral Nacional

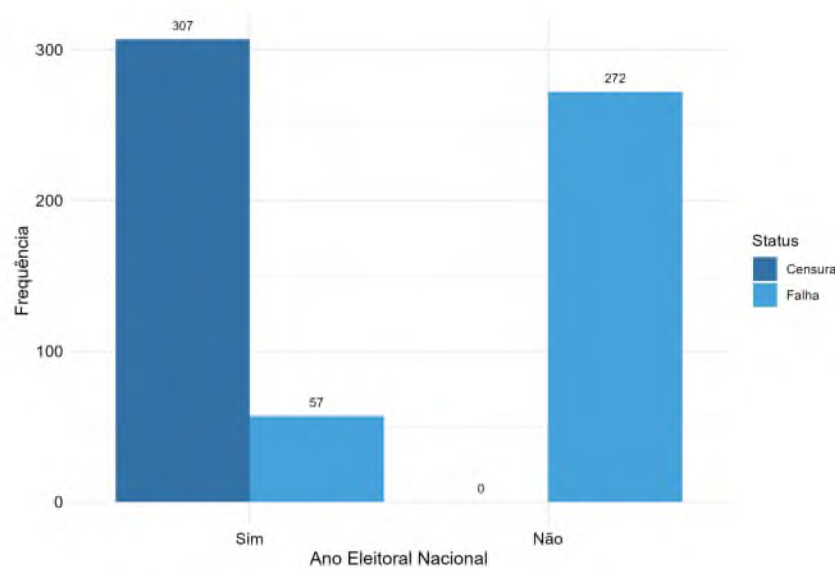


Figura 15: Gráfico de colunas do ano eleitoral nacional pela situação de adoção das políticas

As Figuras 16 e 17, mostram os gráficos da variável ano eleitoral nacional só para os dados do Programa de Aceleração do Crescimento. Percebe-se que o risco de um município falhar dentro de 1350 dias, é maior tanto em anos eleitorais quanto em anos não eleitorais. Destaca-se também a ausência de censuras em anos que não ocorrem eleição nacional.

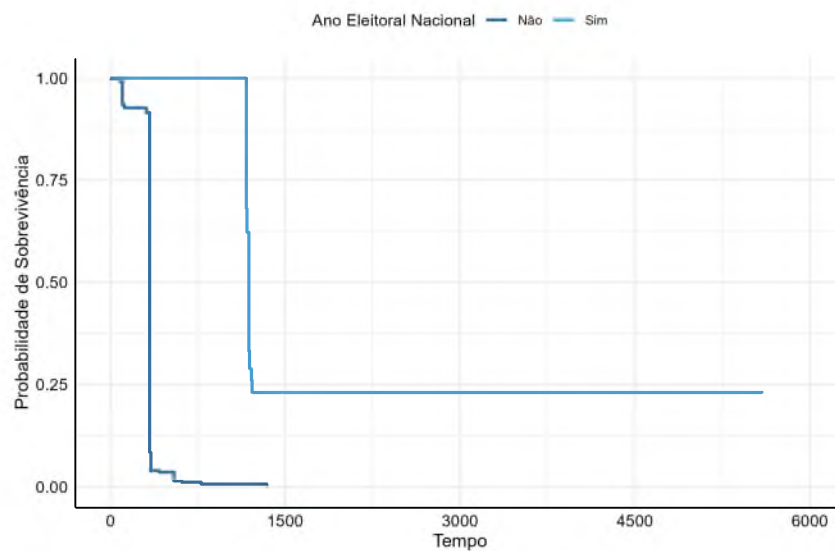


Figura 16: Curva de Kaplan-Meier por Ano Eleitoral Nacional para política PACA

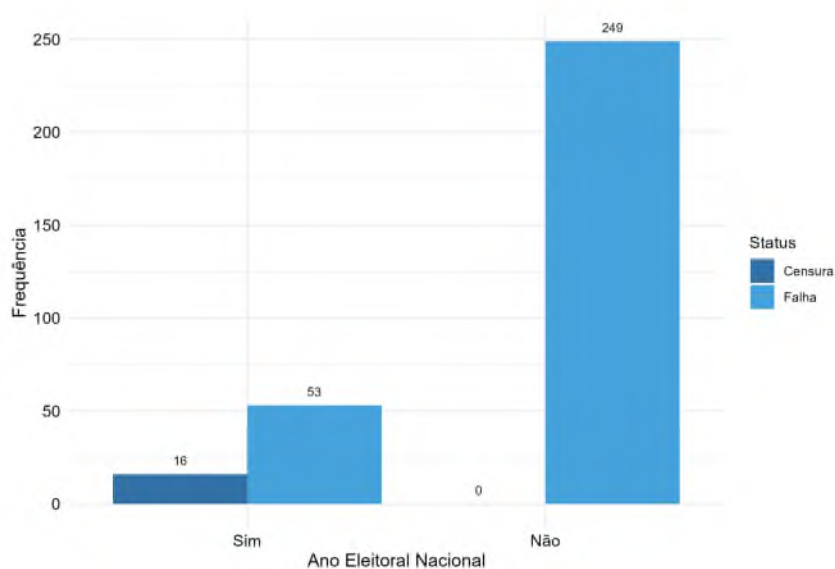


Figura 17: Gráfico de colunas do ano eleitoral nacional pela situação de adoção da política PAC

Os gráficos para o Programa de Parcerias e Investimentos são apresentados nas Figuras 18 e 19. Embora esse programa tenha menos adesão que o Programa de Aceleração do Crescimento, a quantidade de municípios adotantes é maior em anos não eleitorais. Além disso, todas as observações censuradas ocorrem em anos de eleição, mantendo a probabilidade de sobrevivência até o final do estudo próximo de 99%.

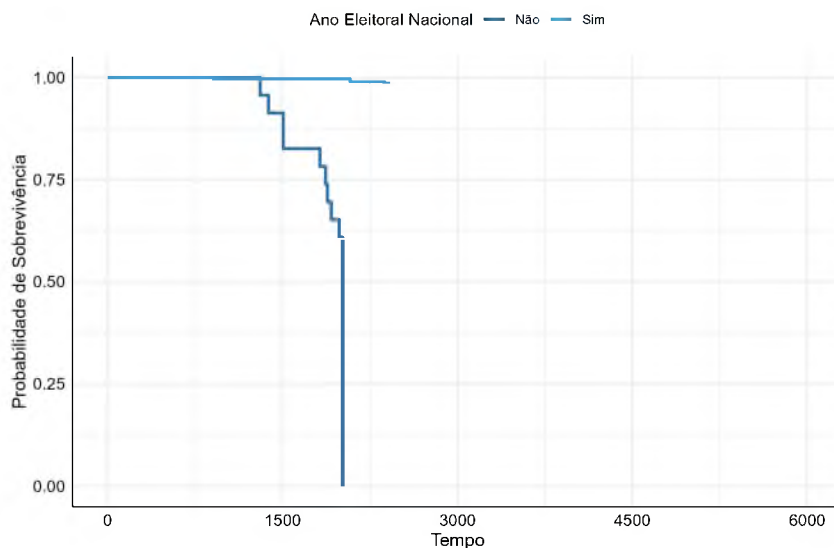


Figura 18: Curva de Kaplan-Meier por Ano Eleitoral Nacional para a política PPI

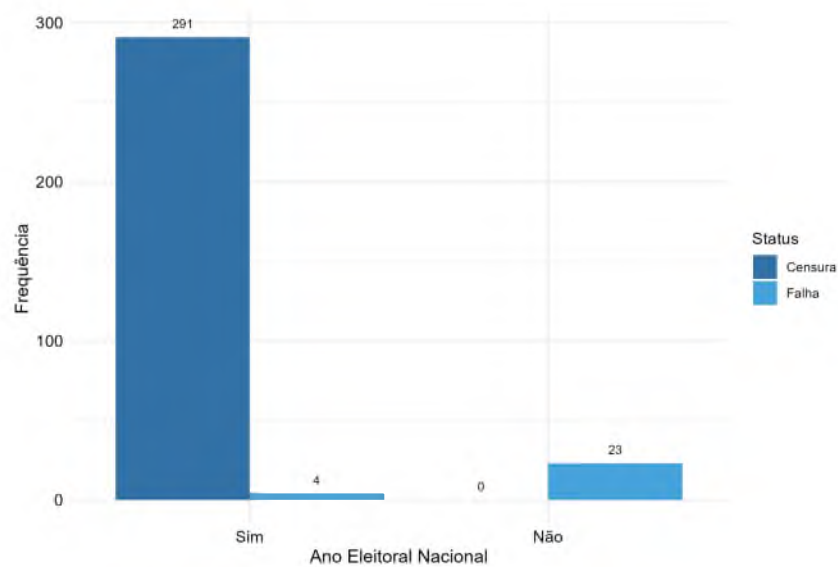


Figura 19: Gráfico de colunas do ano eleitoral nacional pela situação de adoção da política PPI

Ano Eleitoral Municipal

A variável ano eleitoral municipal possui as mesmas características da variável ano eleitoral nacional. Nas Figuras 20 e 21, observam-se resultados que atendem as suposições iniciais, sendo que todas as observações em anos eleitorais municipais apresentam falha dentro de 1500 dias. Em contrapartida, mesmo com os dados dos anos sem eleições apresentando algumas falhas, a maior parte de suas observações são censuradas.

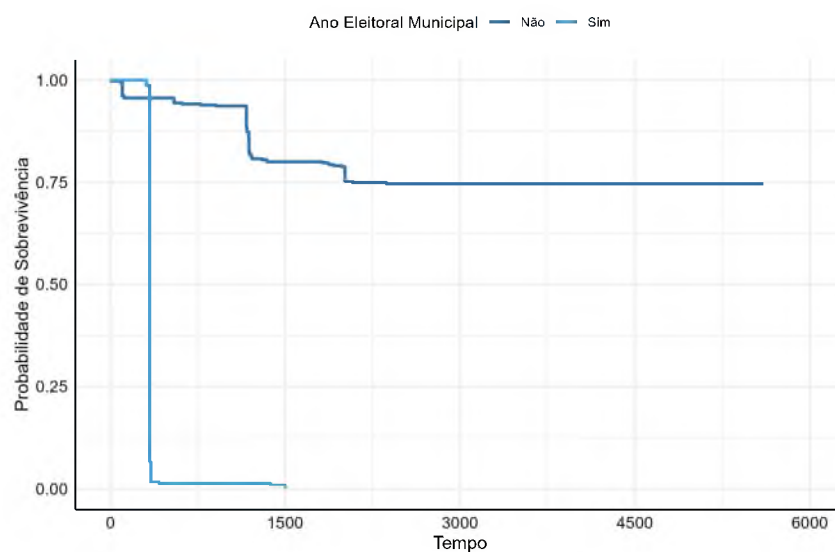


Figura 20: Curva de Kaplan-Meier por ano eleitoral municipal

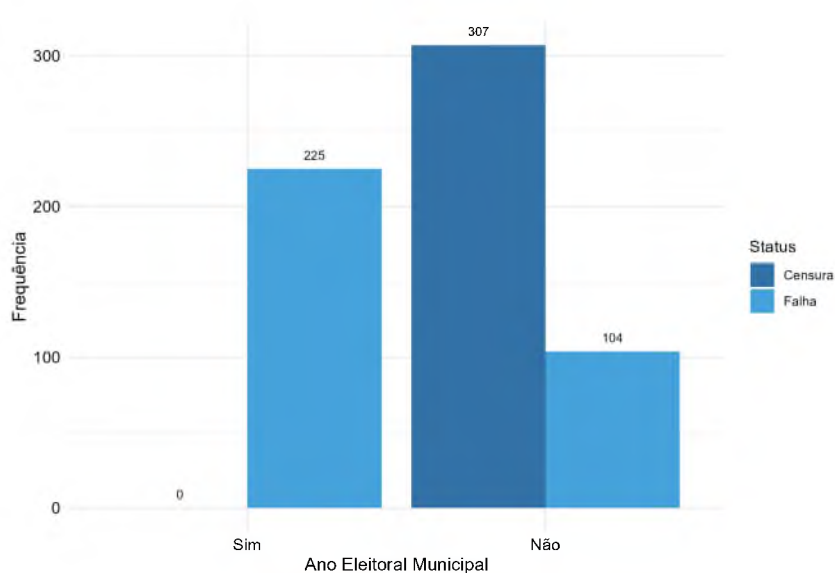


Figura 21: Gráfico de colunas do ano eleitoral municipal pela situação de adoção das políticas

As Figuras 22 e 23 apresentam os gráficos descritivos para o Programa de Aceleração do Crescimento. Em ano eleitoral municipal, as observações de anos eleitorais municipais apresentaram falha dentro de 418 dias. Já nos anos não eleitorais, o risco de um município aderir ao programa é maior dentro de 1350 dias aproximadamente, após esse período a probabilidade da falha ocorrer reduz para 16,7%.

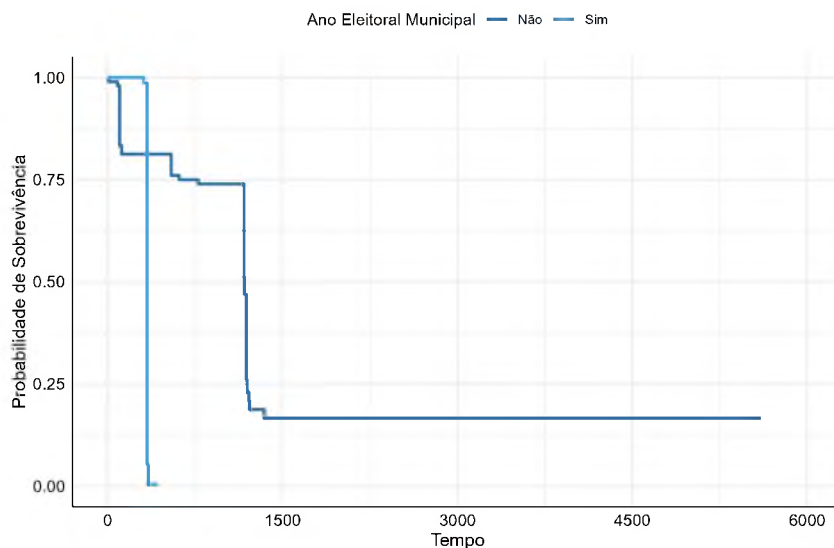


Figura 22: Curva de Kaplan-Meier por ano eleitoral municipal para a política PAC

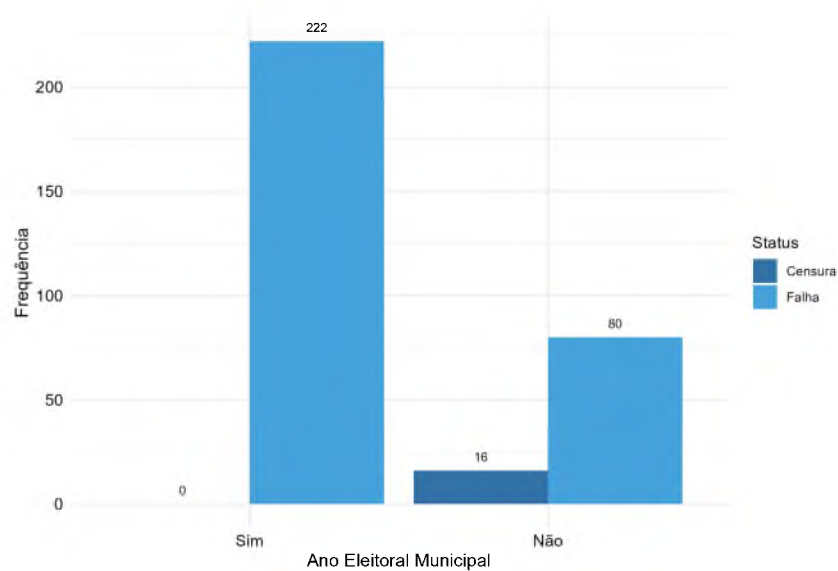


Figura 23: Gráfico de colunas do ano eleitoral municipal pela situação de adoção da política PAC

Já os gráficos para o Programa de Parceiros e Investimentos são apresentados nas Figuras 24 e 25. Em anos eleitorais existem apenas 3 observações, sendo que todas elas falharam dentro de 1500 dias. Enquanto para anos eleitorais, as observações sejam predominantemente censuradas e o risco da falha ocorrer é maior antes de 2015 dias.

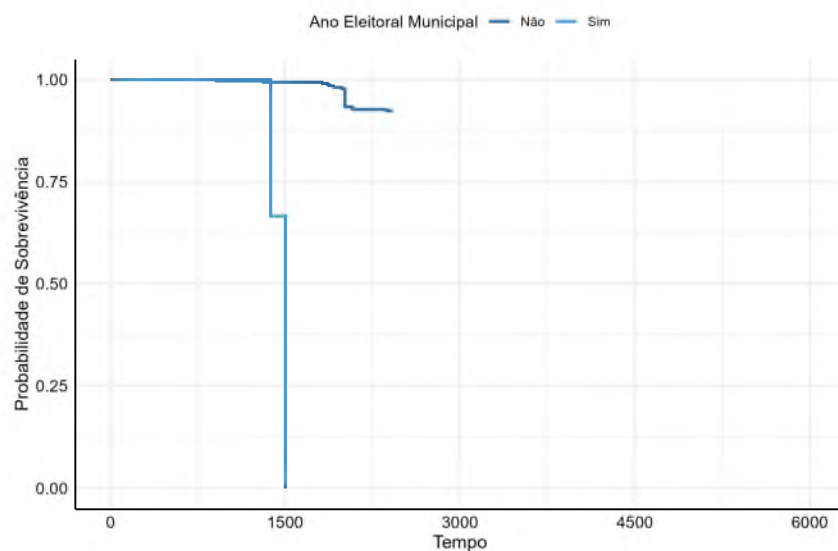


Figura 24: Curva de Kaplan-Meier por ano eleitoral municipal para a política PPI

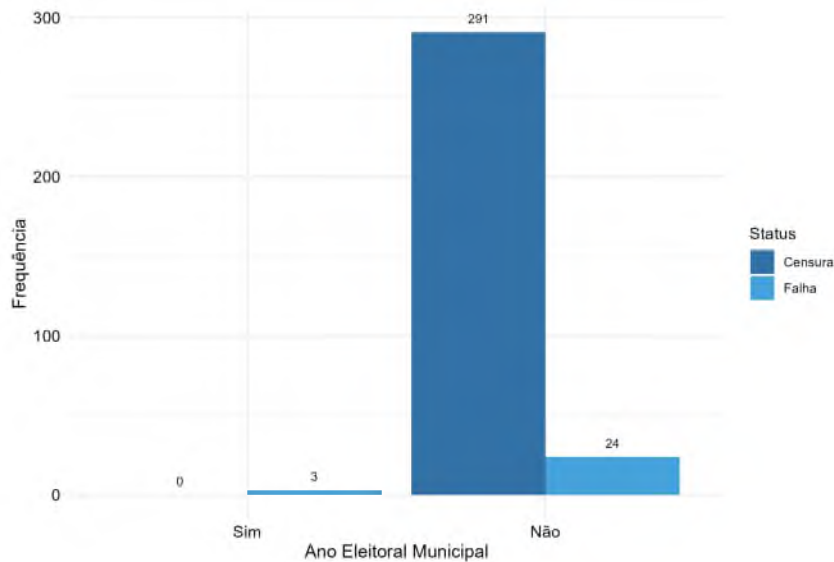


Figura 25: Gráfico de colunas do ano eleitoral municipal pela situação de adoção da política PPI

4.2 Modelagem

O primeiro modelo ajustado ao conjunto de dados foi o de Fragilidade Compartilhada, pois esse trabalho teve início com a suposição da possibilidade de dependência entre os tempos de falha dos municípios. Em seguida, para comparação e investigação da associação entre os tempos das políticas, foi ajustado um modelo bivariado via cópula de Clayton com distribuição Logística e um modelo univariado para cada uma das políticas, ambos com distribuição Logística.

4.2.1 Modelo de Fragilidade Compartilhada

Primeiramente, foi realizada a seleção de variáveis que consistiu das seguintes etapas:

1. Análise de correlação das variáveis;
2. Ajuste do modelo unitário para cada uma das variáveis;
3. Inserção das variáveis que, no modelo unitário, tiveram p-valor inferior a 10%, com exceção daquelas que possuíam correlação superior a 60% com covariáveis já inseridas;
4. Exclusão daquelas que ao serem incluídas no modelo deixaram de ser significativas.

Na análise de correlação, observou-se que a maioria das variáveis estavam associadas entre si. Ainda assim, todas as variáveis foram ajustadas individualmente e, para a etapa seguinte, foram consideradas aquelas que apresentaram o menor p-valor entre as variáveis correlacionadas. Dessa forma, as variáveis utilizadas na terceira etapa foram:

- PIB per capita municipal;
- Percentual de orçamento discricionário empenhado
- Percentual de conflitos em MPs;
- Taxa de disciplina da base;
- Tempo de mandato do ministro;
- Turno de eleição do prefeito;
- Prefeito reeleito;
- Ano eleitoral municipal;
- Ano eleitoral nacional;
- Alinhamento partidário presidente e governador;
- Ideologia do governador;
- Ideologia do prefeito.

Com essas potenciais variáveis, foram realizadas a terceira e quarta etapa. O resultado das covariáveis do modelo é apresentado na Tabela 2.

Tabela 2: Resultado do modelo de Fragilidade Compartilhada

Covariáveis	Coefficiente	Erro-Padrão	Valor-p
PIB	-0,1928	0,1115	0,084
Percentual de orçamento discricionário empenhado	-33,5758	2,0962	<0,001
Percentual de conflitos em Medidas Provisórias	14,4280	1,9013	<0,001
Ideologia partidária do prefeito	-0,6655	0,2000	0,00088
Ano eleitoral nacional	-3,5271	0,2766	<0,001
Ano eleitoral municipal	-1,3333	0,3550	0,00017
Fragilidade			<0,001

O modelo final é composto por 6 covariáveis, sendo que *Percentual de orçamento discricionário empenhado* pertence à dimensão de Capacidade Institucional, *Percentual*

de conflitos em Medidas Provisórias pertence à dimensão Governabilidade, *PIB* é uma covariável de Controle e as demais covariáveis pertencem à dimensão Eleitoral.

No âmbito da capacidade institucional, a taxa de orçamento discricionário empenhado demonstra reduzir a probabilidade de adesão às políticas públicas. Esse resultado mostra que conforme esse percentual aumenta, menor é o risco do município adotar as políticas.

A divergência entre a hipótese inicial e o resultado obtido também é observado na taxa de conflito em MPs, visto que o efeito esperado era a redução do risco. Pode-se supor que em cenários de vulnerabilidade política, a adoção de políticas de infraestrutura aumente.

Na dimensão de controle, o PIB apresentou resultados condizentes com as suposições iniciais, sendo que o aumento do PIB per capita do município reduz a probabilidade da adoção de um programa de infraestrutura em cerca de 17% ($1 - \exp^{-0,1928}$). Uma possível explicação para essa medida é que municípios com renda elevada tendem a depender menos dos recursos federais para melhorias de infraestrutura e, dependendo das circunstâncias, a execução de obras possa não ser relevante para eles.

Quanto às variáveis da dimensão eleitoral, o efeito observado na variável ideologia partidária do prefeito condiz com as expectativas iniciais. É razoável supor que durante mandatos de políticos mais alinhados à esquerda a prioridade seja a adesão de programas mais sociais, como saúde, educação e assistência social. Já no cenário contrário existe incentivo às privatizações, aos investimentos em obras e ao setor produtivo.

Seguindo pelas variáveis do âmbito eleitoral, pode-se dizer que em um ano eleitoral nacional a redução no risco de adesão a uma política é de 97% ($1 - \exp^{-3,5271}$) e em um ano eleitoral municipal o risco é reduzido em 74% ($1 - \exp^{-1,3333}$), aproximadamente. Esses resultados diferem do esperado, o que pode ser explicado pelos gestores municipais preferirem ações que trazem efeitos no curto prazo e mais visibilidade junto ao eleitorado. Além disso, eles podem apresentar certa cautela ao se comprometerem com programas que demandem maior carga administrativa e orçamentária, frente as incertezas dos resultados eleitorais.

Por se tratar do ajuste de um modelo de fragilidade compartilhada, existe ainda a variável de fragilidade. No contexto desse estudo, ela pode ser compreendida como as características individuais dos municípios que não foi possível identificar através dos dados coletados, mas que influenciam no comportamento deles em relação à adesão das políticas. Ela auxilia na explicação do motivo de alguns municípios não adotarem determinado

programa, mesmo tendo condições parecidas a de municípios adotantes.

No modelo ajustado, considerou-se que a fragilidade possui distribuição gaussiana, com média igual a 1 e variância estimada em 0,7748. O teste da fragilidade revelou-se significativo ($p\text{-valor} < 0,0001$), indicando associação entre os tempos de falha de um mesmo município.

A Figura 26 ajuda na avaliação da fragilidade compartilhada, ela apresenta as fragilidades estimadas dos municípios. Valores próximos de 1 indicam que a fragilidade não altera significativamente o risco, enquanto valores maiores que 1 indicam aumento do risco e menores que 1 indicam redução do risco.

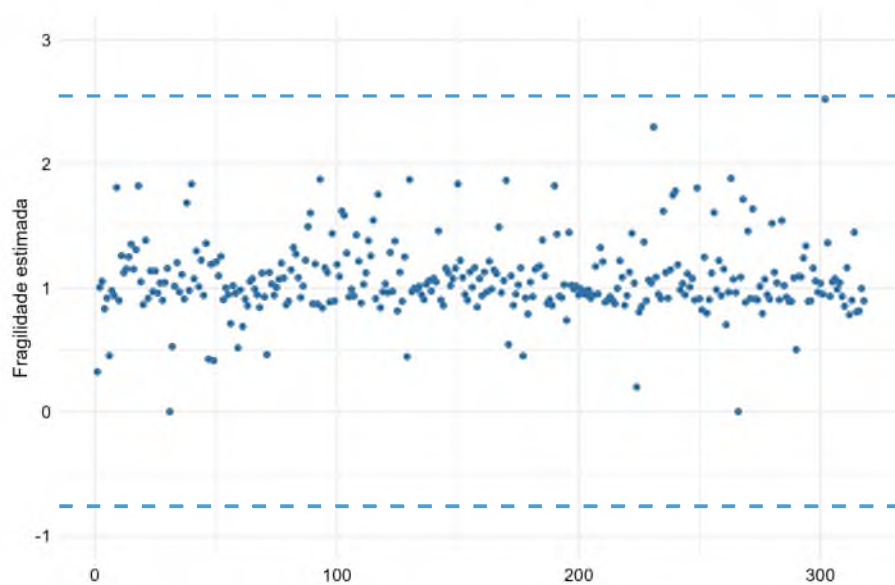


Figura 26: Gráfico da fragilidade compartilhada por município

No gráfico, observa-se que nenhum município teve fragilidade fora do padrão ($1 \pm 2\sigma_{\text{fragilidade}}$), representado pelas linhas tracejadas. Entretanto, os municípios de Tangará da Serra/MT e Sorocaba/SP se destacam por apresentar fragilidade superior a 2, indicando que eles têm risco significativamente maior de aderir às políticas de infraestrutura. Outros municípios que se destacam são Codó/MA e Chapecó/SC, ambos possuem fragilidade muito próxima do 0, demonstrando que eles possuem um risco significativamente menor na adesão de programas de infraestrutura.

Por fim, é necessário avaliar o ajuste global do modelo, para garantir a confiabilidade das estimativas do modelo. Para isso, optou-se pela análise dos resíduos de Cox-Snell, que segundo Lawless (2003), quando o modelo for adequado eles irão seguir uma distribuição exponencial padrão.

Embora as variáveis explicativas do modelo sejam estatisticamente significativas, o ajuste dele não se aproxima como esperado da curva da distribuição exponencial, como é possível observar na Figura 27. Por esse motivo, foram ajustados mais dois modelos para realização de comparação entre um modelo em que há a suposição de dependência entre os tempos dos municípios e entre modelos em que não existe essa suposição.

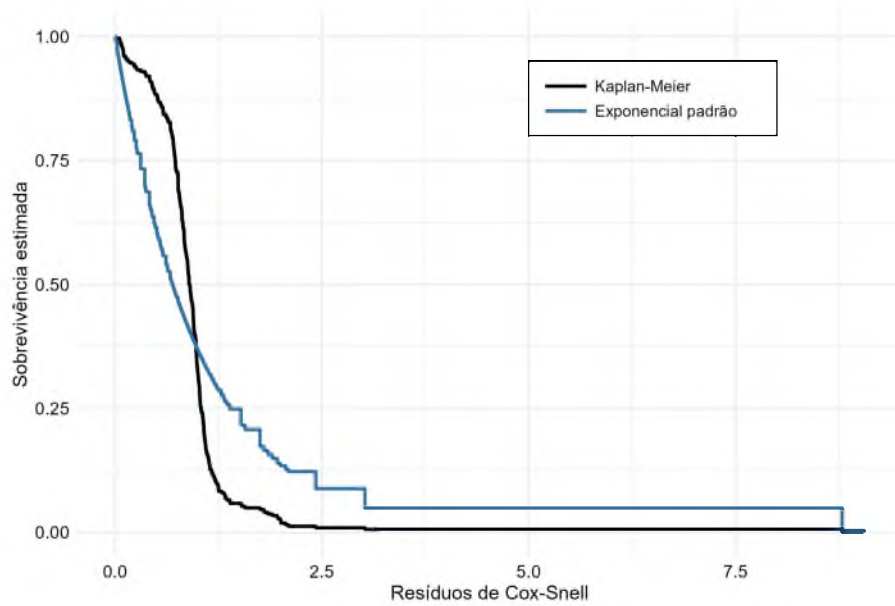


Figura 27: Resíduo de Cox-Snell do modelo de fragilidade compartilhada

4.2.2 Modelos paramétricos

Para o ajuste dos modelos paramétricos, considerou-se que a variável aleatória T segue uma distribuição Log-Logística, com parâmetros α e γ , e que $Y = \log(T)$ segue uma distribuição Logística, com parâmetros $\mu = \log(\alpha)$ e $\sigma = 1/\gamma$. As variáveis explicativas consideradas nesses modelos foram as mesmas utilizadas no modelo final de fragilidade compartilhada.

Modelo de Regressão Bivariado Logístico

Um modelo de regressão bivariado foi ajustado por meio da cópula de Clayton, sendo que a adoção da política PAC e a adoção da política PPI foram os dois eventos de interesse considerados. Além disso, o método de otimização utilizado para o cálculo das estimativas de máxima verossimilhança foi o *Simulated Annealing*. Esse é um método que percorre todo o espaço de soluções, aceitando soluções piores temporariamente para evitar mínimos locais, e a medida que a probabilidade de aceitação diminui, ele converge

para uma solução final.

A Tabela 3 apresenta os resultados desse modelo.

Tabela 3: Resultado do modelo de regressão bivariado através da cópula de Clayton para as políticas de Infraestrutura

Covariáveis	PAC			PPI		
	Coeficiente	Erro-Padrão	Valor-p	Coeficiente	Erro-Padrão	Valor-p
Intercepto	-545,5046	1339,2828	0,6838	272,1516	88,0040	0,0020
PIB	0,1283	0,4348	0,7680	-17,7770	0,8166	<0,0001
Percentual de orçamento discrecionário empenhado	879,0363	3112,8080	0,7776	-32,4137	86,7876	0,7088
Percentual de conflito das medidas provisórias	-388,9536	3305,2334	0,9063	-52,6470	35,0201	0,1328
Ideologia do prefeito	-0,2700	0,5211	0,6044	-6,4587	6,0816	0,2882
Ano eleitoral nacional	0,3402	487,7238	0,9994	4,0811	3,4340	0,2347
Ano eleitoral municipal	38,9632	636,5864	0,9512	-5,5544	8,5467	0,5158
σ	2,2838	0,0006	<0,0001	7,0088	0,9334	<0,0001

O primeiro ponto de destaque dos resultados, são os p-valores das variáveis para o programa PAC, sendo que nenhuma delas foi significativa para o modelo. Ainda assim, as variáveis para o programa PPI também apresentaram esse comportamento, exceto a variável de controle PIB.

Isso pode ter ocorrido devido às variáveis terem sido escolhidas para um modelo de fragilidade compartilhada e devido a indicação de não associação entre os tempos de falha, uma vez que o τ de Kendall calculado antes do ajuste do modelo foi igual a 0,0105.

Após a estimação dos efeitos das covariáveis, o α teve uma estimativa de 0,3224, resultando em um $\tau_{calculado}$ igual a 0,1388, confirmando a indicação de não associação entre os tempos das políticas em estudo.

Dessa forma, a Figura 28 mostra a curva de Kaplan-Meier estimada para a política PAC e o ajuste da função de sobrevivência marginal do modelo. Nela observa-se que nos y_i iniciais tem um bom ajuste do modelo, entretanto, para os valores finais, o ajuste se distancia do esperado.

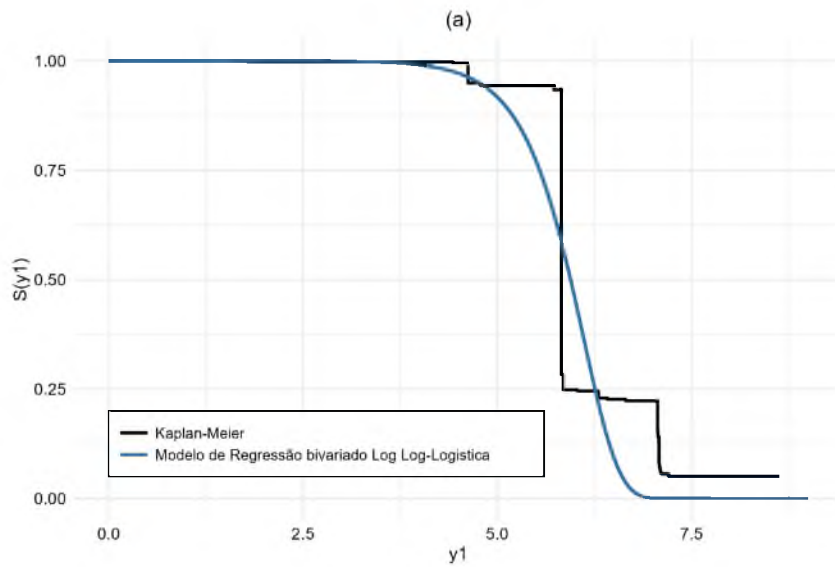


Figura 28: Curva de Kaplan-Meier estimada marginalmente para o Programa de Aceleração do Crescimento - PAC

Já na Figura 29, vendo o gráfico (a.1) o ajuste da função de sobrevivência marginal do modelo aparenta ser adequado, mesmo com a presença de dados censurados. Ao ampliar o gráfico, (a.2), o modelo apresenta, de fato, uma boa adequação aos dados, indicando que com outras variáveis pode ser que se chegue a um modelo mais adequado para explicar a difusão dessa política.

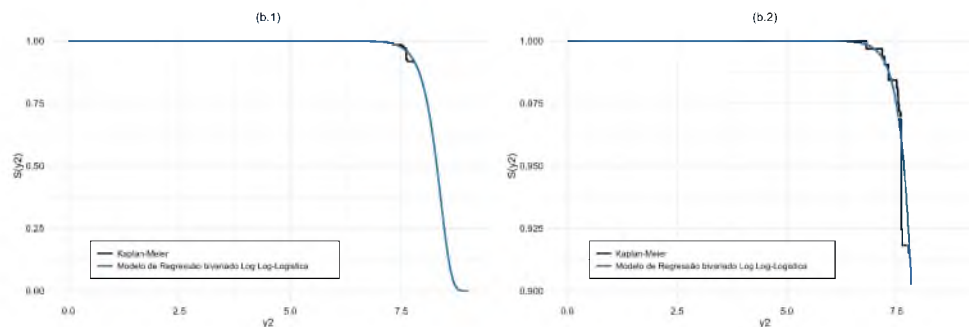


Figura 29: Curva de Kaplan-Meier estimada marginalmente para o Programa de Parcerias e Investimentos - PPI. (a.1) Curva estimada completa e (a.2) Curva estimada com visualização em $0.9 \leq S(y_1) \leq 1$

Modelos de Regressão Univariados Logísticos

A princípio, foram ajustados modelos para cada uma das políticas com as variáveis selecionadas previamente no modelo de fragilidade compartilhada e será denotado "Primeira seleção de variáveis" nas Tabelas 4 e 5. Entretanto, como os resultados obtidos não foram satisfatórios, optou-se por realizar uma nova seleção de variáveis considerando o

modelo de regressão univariado logístico. A Tabela 4 apresenta o resultado dos modelos para a política PAC.

Tabela 4: Resultados dos modelos de regressão univariado logístico para o Programa de Aceleração do Crescimento

Modelo	Covariáveis	Coefficiente	Erro Padrão	Valor-p
Primeira seleção de variáveis	Intercepto	-589,3981	<0,0001	<0,0001
	PIB	-2,6992	<0,0001	<0,0001
	Percentual de orçamento discricionário empenhado	887,4526	<0,0001	<0,0001
	Percentual de conflitos em medidas provisórias	-334,1428	<0,0001	<0,0001
	Ideologia do prefeito	1,8584	<0,0001	<0,0001
	Ano eleitoral municipal	5,159	<0,0001	<0,0001
	Ano eleitoral nacional	64,9334	<0,0001	<0,0001
Segunda seleção de variáveis	Intercepto	-4,6510	0,4127	<0,0001
	PIB	0,0028	0,0111	0,80
	Percentual de orçamento discricionário empenhado	12,0746	0,5093	<0,0001
	Ano eleitoral municipal	0,3528	0,0389	<0,0001
	Ano eleitoral nacional	1,1259	0,0573	<0,0001

Para o primeiro modelo, todas as variáveis são significativas e apenas a ideologia do prefeito apresenta o sentido esperado de acordo com a hipótese inicial. Seu coeficiente positivo indica quanto mais de extrema direita a ideologia do prefeito for menor é a chance do município adotar ao programa, pois coeficientes positivos nesse modelo indicam menor probabilidade de falha. Outro importante destaque nesse modelo são os coeficientes do intercepto, percentual de orçamento discricionário empenhado e percentual de conflitos em MPs, que além do sentido invertido das suposições iniciais também possuem grande magnitude.

O modelo com a nova seleção de variável é composto pelas variáveis PIB, Percentual de orçamento discricionário empenhado, ano eleitoral municipal e nacional. Embora a variável PIB não seja significativa, ela é mantida por ser uma variável de controle. Além disso, apenas ela apresentou um coeficiente que condiz com o esperado.

No novo modelo ajustado, todos os coeficientes possuem uma magnitude menor, facilitando sua interpretação. Por exemplo, no primeiro modelo ajustado a taxa de orçamento discricionário empenhado possui um coeficiente de 887,4526 com sua interpretação de $e^{887,4526}$ sendo um valor muito grande para poder ser interpretado. Já no segundo modelo, a mesma variável apresenta o coeficiente igual a 12,0746, indicando que o aumento de 1 ponto percentual dessa variável pode reduzir o tempo de adesão em pelo menos 175 mil dias aproximadamente ($e^{12,0746}$).

Para comparar a adequação de ambos os modelos, foi utilizada a análise de

resíduos de Cox-Snell, assim como no modelo de fragilidade compartilhada. A Figura 30 apresenta essas análises.

No gráfico (a), observa-se que o modelo não teve um ajuste como o esperado, mesmo com todas as covariáveis sendo estatisticamente significativas. Já o gráfico (b) apresenta um ajuste mais satisfatório e mais próximo do esperado, indicando que entre os dois modelos, o segundo é mais robusto para explicar o tempo até a adesão do Programa de Aceleração do Crescimento.

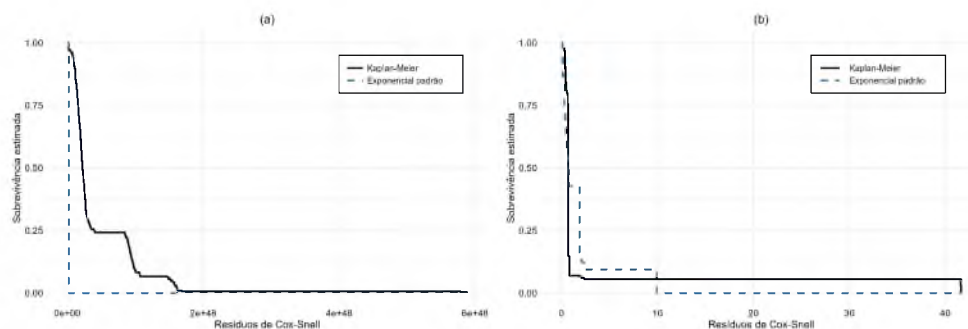


Figura 30: Resíduos de Cox-Snell para os modelos de regressão univariados da política PAC. (a) Resíduos da política PAC com a primeira seleção de variáveis (b) Resíduos da política PAC com a segunda seleção de variáveis

Em seguida, na Tabela 5 são apresentados os resultados dos modelos ajustados da política PPI.

Tabela 5: Resultados dos modelos de regressão univariado logístico para o Programa de Parcerias e Investimento

Modelo	Covariáveis	Coefficiente	Erro Padrão	Valor-p
Primeira seleção de variáveis	Intercepto	14,2409	1,1357	<0,0001
	PIB	-0,0164	0,0148	0,27
	Percentual de orçamento discricionário empenhado	-4,7702	1,0943	<0,0001
	Percentual de conflitos em medidas provisórias	-3,0283	0,1081	<0,0001
	Ideologia do prefeito	0,0104	0,0276	0,71
	Ano eleitoral municipal	-0,3806	0,0661	<0,0001
	Ano eleitoral nacional	0,1835	0,0325	<0,0001
Segunda seleção de variáveis	Intercepto	7,92927	0,2909	<0,0001
	PIB	-0,0238	0,0159	0,135
	Percentual de conflitos em medidas provisórias	-1,9712	0,1668	<0,0001
	Tempo de mandato do ministro	0,0348	0,0053	<0,0001
	Ano eleitoral nacional	0,0768	0,0365	0,035

O modelo ajustado com as variáveis selecionadas no modelo de fragilidade compartilhada apresentou as covariáveis PIB e Ideologia do prefeito com efeitos não significativos. Enquanto isso, o percentual de conflitos em MPs possui influência positiva sobre a

probabilidade de adesão ao programa e o ano eleitoral nacional possui influência negativa, ambos com sentidos diferentes das suposições iniciais.

O segundo modelo ajustado apresenta todas as covariáveis, com exceção do PIB, com efeito diferente do esperado inicialmente. Diferentemente dos modelos ajustados anteriormente, esse modelo possui como covariável o tempo de mandato do ministro, sendo que a cada mês que o ministro permanece em seu cargo o tempo até a adesão da política PPI reduz em pelo menos 1 dia ($e^{0,0348}$).

A Figura 31 apresenta a análise de resíduos de ambos os modelos ajustados para a política PPI. Embora nenhum dos modelos tenham um ajuste totalmente satisfatório, a curva da sobrevivência estimada dos resíduos do segundo modelo se aproximam mais da curva de sobrevivência da exponencial padrão, indicando que dentre os dois modelos, ele tem um ajuste melhor para o Programa de Parcerias e Investimentos.

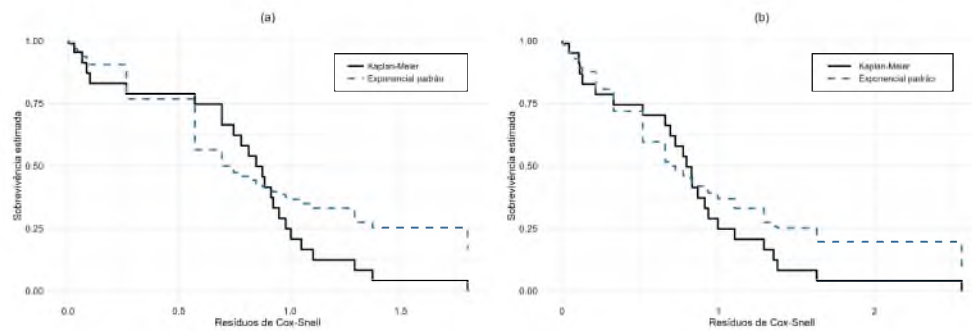


Figura 31: Resíduos de Cox-Snell para os modelos de regressão univariados da política PAC. (a) Resíduos da política PPI com a primeira seleção de variáveis (b) Resíduos da política PPI com a segunda seleção de variáveis

5 Conclusão

No início desse estudo, com a possibilidade de um município sofrer mais de uma vez o evento de interesse, existia a suposição de dependência entre os tempos de falha das políticas de infraestrutura para um mesmo município. Com isso, o modelo de fragilidade compartilhada era um modelo potencial para explicar a difusão dessas políticas.

Em seu ajuste, os efeitos das covariáveis selecionadas foram significativos, sendo que o percentual de conflitos em medidas provisórias tinha influência positiva sobre a adesão das políticas, enquanto o percentual de orçamento discricionário empenhado, a ideologia partidária, o PIB e anos eleitorais municipais e nacionais reduzem a probabilidade de um município adotar uma das políticas. O teste de fragilidade também foi significativo, indicando que realmente poderia haver dependência entre os tempos de falha de um mesmo município.

Entretanto, a curva de sobrevivência estimada dos resíduos de Cox-Snell não se aproximou da curva da exponencial padrão como esperado. Dessa forma, buscou-se um método alternativo que também pudesse levar em conta a possível dependência entre os tempos, sendo ele o modelo de regressão bivariado via cópulas.

Antes de começar o ajuste do modelo de regressão bivariado, foi calculado o τ de Kendall e seu resultado foi de 1,05%, indicando que não existe relação entre os tempos de falha do Programa de Aceleração do Crescimento e o Programa de Parcerias e Investimentos.

Mesmo com esse resultado inicial, seguiu-se com o ajuste do modelo e as estimativas obtidas não foram significativas estatisticamente, possivelmente pelas variáveis utilizadas terem sido selecionadas para um modelo de fragilidade compartilhada. Para avaliação do ajuste do modelo, comparou-se as curvas de Kaplan-Meier com as sobrevivências marginais do modelo bivariado para cada política, sendo possível concluir que o modelo não apresentava um ajuste satisfatório para explicar a difusão dos programas estudados.

Como foi observada a indicação de não associação entre os tempos das políticas PAC e PPI, ajustou-se um modelo de regressão logístico univariado para cada política considerando as mesmas variáveis selecionadas no modelo de fragilidade compartilhada para possível comparação dos resultados.

O modelo de regressão logístico para o Programa de Aceleração do Crescimento apresentou variáveis com efeitos significativos. Já o modelo do Programa de Parcerias e

Investimentos teve resultados que apontaram que as variáveis PIB e Ideologia do Prefeito não são influentes na adoção dessa política.

Para a avaliação do ajuste do modelo, também utilizou-se os resíduos de Cox-Snell. Nessa análise de resíduos, foi possível observar que as curvas empíricas não se aproximaram das curvas teóricas como esperado, indicando que ambos os modelos não são indicados para explicar a adesão às políticas.

Com isso, realizou-se uma nova seleção de variáveis para os modelos de regressão logístico univariado, em que mesmo com as variáveis apresentando efeitos diferentes das suposições iniciais, os modelos apresentaram bons resultados.

Pela análise de resíduos de Cox-Snell, observou que nesses últimos modelos ajustados, os resultados obtidos foram mais satisfatórios que os observados anteriormente.

Além disso, observou-se que em todos os modelos ajustados nesse trabalho tiveram o Ano Eleitoral Nacional com efeito significativo. Assim, pode-se supor que anos de eleição nacional tem grande influência sobre a adesão de políticas públicas de infraestrutura, seja reduzindo ou aumentando a probabilidade da adoção dos programas.

Portanto, esse trabalho teve como objetivo estudar os fatores que influenciam a adoção das políticas de infraestrutura pelos municípios brasileiros. Para isso, uma série de modelos foram ajustados para entender o comportamento dos tempos de adoção das políticas.

O modelo de fragilidade compartilhada e o modelo de regressão bivariado via cópulas promoveram um estudo considerando a associação entre os tempos das políticas. A comparação de ambos permitiu concluir que os tempos de adesão do Programa de Aceleração do Crescimento e do Programa de Parcerias e Investimentos não estão associados. Com isso, a suposição inicial do estudo, que considerava uma possível dependência entre os tempos observados, foi rejeitada ocasionando no ajuste dos modelos univariados para entender o comportamento do tempo e das covariáveis em cada uma das políticas.

A comparação dos quatro modelos ajustados permitiu concluir que para a área de infraestrutura é possível analisar a adesão das políticas de forma individual. Por isso, como continuidade deste trabalho, recomenda-se uma nova seleção de dados para ajuste dos modelos paramétricos, a fim de encontrar um modelo mais robusto, assim como testar a inclusão de uma nova variável controle, sendo ela a região dos municípios (Norte, Sul, Nordeste, Sudeste e Centro-Oeste). Além disso, ressalta-se a importância de estudar mais profundamente as dimensões analíticas que compõem as políticas e sua difusão.

Referências

- BERRY, F. S. *Tax Policy Innovation in the American States*. [S.l.]: University of Minnesota, 1987.
- BERRY, F. S.; BERRY, W. D. State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review*, v. 84, n. 2, p. 395–415, 1990. Disponível em: <https://doi.org/10.2307/1963526>.
- BOEHMKE, F. e. a. Spid: A new database for inferring public policy innovativeness and diffusion networks. *Policy Studies Journal*, v. 0, n. 0, 2019.
- BOX-STEFFENSMEIER, J. M.; JONES, B. S. Time is of the essence: Event history models in political science. *American Journal of Political Science*, v. 41, n. 4, p. 1414–1461, 1997.
- BRESLOW, N. Discussion of professor cox's paper. *Journal of the Royal Statistical Society B*, v. 34, p. 216–217, 1972.
- COÊLHO, D. B.; CAVALCANTE, P.; TURGEON, M. Mecanismos de difusão de políticas sociais no brasil: uma análise do programa saúde da família. *Revista de Sociologia e Política*, v. 24, n. 58, p. 145–165, 2016.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. [S.l.]: Editora Blucher, 2006.
- COX, D. R. Regression models and life tables. *Journal of the Royal Statistical Society, Oxford University Express*, v. 34, n. 2, p. 187–220, 1972.
- COX, D. R. Partial likelihood. *Biometrika*, v. 62, p. 269–276, 1975.
- COX, D. R.; OAKES, D. *Analysis of Survival data*. [S.l.]: Chapman and Hall/CRC, 2018.
- FACHINI, J. B. *Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência*. Tese (Doutorado) — Universidade de São Paulo, 2011.
- GOMES, E. M. de C. *Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas*. Dissertação (Mestrado) — Universidade de São Paulo - Escola Superior de Agricultura "Luiz Queiroz", 2007.
- GONÇALVES, M. S. *Estudo da adesão de políticas públicas da área de saúde por municípios brasileiros via Modelo de Fragilidade Compartilhada* — Universidade de Brasília, Brasília, 2025. Trabalho de Conclusão de Curso de Estatística (Bacharelado).
- HOUGAARD, P. *Analysis of Multivariate Survival Data*. [S.l.]: Springer-Verlag, 2000.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Taylor Francis, Ltd., v. 53, n. 282, p. 457–481, 1958.

LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. 2nd. ed. [S.l.]: John Wiley Sons, Inc, 2003.

MELLO, M. C. d. *Modelando dados de sobrevivência na Ciência Política: uma análise da Difusão da Agenda Urbana no Brasil*. Tese (Doutorado) — Universidade de Brasília, 2021. 94 f., il.

NELSEN, R. B. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics - Theory and Methods*, v. 15, p. 3277–3285, 1986.

NÚÑEZ, J. S. R. *Modelagem Bayesiana para dados de sobrevivência bivariados através de cópulas*. Tese (Doutorado) — Universidade de São Paulo, 2005.

PEREIRA, A. de Ávila Gomide e A. K. *Governança da política de infraestrutura: condicionantes institucionais ao investimento*. [S.l.]: Instituto de Pesquisa Econômica Aplicada, 2018.

PEREIRA, C. *DA POLÍTICA ÀS POLÍTICAS: o que faz com que os programas federais cheguem à ponta*. Tese (Doutorado) — Universidade de Brasília, 2025.

PETO, R. Contribuição à discussão do artido de d.r. cox. *Journal of the Royal Statistical Society B*, v. 34, p. 205–207, 1972.

ROGERS, E. M. *Diffusion of Innovations*. 3rd. ed. [S.l.]: Free Press, 1983.

SKLAR, A. Fonctions de répartition à n dimensios et leurs marges. *Publ. Inst. Statist. Univ. Paris*, v. 8, p. 229–231, 1959.

THERNEAU, P. M. G. T. M. *Modeling Survival Data: Extending the Cox Model*. [S.l.]: Springer-Verlag, 2000.