



Universidade de Brasília
Departamento de Estatística

**Estudo da Adesão do Programa Nacional de Melhoria do Acesso e da
Qualidade da Atenção Básica pelos Municípios Brasileiros via Modelo de
Regressão Log-Logístico para Dados Censurados**

Letícia Fonseca Souza

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

**Brasília
2024**

Letícia Fonseca Souza

**Estudo da Adesão do Programa Nacional de Melhoria do Acesso e da
Qualidade da Atenção Básica pelos Municípios Brasileiros via Modelo de
Regressão Log-Logístico para Dados Censurados**

Orientadora: Prof^ª. Dra. Juliana Betini Fachini Gomes

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

**Brasília
2024**

Sumário

1 Introdução	5
2 Referencial Teórico	7
2.1 Conceitos Básicos.	7
2.2 Função de probabilidade, de sobrevivência e de risco.	8
2.3 Estimador de Kaplan - Meier	11
2.4 Modelos de Probabilidade	12
2.4.1 Distribuição Weibull e Distribuição Gumbel	12
2.4.2 Distribuição Log-Normal	13
2.4.3 Distribuição Log- logística	13
2.5 Método de Máxima Verossimilhança	14
3 Metodologia	16
3.1 Material	16
3.2 Método	17
3.2.1 Modelo de regressão Log-Logístico	17
3.3 Resíduos de Cox-Snell	18
4 Análise e Resultados	19
4.1 Análise Descritiva	19
4.1.1 Unidade da Federação - UF	19
4.1.2 Porte	19
4.1.3 Adesão ao programa	21
4.1.4 Partido do prefeito	22
4.1.5 Margem de vitória	22
4.1.6 Reeleição	23
4.1.7 Ideologia	24
4.1.8 MetrÓpole	25
4.1.9 Tempo	26
4.2 Modelagem	26
4.2.1 Distribuição de probabilidade	26

4.2.2 Seleção de variáveis	29
5 Conclusão	32
Referências.	33

1 Introdução

A Estatística é uma Ciência que trata desde a coleta até a análise e interpretação de dados que pode influenciar na tomada de decisões. Esta ciência pode atuar em diferentes áreas de conhecimento como economia, medicina, esportes, psicologia, administração, ciência política, entre outros.

Uma esfera de conhecimento muito importante da estatística é o estudo em análise de sobrevivência, também conhecida como análise de sobrevida. Essa área tem como objetivo estudar os dados relacionados ao tempo até determinado fato acontecer. Na ciência política, por exemplo, vamos estudar o tempo até a adesão a uma determinada política. O estudo de sobrevivência em conjunto com a ciência política pode ajudar a identificar e avaliar as variáveis que levam os estados e municípios a aderirem a uma determinada política pública.

As políticas públicas são essenciais para o funcionamento eficaz de um país e para atender as necessidades de sua população. Além disso, as políticas públicas são fundamentais para promover a equidade e reduzir as desigualdades sociais. Elas podem fornecer serviços e recursos necessários para grupos vulneráveis, melhorando o acesso à saúde e à educação, por exemplo.

As políticas podem ajudar a gerenciar crises econômicas e assegurar a sustentabilidade a longo prazo. A previsão e a consistência nas políticas públicas são alguns dos fatores que podem atrair investimentos, criar empregos e manter a economia estável, beneficiando tanto o setor público quanto o privado.

A política pública escolhida para fazer esse estudo é o Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ-AB) pois influenciou no avanço e na melhora dos serviços básico de saúde para a população brasileira. O PMAQ-AB foi uma iniciativa do governo brasileiro lançado por meio da portaria n° 1.654 de 19 de julho de 2011 do Ministério da Saúde. O objetivo desse programa é melhorar a qualidade do serviço de atenção básica no Sistema Único de Saúde (SUS).

O PMAQ-AQ busca aumentar a acessibilidade dos serviços de atenção básica para a população e melhorar a qualidade dos serviços prestados e para isso, o programa oferece um incentivo financeiro que aumenta conforme os resultados. A adesão ao programa é voluntária, onde as equipes e unidades básicas de saúde se submetem a um processo de avaliação externa para que recebam incentivos financeiros conforme objetivos tenham sido alcançados.

O PMAQ-AB se desenvolve em ciclos, onde as equipes de saúde passam por diversas fases, como adesão, desenvolvimento de ações de melhoria, autoavaliação, avaliação

externa, análise dos resultados e recontractualização. Os indicadores avaliados incluem a estrutura das UBS, o processo de trabalho das equipes, o cuidado centrado no paciente, entre outros aspectos.

No dia 11 de novembro de 2019 o Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica foi extinto e substituído pelo programa Previne Brasil, instituído pela portaria nº 2.979 de 12 de novembro de 2019. O Previne Brasil alterou a forma de financiamento da atenção básica e introduziu novos critérios para o repasse de recursos.

O fim do PMAQ-AB teve um grande impacto na qualidade dos serviços básicos de saúde visto que o novo programa trouxe mudanças significativas na lógica de financiamento e na avaliação das equipes.

Será que variáveis relacionadas com a política e com características geográficas influenciam em aumentar ou diminuir o risco dos municípios brasileiros falharem? Com o objetivo de responder a essa pergunta o presente trabalho buscou algumas das variáveis no projeto de pesquisa “Estudos e Pesquisas para o aprimoramento das Capacidades Institucionais de Órgãos da Administração Pública para a elaboração, implementação e avaliação de políticas públicas e seus mecanismos de governança” da Escola Nacional de Administração (ENAP) em parceria com a Universidade de Brasília (UnB).

Para realizar uma análise aprofundada dos dados, optou-se pelo uso do software R. O objetivo é investigar as variáveis que podem influenciar o tempo de adesão, utilizando um modelo de regressão para dados censurados. Este trabalho está organizado em mais quatro seções principais: começando com o referencial teórico, seguido pela metodologia, análise dos resultados e, por fim, as conclusões.

2 Referencial Teórico

A análise de sobrevivência constitui uma área de estatística dedicada à investigação de dados vinculados ao período de tempo até a manifestação de um evento. Nesse contexto, designamos o intervalo transcorrido até a ocorrência do evento como "tempo de falha", atribuindo à variável resposta do estudo uma natureza longitudinal.

A principal característica em análise de sobrevivência é a presença de censuras, ou seja, observações incompletas, ao longo do estudo. A presença desse tipo de observação no estudo dificulta a utilização de técnicas estatística convencionais pois tais técnicas não consideram a presença dessas observações, criando, assim, um possível viés nas conclusões.

Com o intuito de entender sobre o processo de difusão do PMAQ-AB no Brasil, buscando identificar os fatores que influenciam na decisão e na velocidade da adoção ou não do programa será utilizado o conjunto de técnicas de análise de sobrevivência. Essas técnicas permitem incorporar no estudo as censuras, aumentando a confiabilidade das conclusões.

2.1 Conceitos Básicos

Ao iniciar o estudo utilizando a análise de sobrevivência é necessário entender alguns conceitos antes como (COLOSIMO; GIOLO,2006):

- **Falha:** é o evento de interesse que será estudado.
- **Tempo de falha:** é o intervalo de tempo decorrido desde o início da observação até o momento em que a falha ocorre. Para isso é necessário fixar o tempo inicial do estudo e a escala a ser usada.
- **Censura :** é quando o tempo de falha não é totalmente observado, ou seja, é quando a observação é incompleta, seja pelo fim do estudo, por não falhar ou qualquer outro motivo.

Existem três tipos de censura, a censura à direita do tipo I, do tipo II e aleatória, à esquerda e a intervalar.

- **Censura à direita:** acontece quando a observação é encerrada antes que a falha ocorra. Esse tipo de censura pode ser dividida em três categorias:
 - * **Tipo I:** o estudo tem um tempo pré-definido de duração e, os indivíduos que não experimentaram a falha até a conclusão do estudo, são censuras a direita deste tipo.

- * **Tipo II:** o estudo termina após uma quantidade pré-determinada de falhas.
- * **Aleatória:** ocorre quando um indivíduo é retirado do estudo sem ter experimentado a falha por motivos não controláveis.
- **Censura à esquerda:** é quando o evento de interesse já ocorreu quando o indivíduo foi observado.
- **Censura intervalar:** ocorre quando a falha não é observada com precisão, mas tem conhecimento do intervalo que o indivíduo experimentou o evento de interesse.
- **Variável resposta :** é dada pelo par (t, δ) , em que "t" é o tempo registrado, podendo ser tempo de falha ou censura, e δ é a variável indicadora que mostra se a observação é de falha ou de censura. Essa variável indicadora é dada por:

$$\delta_i = \begin{cases} 1, & \text{se a observação } i \text{ é de falha.} \\ 0, & \text{se a observação } i \text{ é de censura.} \end{cases}$$

- **Fração de cura :** esse termo é usado para se referir as observações não suscetíveis a falha, isto é, são entidades que não experimentam a falha em nenhum momento. São observações consideradas curadas.

2.2 Função de probabilidade, de sobrevivência e de risco

A variável aleatória T é definida como uma variável aleatória não negativa e é comumente modelada como contínua e corresponde ao tempo de sobrevivência, ou seja, o tempo decorrido até que um indivíduo experimente uma falha. Essa variável pode ser caracterizada pela sua função densidade de probabilidade $f(t)$, pela função de sobrevivência $S(t)$ e pela função de risco $h(t)$. Além dessas funções, T também pode ser descrito pelas inter-relações entre essas três funções.

A função densidade de probabilidade $f(t)$ é expressa como a probabilidade do indivíduo experimentar a falha em um determinado intervalo de tempo e é dada por:

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

com $f(t) \geq 0$ e $\int_{-\infty}^{\infty} f(t) dt = 1$.

A função de sobrevivência é uma função decrescente definida pela probabilidade de um indivíduo não falhar até um certo instante t , isto é, a probabilidade do indivíduo

sobreviver até o tempo t . Essa função pode ser descrita como (COLOSIMO; GIOLO, 2006):

$$S(t) = P(T \geq t),$$

com $t > 0$. Como consequência, a função de sobrevivência pode ser descrita em termos da função de probabilidade acumulada:

$$S(t) = 1 - P(T < t) = 1 - F(t),$$

em que $F(t) = P(T < t)$ é a função de probabilidade acumulada.

Já a função de risco, ou função taxa de falha é definida como a probabilidade de que a falha ocorra num determinado intervalo $[t, t + \Delta t)$. Essa função pode ser expressa por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t \cdot S(t)}.$$

Outras maneiras de escrever a função de risco, é em termos da função de sobrevivência e da função densidade de probabilidade:

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

Uma medida importante associada à função taxa de falha é a função de risco acumulada, $H(t)$, que mede o risco de ocorrer a falha até um determinado instante t . A função acumulada pode ser escrita em termos da função de risco e, também, em termos da função de sobrevivência:

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

Essa função não apresenta uma interpretação direta, mas pode ser bastante útil na obtenção de $h(t)$, principalmente nas estimações não paramétricas. Além disso, a $H(t)$

também é utilizada no processo de obtenção do modelo de probabilidade mais adequado quando há um grande número de censuras.

Nas situações com muitas censuras, é construído um gráfico da função de risco acumulada estimada, $\hat{H}(t)$. Este gráfico compartilha uma interpretação semelhante à curva de Tempo Total do Teste (TTT) proposto por Aarset(1987).

A elaboração da curva TTT permite visualizar comportamentos esperados da função de risco ao longo do tempo. Esta curva é construída a partir da expressão:

$$G(r/n) = \frac{\sum_{i=1}^n T_{i:n} + (n-r)T_{r:n}}{\sum_{i=1}^n T_i},$$

em que $r = 1, \dots, n$ e $T_{r:n} = 1, \dots, n$ são estatísticas de ordem da amostra. A curva TTT e o gráfico da função de risco acumulada estimada podem ter alguns comportamentos como:

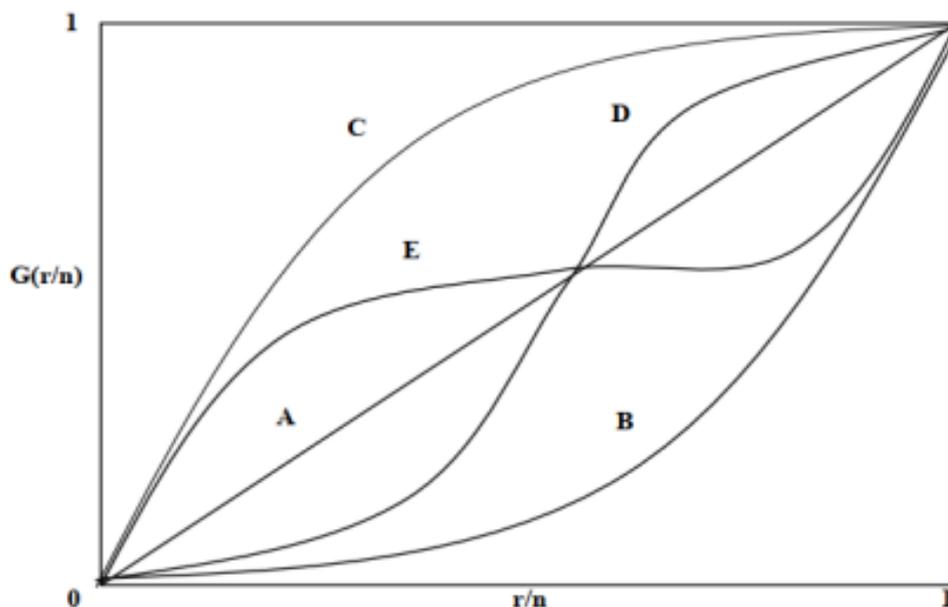


Figura 1: Ilustração de alguns comportamentos da TTT e da $\hat{H}(t)$.Fonte: AARSET, 1987

em que os significados variam de acordo com a técnica usada. Logo, através dos comportamentos da Figura 1, é possível identificar os seguintes comportamentos da função de risco:

- Reta diagonal (A):
 - TTT e $\hat{H}(t)$: função de risco constante.
- Curva convexa (B):

- TTT: função de risco monotonicamente decrescente.
- $\hat{H}(t)$: função de risco monotonicamente crescente.
- Curva côncava (C):
 - TTT: função de risco monotonicamente crescente.
 - $\hat{H}(t)$: função de risco monotonicamente decrescente.
- Curva convexa e depois côncava (D):
 - TTT : função de risco em formato de "U".
 - $\hat{H}(t)$: função de risco unimodal.
- Curva côncava e depois convexa (E):
 - TTT: função de risco unimodal.
 - $\hat{H}(t)$: função de risco em formato de "U".

2.3 Estimador de Kaplan - Meier

O estimador de Kaplan-Meier, também chamado de limite-produto, é uma das técnicas mais utilizadas para estimar a função de sobrevivência em dados censurados (KAPLAN,MEIER,1958). É possível mostrar que a função de sobrevivência estimada, $\hat{S}(t)$, é um estimador de máxima verossimilhança não paramétrico de $S(t)$. Além disso, ele é não viciado para amostras grandes, fracamente consistente, ou seja, converge para o valor real do parâmetro a medida que a amostra cresce, e converge assintoticamente para uma distribuição normal (BRESLOW; CROWLEY, 1974).

No caso de haver n observações e k falhas, com $k \leq n$, o estimador é definido por (KAPLAN,MEIER,1958):

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

em que:

- $t_1 < t_2 < \dots < t_k$ são os k tempos distintos de falha.
- d_j é o número de falhas no tempo t_j em que $j = 1, \dots, k$.
- n_j é o número de indivíduos sob risco em t_j , isto é, número de observações que não falharam ou foram censurados até o momento anterior a t_j .

Na ausência de censura, o estimador de Kaplan-Meier pode ser definida como:

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

no qual, para se obter $\hat{S}(t)$, é necessário a ordenação das observações primeiro.

2.4 Modelos de Probabilidade

2.4.1 Distribuição Weibull e Distribuição Gumbel

A distribuição Weibull é uma distribuição de probabilidade contínua amplamente utilizada em modelagem estatística e análise de dados. É particularmente valiosa para descrever diversas distribuições por apresentar uma grande variedade de formas.

Para T, a função de distribuição de probabilidade da Weibull pode ser descrita como:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \text{ com } t \geq 0,$$

em que $\alpha > 0$ é o parâmetro de escala e $\gamma > 0$ é o parâmetro de forma. Conseqüentemente, a função de sobrevivência e a função de risco para a distribuição Weibull são dadas por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}$$

e

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}.$$

Sabendo como a função de risco é definida, é possível reconhecer que $h(t)$ é estritamente crescente quando $\gamma > 1$, estritamente decrescente quando $\gamma < 1$ e constante quando $\gamma = 1$, levando em conta que o tempo é sempre maior ou igual a 0 e tanto α quanto γ são maior que zero.

Outra maneira de analisar a distribuição de T é considerando a transformação logarítmica dos tempos de falha, que é frequentemente preferida por suas propriedades estatísticas. Ao se aplicar o log, os dados podem resultar em uma distribuição mais simétrica, estabilizando a variância e facilitando a interpretação multiplicativa dos coeficientes nos modelos lineares.

Quando dizemos que T tem distribuição Weibull e fazemos $Y = \log(T)$, assumimos que Y tem distribuição Gumbel, também chamada de distribuição do valor extremo, com função de densidade:

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

com $y, \mu \in \mathbb{R}$ e $\sigma > 0$. A relação dos parâmetros da distribuição Weibull e da distribuição Gumbel pode ser escrita como:

$$\gamma = \frac{1}{\sigma}$$

e

$$\alpha = \exp(\mu).$$

Em decorrência disto, Y apresenta função de sobrevivência

$$S(y) = \exp \left\{ - \exp \left(\frac{y - \mu}{\sigma} \right) \right\},$$

e função de risco

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

2.4.2 Distribuição Log-Normal

Se T tem distribuição log-normal, apresenta função densidade de probabilidade:

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

em que $t > 0$, μ é a média do logaritmo dos tempos de falha e σ é o desvio padrão.

Logo, apresenta função de sobrevivência:

$$S(t) = 1 - \phi \left(\frac{\log(t) - \mu}{\sigma} \right) = \phi \left(\frac{-\log(t) + \mu}{\sigma} \right),$$

em que $\phi(\cdot)$ é a função de distribuição acumulada de uma distribuição normal padrão. Em posse dessas informações, é possível escrever a função de risco dividindo a $f(t)$ pela $S(t)$.

2.4.3 Distribuição Log-logística

A distribuição log-logística é uma outra distribuição bastante utilizada por ser uma alternativa aos outros modelos de probabilidade. As funções densidade de probabilidade, sobrevivência e de risco são, respectivamente (COLOSIMO; GIOLO,2006):

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

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

e

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

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

2.5 Método de Máxima Verossimilhança

Seja uma amostra t_1, t_2, \dots, t_n de uma variável aleatória T , $f(t, \theta)$ representa a função densidade de probabilidade e θ é um vetor de parâmetros. A função de verossimilhança é expressa por:

$$L(\theta) = \prod_{i=1}^n f(t_i, \theta).$$

A contribuição de cada elemento da amostra para a função de verossimilhança pode ser dada por:

$$\begin{cases} f(t_i, \theta) & \text{se } t_i \text{ é tempo de falha.} \\ S(t_i, \theta) & \text{se } t_i \text{ é tempo de censura.} \end{cases}$$

Logo, as observações podem ser divididas em dois conjuntos, as r primeiras observações ordenadas são as não censuradas, e as $n-r$ seguintes são as censuradas. Assim, a função de verossimilhança adaptada para estudos em análise de sobrevivência pode ser elucidada como:

$$L(\theta) \propto \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i},$$

em que δ_i é variável indicadora de falha, $f(t, \theta)$ é a função densidade de probabilidade e $S(t, \theta)$ é a função de sobrevivência do modelo de probabilidade considerado.

Os estimadores de máxima verossimilhança são os valores de θ encontrados a partir de:

$$\frac{d}{d\theta} \log(L(\theta)) = 0.$$

Ao resolver essa equação, os estimadores encontrados são os valores de θ que maximizam $L(\theta)$.

3 Metodologia

3.1 Material

O banco de dados em estudo possui uma observação referente a cada município do Brasil. Em 2024, de acordo com o Instituto Brasileiro de Geografia e Estatística - IBGE, o Brasil possui 5.570 municípios. Dentre esses, vamos desconsiderar o Distrito Federal e o arquipélago de Fernando de Noronha por não possuírem prefeito. Com isso, temos 5.568 municípios com observações relativas à última eleição municipal anterior à data da adesão do município à política pública, e para casos em que não houve a adesão à política, as observações são relativas às eleições municipais de 2016 pois os candidatos eleitos estavam no poder no ano de 2019, que foi o ano em que o programa acabou, não sendo mais possível a sua adesão.

As variáveis em estudo são:

- Unidade da Federação - UF
- Porte
- Adesão ao programa
- Partido do prefeito
- Margem de vitória
- Reeleição
- Ideologia
- Metrópole
- Tempo

As informações referente às variáveis eleitorais foram retiradas e calculadas a partir dos dados disponíveis no site do Tribunal Superior Eleitoral (TSE) e dos Tribunais Regionais Eleitorais (TREs). Já a variável ideologia foi retirada do artigo *Fragmentação sem clivagens Ideologia, preferências políticas e polarização política no Brasil (Fragmentation Without Cleavages Ideology, Policy Preferences, and Political Polarization in Brazil)* de César Zucco e Timothy Power.

Visto que a adesão ou não dos municípios ao programa não dependem uns dos outros, pode-se excluir as observação que não possuem todas as informações. Após as

exclusões, temos 5.168 municípios no estudo que possuem todos os dados. Foram retirados 2 por não se caracterizarem como municípios e não possuírem eleições municipais e 400 por não terem a informação sobre ideologia.

3.2 Método

3.2.1 Modelo de regressão Log-Logístico

Em análise de sobrevivência é comum utilizar modelos de regressão para entender se existem variáveis que influenciam no tempo até a ocorrência do evento estudado e como elas influenciam. Para fazer este estudo foi escolhida a distribuição log-logística citada na subseção 2.5.3.

Com o objetivo de conectar as covariáveis com a variável resposta é preciso usar um preditor linear definido como o produto da matriz transposta de covariáveis com o vetor de parâmetros. O uso desse vetor de covariáveis no modelo de regressão é uma maneira importante de representar toda a heterogeneidade em uma população (LAWLESS, 2011).

A reparametrização do parâmetro de escala pode ser descrita como:

$$\alpha = \exp(\mathbf{x}^T \boldsymbol{\beta}),$$

em que $\mathbf{x}^T = (1, x_1, \dots, x_p)$ é o vetor de covariáveis e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ é o vetor dos coeficientes associados a cada covariável.

Sabendo da reparametrização, é possível reescrever as funções da seção 2.5.3 com a adaptação do parâmetro de escala. Dessa forma, ao fazer a reparametrização o modelo será dado por:

$$f(t) = \frac{\gamma}{(\exp(\mathbf{x}^T \boldsymbol{\beta}))^\gamma} \cdot t^{\gamma-1} \left[1 + \left(\frac{t}{\exp(\mathbf{x}^T \boldsymbol{\beta})} \right)^\gamma \right]^{-2},$$

em que $t > 0$.

A função de sobrevivência do modelo de regressão log-logístico pode ser escrita como:

$$S(t) = \frac{1}{1 + \left(\frac{t}{\exp(\mathbf{x}^T \boldsymbol{\beta})} \right)^\gamma},$$

onde γ é o parâmetro de forma e $\gamma > 0$. Além disso, $\alpha > 0$ é o parâmetro de escala, $\mathbf{x}^T = (1, x_1, \dots, x_p)$ e $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$

3.3 Resíduos de Cox-Snell

Após o ajuste de um modelo de regressão, é necessário avaliar a qualidade do modelo ajustado através da análise dos resíduos (COLOSIMO; GIOLO, 2006). Os resíduos de Cox-Snell são uma forma de analisar o ajuste global do modelo.

Os resíduos de Cox-Snell são encontrados a partir da função de taxa de falha acumulada obtida do modelo ajustado. Pode ser matematicamente escrita por (COX;SNELL,1968):

$$\hat{e}_i = \hat{H}(t_i | x_i).$$

Levando em consideração que $H(t) = -\log S(t)$, o resíduo de Cox-Snell para os modelos de regressão log-logísticos podem ser definidos por:

$$\hat{e}_i = -\log \left(\frac{1}{1 + \left(\frac{t+1}{\exp(\mathbf{x}^T \hat{\beta})} \right)^\gamma} \right).$$

Segundo Lawless (1982) esses resíduos devem seguir uma distribuição exponencial padrão para que o ajuste do modelo esteja adequado. Assim, ao construir um gráfico com a função de sobrevivência dos resíduos através da estimação por Kaplan-Meier e a distribuição exponencial padrão, as curvas precisam estar próximas para que o ajuste seja considerado bom.

4 Análise e Resultados

4.1 Análise Descritiva

4.1.1 Unidade da Federação - UF

Essa variável mostra a distribuição dos 5.168 municípios repartidos nas Unidades das Federações.

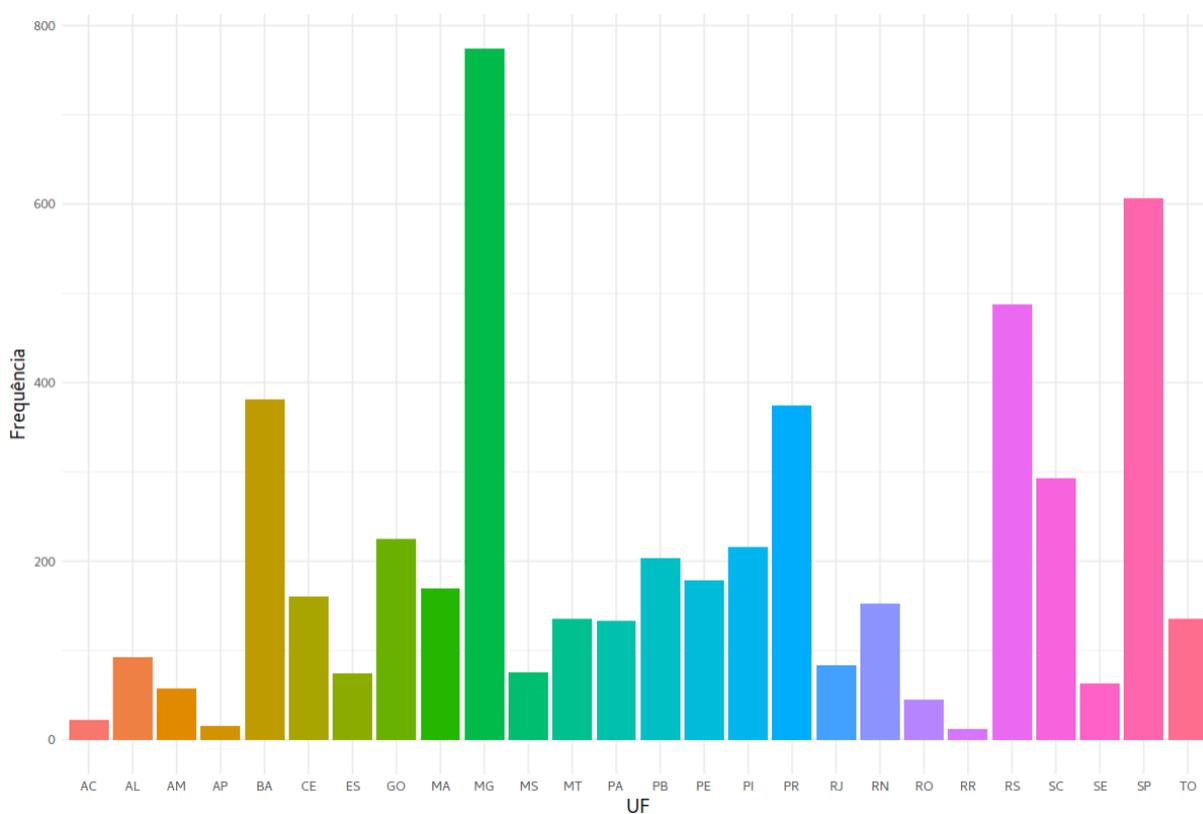


Figura 2: Número de municípios por UF

De acordo com a Figura 2 pode-se perceber que o estado com mais municípios no estudo é Minas Gerais (MG) seguido de São Paulo (SP) e Rio Grande do Sul (RS). Além disso, é nítido que os estados com menos municípios são Acre (AC), Amapá (AP) e Roraima (RR).

4.1.2 Porte

O porte de cada município é determinado pela quantidade de habitantes e foi dividido em sete faixas:

- 1 - Até 5.000
- 2 - 5.001 até 10.000
- 3 - 10.001 até 20.000
- 4 - 20.001 até 50.000
- 5 - 50.001 até 100.000
- 6 - 100.001 até 500.000
- 7 - Maior que 500.000

O porte de cada município gera a seguinte distribuição:

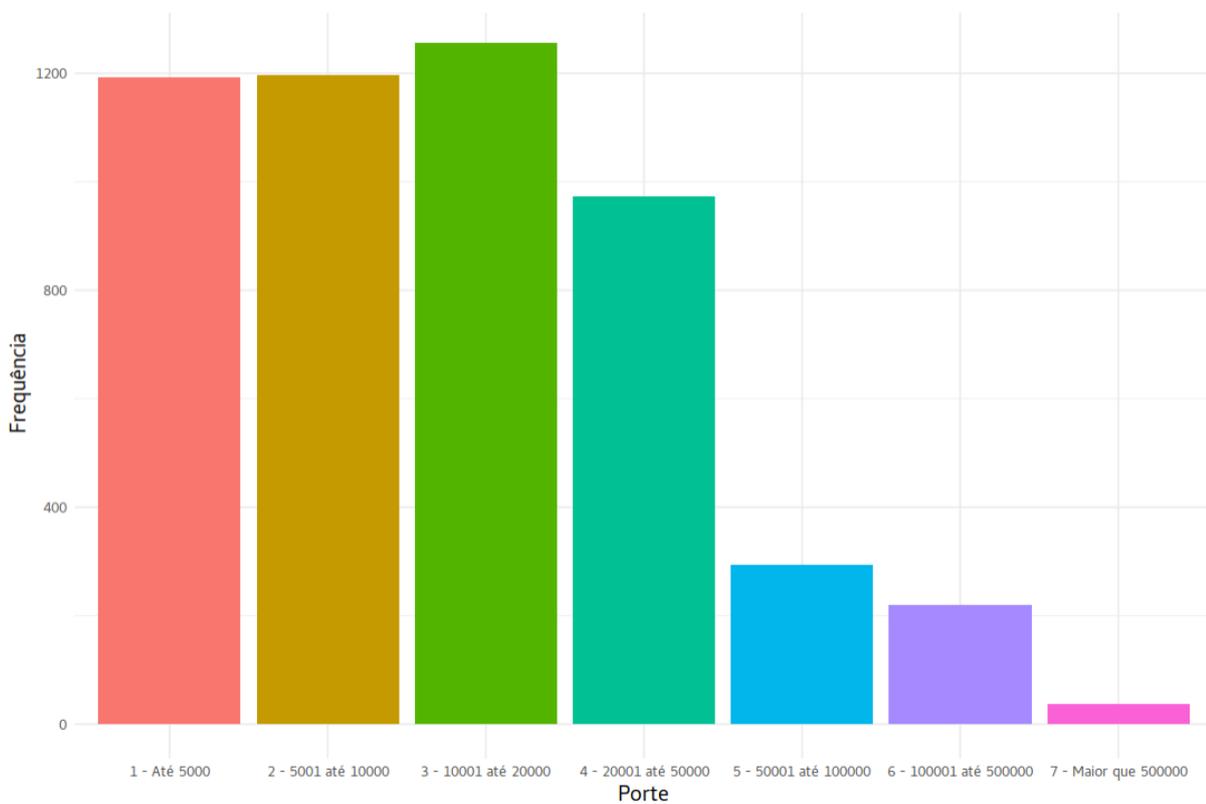


Figura 3: Número de municípios por porte

Através da Figura 3 é perceptível que a maior parte dos municípios do estudo possuem até 50 mil habitantes e que menos de 100 municípios tem população maior que 500 mil.

4.1.3 Adesão ao programa

Essa variável mostra quais municípios aderiram e quais não aderiram ao Programa de Melhoria do Acesso e da Qualidade da Atenção Básica, PMAQ-AB. Está dividida em duas categorias:

- 0 - Não aderiu
- 1 - Aderiu ao programa

Os municípios tiveram desde o dia 19 de julho de 2011 para aderir ao programa. Dos 5.168 municípios, 4.913 aderiram ao programa ao longo de sua existência, conforme pode ser observado abaixo:

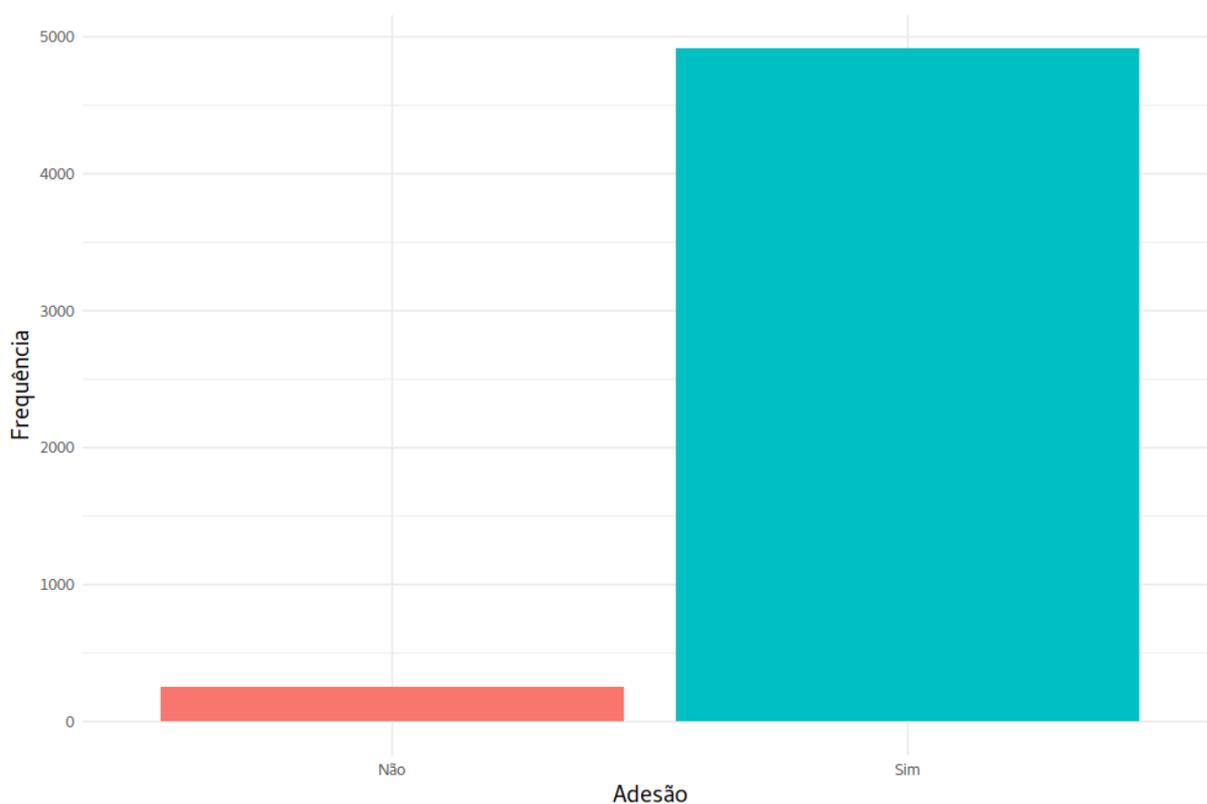


Figura 4: Número de municípios que aderiram ou não ao programa

A Figura 4 mostra que até o dia 11 de novembro de 2019, apenas 255 municípios não adotaram o PMAQ-AB, que foi a data em que foi substituído pelo Programa Previne Brasil.

4.1.4 Partido do prefeito

Essa variável informa a qual partido o prefeito era filiado na eleição anterior a data da adoção ao programa.

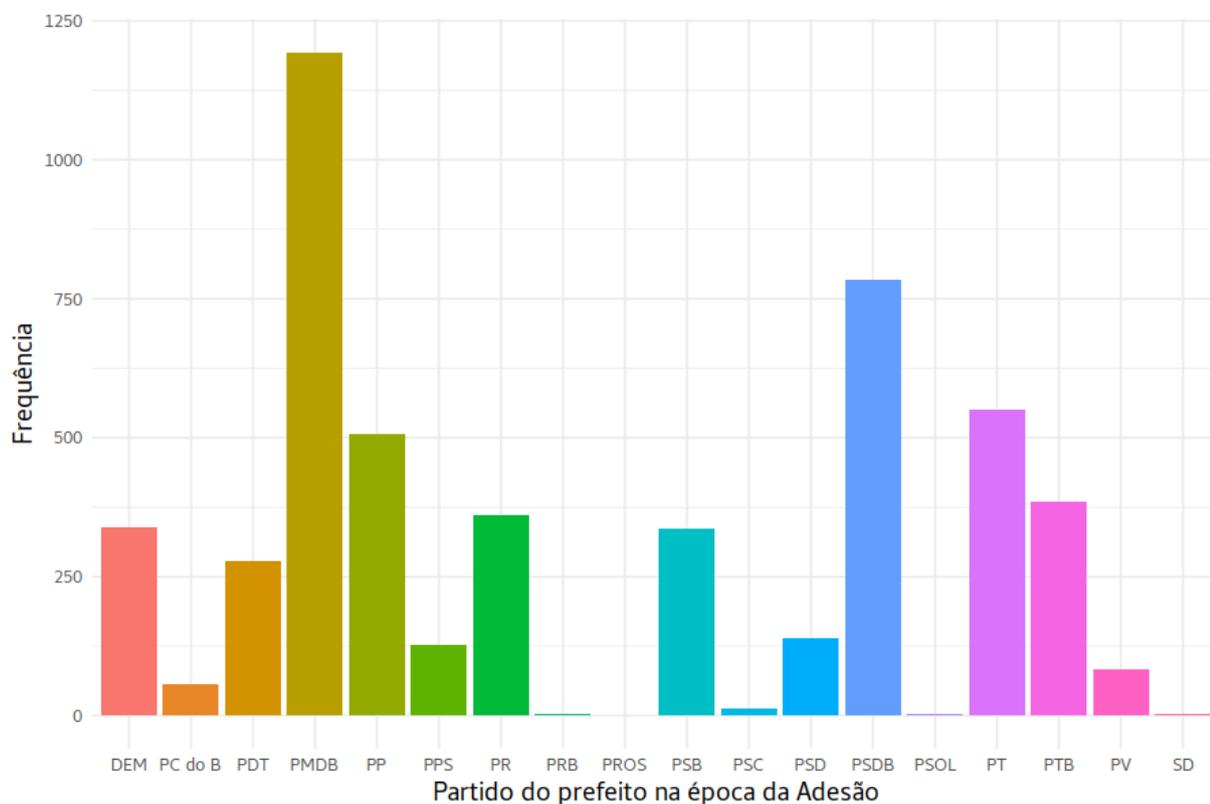


Figura 5: Número de municípios por partido político do prefeito

O Partido do Movimento Democrático Brasileiro - PMDB e o Partido da Social Democracia Brasileira - PSDB são os partidos com maior número de municípios, enquanto o Partido Republicano da Ordem Social - PROS e o Solidariedade - SD são os com menores números de municípios, como mostrado na Figura 5.

4.1.5 Margem de vitória

A margem de vitória foi calculada com base nas informações disponíveis no site do Tribunal Superior Eleitoral - TSE e dos Tribunais Regionais Eleitorais - TREs e é a diferença entre a porcentagem de votos válidos do primeiro colocado e do segundo colocado. No banco de dados varia de 0,02% até 100%

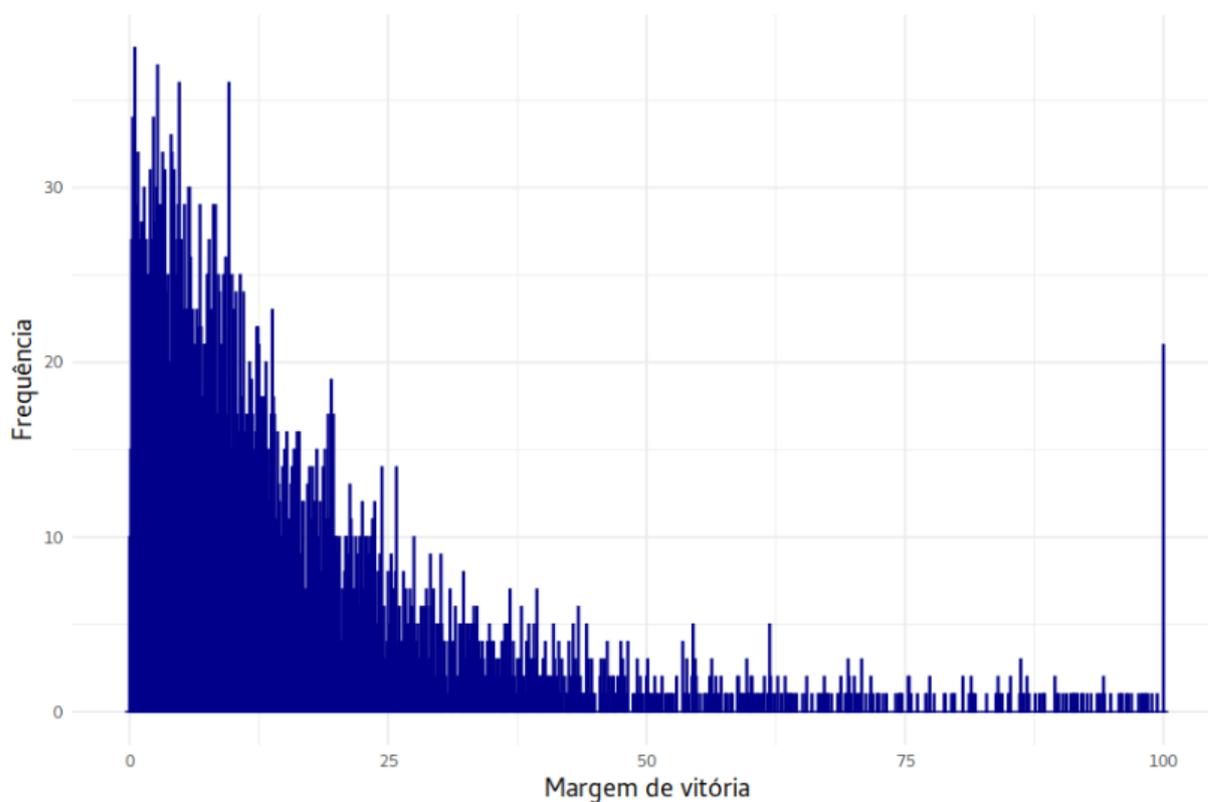


Figura 6: Histograma das margens de vitória

É possível observar na Figura 6 que a maior parte dos prefeitos eleitos tiveram uma margem de vitória abaixo de 50% enquanto apenas 281 prefeitos foram eleitos com mais de 50%. Desses 281 prefeitos, 21 obtiveram 100% de margem de vitória, isso pode ter ocorrido por ser o único candidato no município e/ou o concorrente ter se tornado inelegível durante o processo de eleição.

4.1.6 Reeleição

A variável sobre reeleição é aquela que mostra se o prefeito foi ou não reeleito na eleição anterior à data da adesão ao programa.

- 0 - Não foi reeleito
- 1 - Reeleito

Dos 5.168 municípios apenas 1.075 tiveram um prefeito que foi reeleito na eleição anterior a data da adesão ou, nos casos que não houve a adesão, anterior a novembro de 2019. A Figura 7 demonstra a proporção de prefeitos reeleitos e não reeleitos nos referidos municípios.

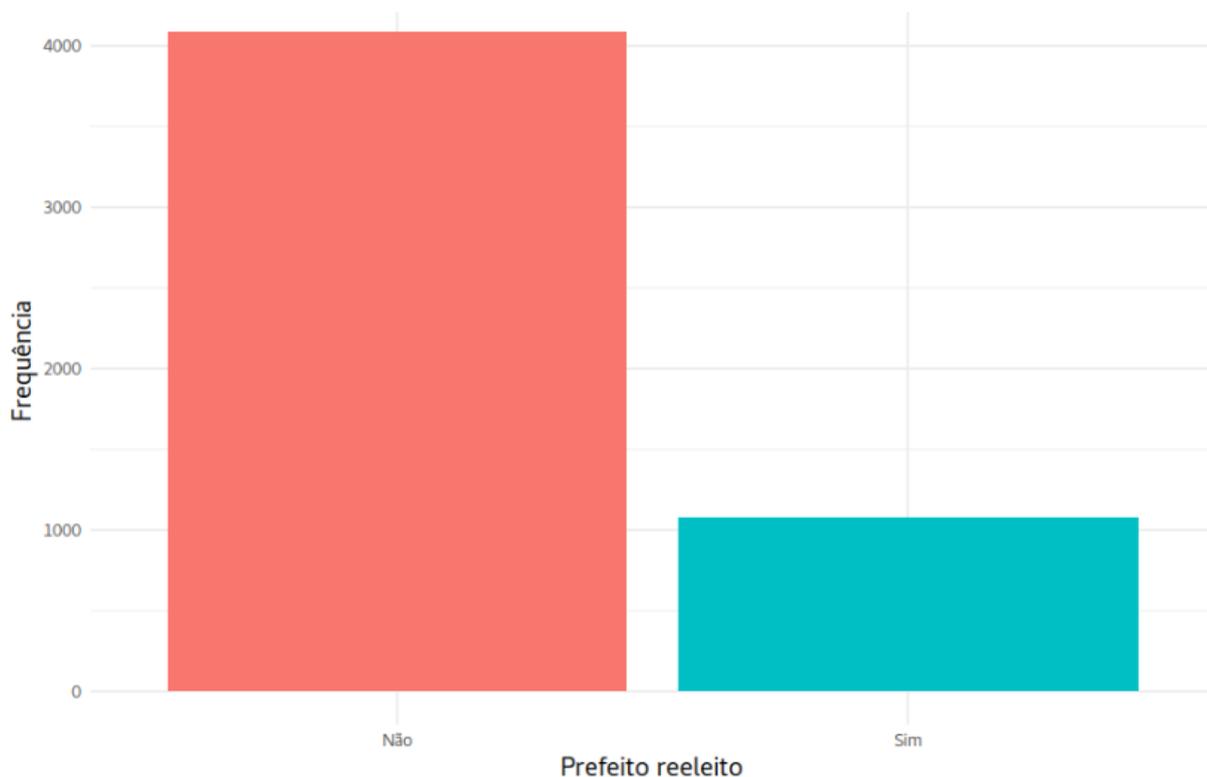


Figura 7: Número de municípios por reeleição do prefeito

4.1.7 Ideologia

As ideologias de direita e de esquerda possuem visões diferentes sobre o papel do Estado, da economia e da sociedade. A direita tende a defender uma economia de mercado com menor intervenção estatal, promovendo privatizações, responsabilidade individual e valores tradicionais, frequentemente adotando posições conservadoras em questões sociais. Já a esquerda defende maior participação do Estado na economia para promover redistribuição de renda e combater desigualdades, além de adotar posturas progressistas em temas sociais, como direitos das minorias e mudanças nos costumes. Enquanto a direita valoriza a liberdade econômica e a ordem social, a esquerda prioriza a igualdade social e justiça distributiva.

A variável Ideologia tenta medir os ideais do partido político com base nas suas ações e falas e coloca a ideologia dele em uma escala que varia de -1 a 1. Quanto mais perto de -1, mais características de esquerda o partido tem e quanto mais perto de 1, de direita. O zero seria o equilíbrio perfeito entre as duas ideologias.

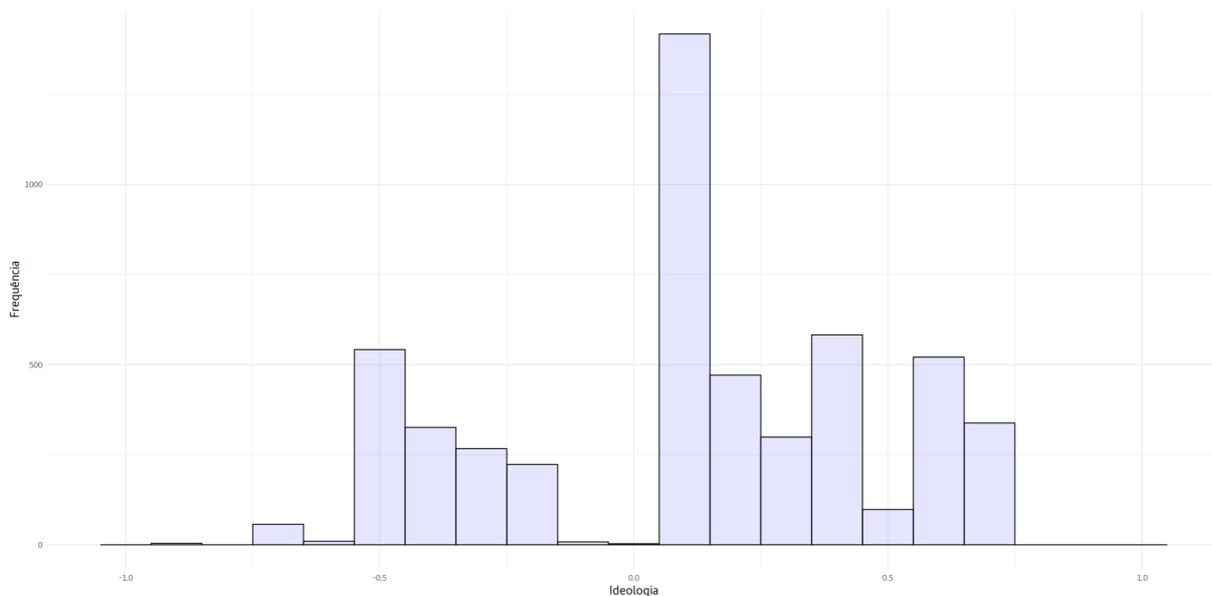


Figura 8: Histograma das ideologias dos partidos políticos dos prefeitos

A Figura 8 mostra que, no momento da observação, havia mais partidos com ideologias de direitas do que de esquerda. O gráfico também mostra que 4 municípios em que a ideologia é abaixo de -0,75 e que não há municípios com ideologias acima de 0,75.

4.1.8 MetrÓpole

Uma metrÓpole é caracterizada por ser uma grande cidade, geralmente com uma densa população e significativa influência econômica, cultural e política na região onde está inserida. Essa variável nos mostra se o município é ou não considerado metrÓpole da seguinte forma:

- 0 - Não é uma metrÓpole
- 1 - MetrÓpole

A Tabela 1 mostra o quantitativo dos municípios que são considerados, ou não, metrÓpoles.

Tabela 1: Distribuição dos municípios em metrÓpoles e não metrÓpoles

MetrÓpole	Não MetrÓpole
16	5152

4.1.9 Tempo

A variável tempo traz a quantidade de meses que o município levou para aderir ao programa ou não. Esse tempo pode variar de 1 a 102 sendo que 1 é quando o município aderiu ao programa em julho de 2011 e 102 em novembro de 2019, que foi quando o programa foi substituído por outra política.

Tabela 2: Distribuição dos municípios de acordo com o tempo de adesão

Meses até a adesão	5	6	7	9	22	29	30	102
Valores	3607	4	20	10	1217	12	40	258

Na distribuição da variável demonstrada na Tabela 2, é possível perceber que 3.607 municípios adotaram o programa em novembro de 2011, 4 municípios em dezembro de 2011, 30 aderiram em 2012, 1.269 em 2013 e 258 não aderiram ao programa até novembro de 2019 quando foi substituído pelo Programa Previne Brasil .

4.2 Modelagem

4.2.1 Distribuição de probabilidade

Antes de partir para o modelo de regressão é necessário encontrar uma distribuição de probabilidade que se adapte à variável resposta. Para auxiliar nessa decisão, é preciso analisar os gráficos TTT e $\hat{H}(t)$.

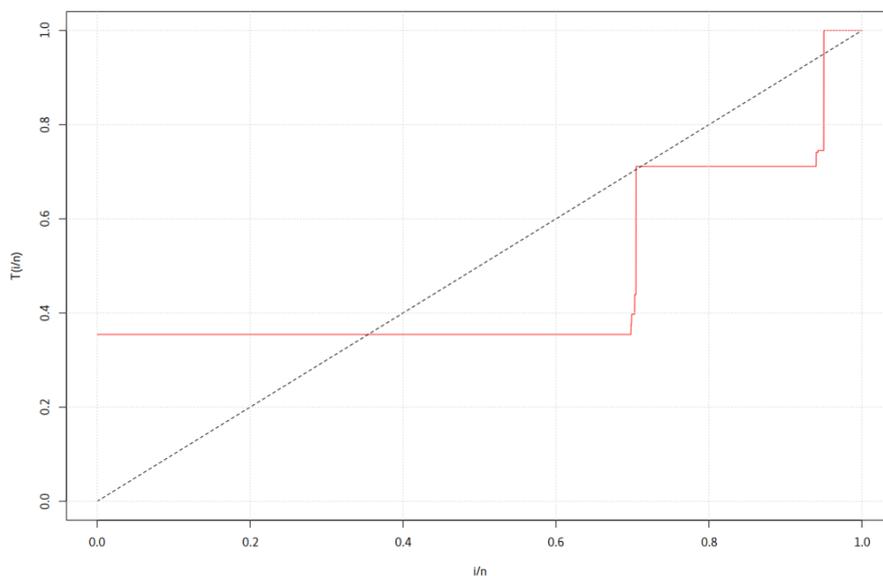


Figura 9: Gráfico TTT

A Figura 9 apresenta um comportamento bem diferente dos apresentados na seção 2.3. Ao analisar o comportamento do Tempo Total de Teste (TTT), observamos que a forma que mais se assemelha à observada é a de uma curva que inicialmente apresenta concavidade e, posteriormente, se torna convexa. Esse padrão é característico de funções de risco unimodais, como as distribuídas log-logística e log-normal.

Ao observar a Figura 10, é possível identificar um comportamento anômalo, embora a curva exiba predominantemente um padrão côncavo. Esse padrão é característico de funções de risco monotonicamente decrescentes, como a distribuição Weibull por exemplo.

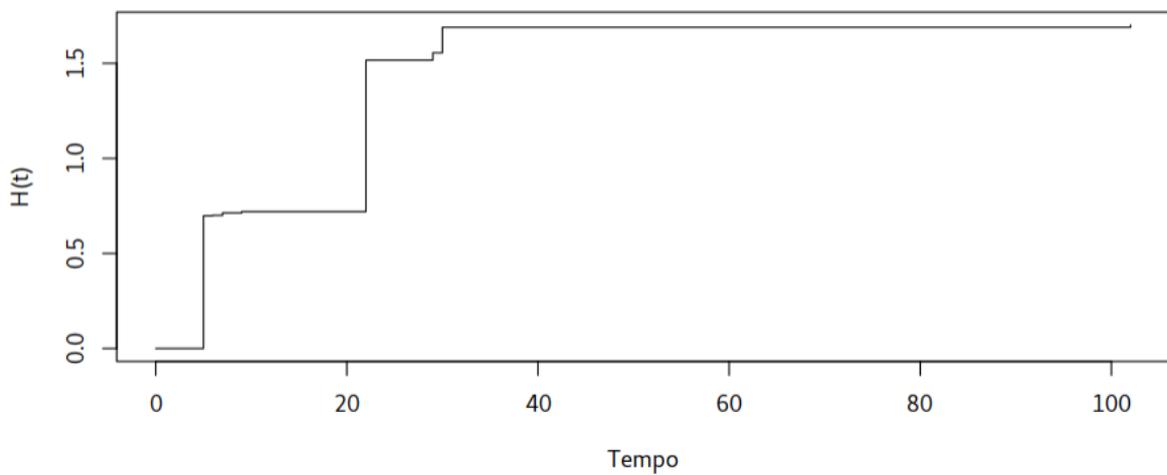


Figura 10: Gráfico $\hat{H}(t)$

Como os gráficos TTT e $\hat{H}(t)$ não foram bem conclusivos, é necessário analisar todas as distribuições juntamente com Kaplan-Meier.

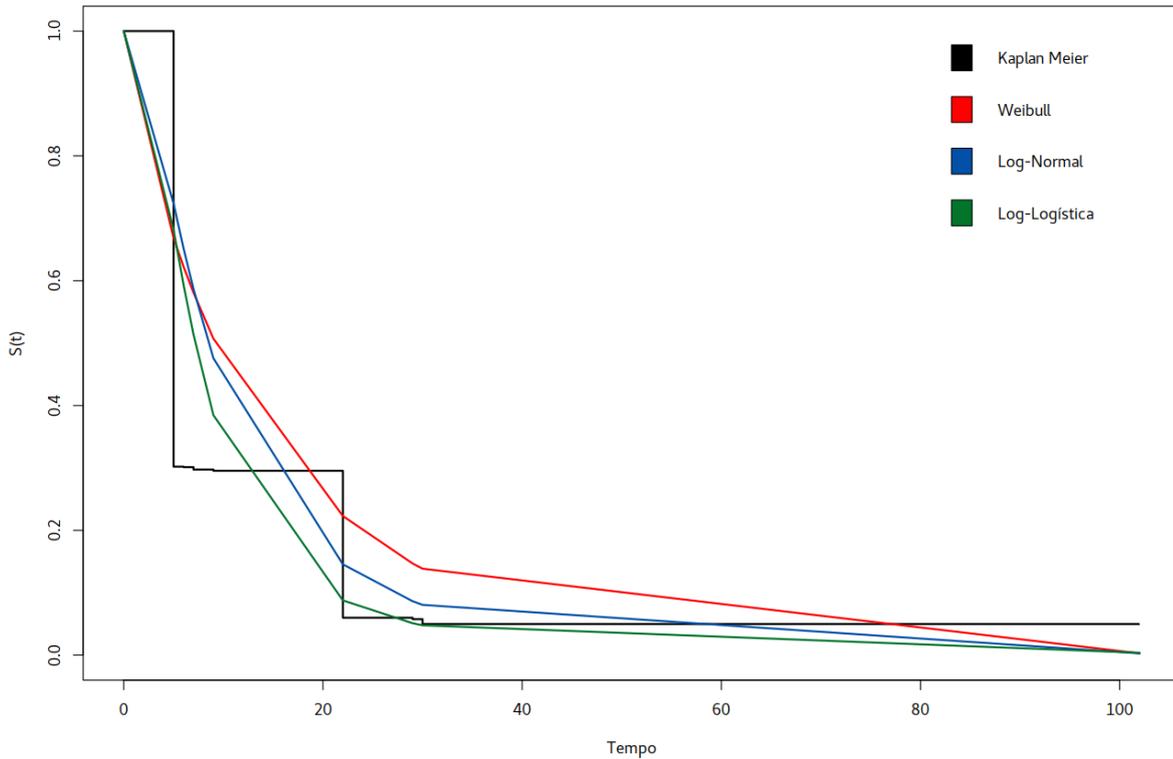


Figura 11: Kaplan-Meier dos tempos e possíveis distribuições de probabilidade

Observando a Figura 11, é possível notar que, visualmente, a curva da distribuição que mais se aproxima de Kaplan-Meier é a curva da log-logística, mas é necessário observar, também, as medidas de informação.

Tabela 3: Medidas de informação por distribuição

Distribuição	AIC	BIC	AIC _c
Weibull	36178.24	36191.34	36178.24
Log-normal	33370.16	33383.27	33370.17
Log-logística	32842.86	32855.96	32842.86

Levando em consideração a Figura 11 e as medidas de informação na Tabela 3, a distribuição que mais se adequa aos dados é a Log-logística pois, comparada com as demais, possui os menores valores quando se olha para os critérios de informação. Após decidir pela distribuição log-logística, a próxima etapa é a seleção de variáveis, sendo possível identificar quais são as variáveis que influenciam na variável resposta deste estudo, ou seja, no tempo de adesão.

4.2.2 Seleção de variáveis

O processo de seleção de variáveis foi iniciado com as variáveis sobre região, porte, margem de vitória, reeleição, turno, ideologia e metrópole. As variáveis relativas a Unidade da Federação - UF e partido do prefeito foram retiradas dessa etapa visto que UF está ligada à região do município e a ideologia é calculada com base no partido do prefeito.

A seleção das variáveis explicativas foi feita utilizando a abordagem Backward. Essa técnica começa com todas as variáveis no modelo e, em cada etapa, há a remoção daquela variável que possui maior p-valor até que todas estejam com p-valor menor que 10%.

Tabela 4: Primeira etapa na seleção de variáveis

Etapa 1			
Variável	Estimativa	Erro Padrão	P-valor
β_0	2.255775	0.049167	$< 2e - 16$
$\beta_{\text{Região Nordeste}}$	-0.141811	0.046555	0.00232
$\beta_{\text{Região Sudeste}}$	-0.206229	0.046474	9.1e-06
$\beta_{\text{Região Sul}}$	-0.242708	0.048219	4.8e-07
$\beta_{\text{Região Centro Oeste}}$	-0.159020	0.057610	0.00577
$\beta_{\text{Porte 2}}$	0.007036	0.034098	0.83651
$\beta_{\text{Porte 3}}$	-0.038629	0.033828	0.25349
$\beta_{\text{Porte 4}}$	-0.147255	0.035254	3.0e-05
$\beta_{\text{Porte 5}}$	-0.169547	0.050809	0.00085
$\beta_{\text{Porte 6}}$	-0.222882	0.055433	5.8e-05
$\beta_{\text{Porte 7}}$	-0.251985	0.160222	0.11578
$\beta_{\text{Margem de Vitória}}$	-0.001095	0.000642	0.08796
β_{Reeleito}	-0.170291	0.026365	1.1e-10
$\beta_{\text{Ideologia}}$	0.094939	0.030619	0.00193
$\beta_{\text{Metrópole}}$	-0.019835	0.237933	0.93356
Log(scale)	-0.750736	0.012183	$< 2e - 16$

A primeira etapa resultou a Tabela 4 mostrando que a variável Metrópole e a primeira a sair do modelo e a próxima etapa é mostrada na tabela abaixo.

Tabela 5: Segunda etapa na seleção de variáveis

Etapa 2			
Variável	Estimativa	Erro Padrão	P-valor
β_0	2.255746	0.049166	$< 2e - 16$
$\beta_{\text{Região Nordeste}}$	-0.141801	0.046555	0.00232
$\beta_{\text{Região Sudeste}}$	-0.206214	0.046474	9.1e-06
$\beta_{\text{Região Sul}}$	-0.242717	0.048219	4.8e-07
$\beta_{\text{Região Centro Oeste}}$	-0.159040	0.057609	0.00577
$\beta_{\text{Porte 2}}$	0.007032	0.034098	0.83661
$\beta_{\text{Porte 3}}$	-0.038634	0.033828	0.25343
$\beta_{\text{Porte 4}}$	-0.147259	0.035254	3.0e-05
$\beta_{\text{Porte 5}}$	-0.169547	0.050809	0.00085
$\beta_{\text{Porte 6}}$	-0.222886	0.055433	5.8e-05
$\beta_{\text{Porte 7}}$	-0.260749	0.120786	0.03087
$\beta_{\text{Margem de Vitória}}$	-0.001094	0.000642	0.08828
β_{Reeleito}	-0.170298	0.026365	1.1e-10
$\beta_{\text{Ideologia}}$	0.094987	0.030613	0.00192
Log(scale)	-0.750737	0.012183	$< 2e - 16$

A seleção de variáveis é finalizada na segunda etapa pois apesar de terem p-valores acima de 10% eles representam categorias de uma variável que também possui categorias significativas.

Além disso, ao considerar a variável região, a sobrevivência em todas as regiões é menor do que na região de referência, ou seja, do que na região Norte. Ou seja, as regiões Nordeste, Sudeste, Sul e Centro Oeste tem maior risco de aderir ao Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ-AB) do que a região Norte.

Enquanto isso, na variável porte, o porte 2 e 3 não têm um efeito significativo à categoria de referência (porte 1) e as demais estimativas para os portes apresentam sinal negativo, reduzindo a sobrevivência. Já a margem de vitória e a reeleição do prefeito também diminuem a sobrevivência do município, ou seja, aumentam o risco do município aderir ao PMAQ-AB. E a ideologia, por ter sinal positivo e ser significativa, impacta em diminuir o risco da adesão ao PMAQ-AB.

Para verificar se o ajuste do modelo da etapa 3 está bom é necessário realizar a análise dos resíduos:

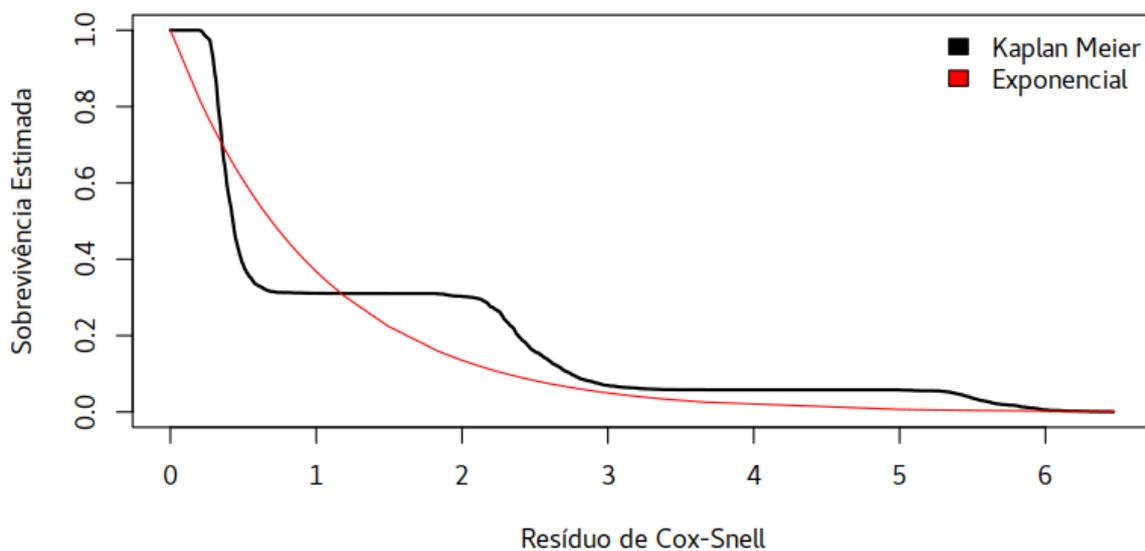


Figura 12: Curva de sobrevivência estimada e resíduos de Cox-Snell

Observando a Figura 12 é possível notar que o resíduo de Cox-Snell tem um comportamento um pouco anormal, mas, de maneira geral, ele tende a seguir a distribuição exponencial padrão. Dessa forma, o modelo pode ser utilizado.

5 Conclusão

O estudo sobre o Programa de Melhoria do Acesso e da Qualidade da Atenção Básica -PMAQ-AB é importante para poder entender as variáveis que influenciam a adoção do programa em diferentes contextos regionais e institucionais. Ao estudar o impacto de características como "região", "porte", "margem de vitória" e "ideologia" sobre a sobrevivência do programa, é possível identificar fatores que podem contribuir para a eficiência das políticas no setor de saúde.

O modelo log-logístico utilizado proporcionou um resultado detalhado de como essas variáveis afetam o tempo até a falha. Além disso, os resultados revelam a importância de considerar características regionais e institucionais na formulação de políticas públicas.

A análise indica que, em geral, o programa de saúde foi mais rapidamente adotado em regiões fora da região Norte e que características como a margem de vitória e a reeleição do prefeito também influenciam o tempo de sobrevivência. Esses achados sugerem que fatores regionais e políticos desempenham papéis cruciais no período de adoção ao PMAQ-AB.

No entanto, o estudo apresenta algumas limitações que devem ser consideradas. Primeiro, a análise pode ser limitada pela qualidade dos dados, que pode influenciar a precisão das estimativas e a validade dos resultados. Além disso, o gráfico do Tempo Total de Teste (TTT) e o gráfico da função de risco acumulado não deram um resultado preciso relativo ao comportamento da função de risco.

A análise dos resíduos de Cox-Snell sugere que o ajuste do modelo pode não ser perfeito, o que indica que o modelo pode não capturar todas as complexidades do processo de falha do programa.

Para contornar essas limitações e melhorar a análise, é recomendável a realização de estudos futuros que explorem diferentes variáveis e modelos. Esses estudos podem contribuir para a formulação de políticas e programas mais eficazes e adaptadas às necessidades específicas de cada município.

Referências

COLOSIMO, E. A. e GIOLO, S. R. **Análise de Sobrevivência Aplicada**. ABE - Projeto Fisher, São Paulo, 1ª edição, 2006.

KALBFLEISCH, J. D. and Prentice, R. L. **The Statistical Analysis of Failure Time Data**. John Wiley and sons, New York, 2nd edition, 2011.

CARVALHO, C. G. C. **Proposta de nova taxa de inovação de adoção de políticas públicas por meio do modelo de regressão de Cox**. Universidade de Brasília (UnB), 2022.

MILHOMEM, M. B. **Proposta de nova taxa de inovação de adoção de políticas públicas por meio do modelo de regressão paramétrico para dados censurados**. Universidade de Brasília (UnB), 2022.

ZUCCO, Cesar; Power, Timothy J. **Fragmentation without cleavages? Endogenous fractionalization in the Brazilian party system**. Comparative Politics, 2019. Disponível em: <https://ssrn.com/abstract=3466149>. Acesso em: 27 ago. 2019.

SANTOS, A. F. D. C. e WERNER, F. **Análise Estatística de Sobrevivência: um estudo de pacientes com câncer de mama tratados no município de Curitiba**. Trabalho de Conclusão de Curso, Universidade Federal do Paraná, 2016.

NAKANO, E. Y. **Um curso de análise de sobrevivência**. 2017.

SILVA, G. O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificada**. Tese (Doutorado) — Universidade de São Paulo, 2008.

BRESLOW, N.; CROWLEY, J. **A large sample study of the life table and product limit estimates under random censorship**. The Annals of statistics, JSTOR, p. 437–453, 1974.

AARSET, M.V. **How to Identify Bathtub Hazard Rate**. IEEE Transactions on Reliability, 1987

KAPLAN, E. L.; MEIER, P. **Nonparametric estimation from incomplete observations**. Journal of the American statistical association, Taylor Francis, v. 53, n. 282, p. 457–481, 1958.

BRASIL. Ministério da Saúde. **Saúde Mais Perto de Você – Acesso e Qualidade Programa Nacional de Melhoria do Acesso e da Qualidade da Atenção Básica (PMAQ)**. Brasília: Ministério da Saúde, 2012.

LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. New York: John Wiley Sons, 2011.