Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Estudo da adesão ao Programa Saúde da Família por meio da metodologia de Análise de Sobrevivência

Leylanne Nayra Figueira de Alencar

Brasília
2018

Leylanne Nayra Figueira de Alencar

Estudo da adesão ao Programa Saúde da Família por meio da metodologia de Análise de Sobrevivência

Orientadora:

Profª. Dra. Juliana Betini Fachini Gomes

Trabalho de Conclusão de Curso apresentado para o Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do título de Bacharel em Estatística.

Brasília

2018

Dedicatória

À minha avó,
Lionidia Dias Figueira *(in memoriam), por ter sido a minha melhor pessoa.*
"Só enquanto eu respirar, vou me lembrar de você"
(O Teatro Mágico).

Agradecimentos

À Deus e à Nossa Senhora, por se fazerem presentes em todos os momentos da minha vida, e por terem colocado pessoas no meu caminho que foram a personificação de Vosso infinito amor e misericórdia por mim.

À professora Dra. Juliana Betini Fachini Gomes por todo apoio, amizade e paciência. Foi uma honra ter sido sua orientanda. Obrigada por tornar essa fase da graduação mais leve.

Aos meus pais e irmãs por todo amor, apoio e compreensão. Vocês foram fundamentais para que eu chegasse até aqui. Aos meus avós *(in memoriam)*, por terem sido tão presentes e tão importantes em todos os momentos em que tive a honra de viver com vocês.

Às pessoas que Brasília me deu! Vocês também são família. Obrigada por compartilharem tanto os momentos bons quanto os difíceis comigo. Lucas, Luíza, Marina, Bruno, Eduarda, Ludimila, Laura, Bruna, Isabella e tantos outros. Eu não teria palavras para expressar o quanto vocês foram e são importantes na minha vida.

Sumário

1 Introdução 5

2 Revisão de Literatura 7

2.1 Análise de Sobrevivência 7

2.1.1 Tempo e Censura 7

2.1.2 Função de sobrevivência 8

2.1.3 O estimador de Kaplan - Meier 9

2.1.4 Função de Risco 9

2.1.5 Função de Risco Acumulada 10

2.1.6 Curva do Tempo Total em Teste 11

2.2 Modelos Probabilísticos 12

2.2.1 Distribuição Weibull 13

2.2.2 Distribuição Log-Normal 13

2.2.3 Distribuição Log-Logística 14

2.3 Estimação dos Parâmetros dos Modelos 14

2.3.1 O Método de Máxima Verossimilhança 15

2.4 Modelos de Regressão 16

2.4.1 Modelo de regressão Weibull 17

2.4.2 Modelo de regressão Log-Normal 17

2.4.3 Modelo de regressão Log-Logístico 18

2.5 Análise de Resíduos 18

2.5.1 Resíduos de Cox-Snell 19

3 Metodologia 20

3.1 Descrição dos dados 20

3.2 Métodos 23

4 Resultados e Discussões 25

4.1 Análise Descritiva 25

4.2 Definição da distribuição de probabilidades 29

4.3 Análise de regressão 30

4.4 Definição dos modelos de regressão para tempo inicial 01/01/1991. 31
4.5 Definição dos modelos de regressão para tempo inicial 01/01/1997. 38
5 Considerações Finais 46
Referências. 48

Resumo

Estudo da adesão ao Programa Saúde da Família
por meio da metodologia de Análise de Sobrevivência

Neste trabalho são propostos os modelos de regressão Weibull e Log-logístico para dados contínuos. A proposta do projeto é identificar, por meio da metodologia de análise de sobrevivência, fatores que exerçam influência na decisão dos governos sobre a adoção ou não de determinada política, considerando, neste caso, o Programa Saúde da Família (PSF). Além disso, a análise considerou dois tempos iniciais diferentes, com foco na obtenção de resultados que possam ser satisfatórios tanto para a estatística quanto para as ciências políticas.

Palavras-chave: Dados contínuos; Dados censurados; Distribuição Weibull; Distribuição Log-logística; Modelos de regressão; Programa Saúde da Família; PSF.

1 Introdução

Com aplicação em diversas áreas do conhecimento, a Estatística é uma ciência exata que reúne um conjunto de técnicas que abrange desde a coleta de dados até a disseminação da informação por eles gerada.

Em muitas pesquisas, tem-se como objeto de análise o tempo até a ocorrência de um evento de interesse, conhecido como falha. É importante, em estudos de sobrevivência, definir de forma clara e precisa o que vem a ser a falha (Colosimo,2006), pois há situações em que o evento de interesse já é descrito de forma bem específica, como quando o foco é o tempo até a morte, por exemplo. No entanto, em outros casos, essa definição é incerta.

Além do estudo das falhas, é comum que em vários desses trabalhos existam observações incompletas ou parciais. Isso se dá, por exemplo, quando há a perda de acompanhamento do objeto de estudo ou a não ocorrência da falha até o término do experimento. Essas observações para as quais não há ocorrência do evento em foco, são chamadas de censuras. E apesar de serem informações incompletas, as censuras são de extrema importância para a análise.

Segundo Colosimo (2006), os conjuntos de dados de sobrevivência são caracterizados pelos tempos de falha e, muito frequentemente, pelas censuras. Esses dois componentes constituem a resposta.

Aplicações de análise de sobrevivência podem ser vistas em diversas áreas, por exemplo, na medicina pode ser o estudo do tempo até a morte ou cura de pacientes, na engenharia pode ser o tempo até a falha de alguma máquina, em ciências políticas pode ser o estudo da difusão de determinada política pública. Ressalta-se aqui uma notável diferença entre os tipos de falhas nas diversas áreas: na saúde, por exemplo, a ocorrência da falha depende do comportamento de quem está sob observação, do tratamento recebido, entre outros fatores, o que torna isso bem individual. Já no âmbito político, a falha depende muito mais, por exemplo, de influência de outros governos do que de alguma característica pessoal do governante, pois há a interdependência entre governos: a falha de um deles pode gerar influência na falha de outros. A tomada de decisão no âmbito político depende também de características geográficas, porte populacional, economia entre outros fatores.

Considerando que análise de sobrevivência busca investigar o tempo até a ocorrência de determinado evento de interesse e que eventos de difusão possuem relação com a questão temporal, pois muitas vezes as causas da adoção de determinado programa se dão por conta de fatores que coexistem. Esse trabalho irá apresentar um estudo sobre a adesão ao Programa Saúde da Família (PSF), projeto

baseado no Programa de Agentes Comunitários de Saúde (PACS), esse criado em 1991, pelo Ministério da Saúde, cujo foco era reduzir a mortalidade infantil e materna, especialmente nas regiões Norte e Nordeste. A partir do PACS, o Ministério da Saúde começou a tirar o foco do indivíduo, passando a dar atenção à família.

Cabe ressaltar, também, que esse movimento de "olhar a família"se deu em muitos países e a formulação do PSF teve a seu favor o desenvolvimento anterior de modelos de assistência à família no Canadá, Cuba, Suécia e Inglaterra que serviram de referência para a formulação do programa brasileiro. (Viana e Poz, 2005)

Segundo Viana e Poz, o PACS serviu de auxílio para a implementação do Sistema Único de Saúde (SUS) e para a organização dos sistemas locais de saúde, uma vez que passou a exigir requisitos para a adesão dos municípios, como o funcionamento dos Conselhos Municipais de Saúde, a existência de uma unidade básica de referência do programa, a disponibilidade de um profissional de nível superior na supervisão e auxílio às ações de saúde e a existência do Fundo Municipal de Saúde.

Com isso, o PACS se desenvolveu nos níveis estadual e municipal, consolidando-se em 1994, ano em que seu financiamento entrou no pagamento por procedimentos operados pelo SUS. Assim, seu sucesso incentivou a formulação do PSF na mesma época, que hoje é conhecido como Estratégia Saúde da Família e busca mudar os aspectos de conhecimento do processo saúde/doença, de forma que o entendimento do contexto em que a doença ocorreu seja objeto de análise. Isso ocorre por meio de grupos de apoio que acompanham as famílias dos municípios participantes do programa.

A base de dados apresenta a data de adesão ao PSF pelos municípios em estudo, bem como informações destas áreas que podem contribuir para a participação ou não do programa, como a região de pertencimento, o porte populacional, entre outras.

Sendo assim, o objetivo deste trabalho é aplicar técnicas de análise de sobrevivência para analisar o tempo entre o início do estudo e a adesão pelos municípios e propor um modelo de regressão que mostre qual o embasamento para a tomada de decisões municipais sobre a participação de um novo programa social, neste caso, o Programa Saúde da Família.

Os modelos de regressão propostos serão com base nas distribuições Log-Logística e Weibull e o objetivo específico é explicar a adesão ao Programa Saúde da Família, tanto do ponto de vista estatístico como político, por meio da utilização do *software* R para a execução da análise de dados.

2 Revisão de Literatura

Nesta seção, serão apresentados conceitos de análise de sobrevivência que irão auxiliar no entendimento da metodologia a ser implementada neste trabalho para o alcance do objetivo final proposto. Que é apresentar modelos de regressão para os dados do Programa Saúde da Família (PSF), de forma que estes modelos apontem quais variáveis tem maior influência na tomada de decisão a respeito da adesão ou não do PSF.

2.1 Análise de Sobrevivência

As técnicas de análise de sobrevivência, também conhecida nas engenharias como Análise de Confiabilidade e em ciências políticas como EHA (*Event History Analysis*), são utilizadas quando o objeto de interesse é o estudo do tempo até a ocorrência de determinado evento ou o risco de ocorrência por unidade de tempo. Sendo assim, nota-se que essa técnica diz respeito ao acompanhamento dos indivíduos observados ao longo do tempo. Sua aplicação pode ser vista em estudos de diferentes áreas, como na medicina, na área financeira, nas ciências políticas, engenharias, dentre outras.

2.1.1 Tempo e Censura

Segundo Colosimo (2006), nos estudos em que o objeto de análise é o tempo até a ocorrência de determinado evento, esse tempo é denominado como tempo de falha e ele irá variar de acordo com o interesse do estudo em questão. Por exemplo, em estudos na área de ciências políticas, o tempo de falha pode ser o tempo até a adoção de determinado programa social.

Como a variável de interesse não será medida instantaneamente, elá será composta por falhas e censuras. Censuras são as observações para as quais não se tem registro do evento de interesse, porém há, apesar de incompletas, informações da variável em estudo. Dessa forma, é necessário que, mesmo censurados, todos os resultados obtidos em um estudo de sobrevivência, sejam usados em sua análise.

Segundo Colosimo (2006), duas razões justificam esse procedimento: (i) mesmo incompletas, as censuras fornecem informações sobre o tempo em estudo; (ii) a omissão das censuras pode acarretar conclusões viciadas no estudo.

Para diferenciar as informações de falha e censura, é necessário inserir uma variável indicadora de censura, que será denotada por δ_i. Essa variável é definida da

seguinte maneira:

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

As censuras podem ser classificadas em: censura à direita, à esquerda ou intervalar. A censura à direita acontece quando o tempo entre o início do estudo e o evento é maior que o tempo observado para determinado indivíduo, consequentemente não é alcançado o desfecho. Neste caso, é aproveitada a informação do tempo em que o objeto de estudo esteve em observação sem a ocorrência do evento.

Esse tipo de censura se divide da seguinte forma: Tipo I, Tipo II e Aleatória.

- Censura Tipo I - quando o estudo termina após um tempo pré-determinado. Assim, serão consideras como censura as observações para as quais não houve registro do evento de interesse até o tempo pré-determinado.
- Censura Tipo II - quando o estudo termina após a ocorrência do evento de interesse para um número pré-determinado de observações. Ou seja, é determinado um número K de falhas e o estudo é finalizado após a ocorrência do evento de interesse para K observações.
- Censura Aleatória - quando o evento de interesse ocorre por um motivo diferente do que está sendo estudado ou quando uma ou mais observações são retiradas do estudo antes da ocorrência do evento de interesse. São censuras que ocorrem sem intervenção do pesquisador.

A censura à esquerda acontece quando, no momento que há a observação do indivíduo, o evento de interesse já ocorreu.

A censura intervalar ocorre quando a falha se encontra em um intervalo de tempo, mas não sabe-se o momento exato de ocorrência no intervalo em questão. As censuras à direita e à esquerda são casos particulares de censura intervalar.

2.1.2 Função de sobrevivência

A função de sobrevivência, S(t), é a probabilidade de uma observação não falhar até determinado tempo t, e é denotada por:

$$S(t) = (T > t) = 1 - F(t), \tag{1}$$

em que T é uma variável aleatória positiva que diz respeito ao tempo de falha e

F(t) é a função de distribuição acumulada de T.

Tem-se, como característica da maioria dos estudos de sobrevivência, a presença de censura. Isso dificulta a análise descritiva convencional que é feita em outros tipos de estudo, como média, desvios-padrão e análise gráfica. É possível fazer uma análise pelo gráfico de dispersão, mas as observações censuradas tornam a análise do gráfico mais complicada, apesar de possível. Sendo assim, nota-se que a presença de censura traz consigo a necessidade do uso de técnicas que possam incorporar as informações incompletas na análise dos dados. Assim, faz-se uso de algumas técnicas descritivas específicas para esse tipo de dados. Comumente, utiliza-se a estimativa para a função de sobrevivência de forma não paramétrica e, a partir dela, busca-se estimar estatísticas de interesse. Para estimar a função de sobrevivência são usados alguns estimadores não-paramétricos. Entre eles, o estimador de Kaplan-Meier, que será utilizado neste trabalho.

2.1.3 O estimador de Kaplan - Meier

Esse estimador considera o número de intervalos como sendo a mesma quantidade de tempos de falha, sendo esses tempos os limites dos intervalos. Assim, a probabilidade de chegar ao tempo t será o produto da probabilidade de chegar até cada tempo anterior. Ele é também conhecido como estimador limite-produto e é uma adaptação da função de sobrevivência sem censuras. Como este estimador envolve uma sequência de passos que são gerados pelos intervalos, após a ordenação dos tempos de falha, ele pode ser feito em termos de probabilidades condicionais. Neste caso, S(t) é escrita como:

$$\hat{S}(t) = \prod_{j:t_j<t} \Big(1 - \frac{d_j}{n_j}\Big), \tag{2}$$

em que d_j é o número de falhas em t_j, $j = 1, 2, ..., k$, em que k é o número de tempos distintos de falha e n_j é o número de observações sob risco em t_j.

As principais propriedades do estimador de Kaplan-Meier são: ele é não viciado para grandes amostras, é fracamente consistente, converge assintoticamente para um processo gaussiano e é estimador de máxima verossimilhança de S(t). Espera-se que, quanto maior for o número de intervalos, melhor a aproximação para o verdadeiro número de falhas.

2.1.4 Função de Risco

Essa função é conhecida como função de risco ou taxa de falha e é denotada por h(t). Ela representa o risco instantâneo que o indivíduo tem de experimentar o

evento de interesse em determinado tempo, ou seja, ela é a probabilidade de ocorrência da falha no intervalo [tj, tj + 1). O valor obtido em h(t) irá mostrar a mudança da taxa de falha de acordo com o tempo em estudo. Por esse motivo, essa função é tão utilizada para descrever o comportamento do tempo de sobrevivência. A taxa de falha é definida por:

$$h(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{3}$$

Ou seja, o risco instantâneo é a razão do limite da probabilidade de uma observação falhar no intervalo de tempo $[t, t + \Delta t]$, sendo que o indivíduo não experimentou o evento de interesse até o tempo t sobre o intervalo de tempo Δt.

Quando é calculada em termos da função densidade de probabilidade, f(t), e da função de sobrevivência, a função de risco é representada da seguinte forma:

$$h(t) = \frac{f(t)}{S(t)}. \tag{4}$$

A partir da h(t), pode-se obter a função de risco acumulada.

2.1.5 Função de Risco Acumulada

A função de risco acumulada, H(t), tem um importante papel para determinar qual distribuição de probabilidade ou classes de distribuições de probabilidade podem ser usadas para modelar a variável resposta. Como $S(t) = \exp\{-H(t)\}$, então $H(t) = -\log S(t)$. Assim, ao utilizar o método de Kaplan-Meier, a estimativa da função de risco acumulada é obtida por:

$$\hat{H}(t) \approx -\log S_{KM}(t). \tag{5}$$

Para ter ideia do comportamento da função de risco a partir do gráfico da função de risco acumulada, deve-se interpretar a Figura 1 da seguinte maneira:

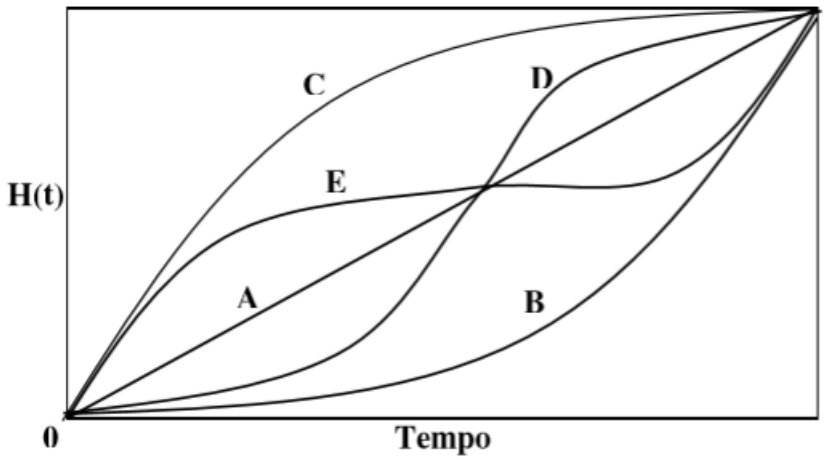

Figura 1: Função de Risco Acumulada.

A) Reta diagonal - indica função de risco constante;

B) Curva convexa - indica que a função de risco é monotonicamente crescente;

C) Curva côncava - indica que a função de risco é monitonicamente decrescente;

D) Curva convexa e côncava - tem-se a função de risco unimodal;

E) Caso reverso - quando a função de risco tem a forma de banheira.

Para o caso **A**, a distribuição sugerida é a Exponencial, para o **B**, a Weibull e a Weibull modificada generalizada, para o caso **C**, Weibull e Log-Logística, para **D** as distribuições candidatas são Log-Normal, Log-Logística, Burr XII, Weibull modificada generalizada e modelos de riscos múltiplos. E, para o caso **E**, as distribuições sugeridas são Weibull modificada generalizada, Kumaraswamy e modelos de riscos múltiplos. Outras distribuições também podem ser consideradas para cada uma dessas situações. Aqui foram definidas apenas alguns exemplos.

2.1.6 Curva do Tempo Total em Teste

Uma outra metodologia gráfica para verificar qual distribuição de probabilidade pode ser usada para modelar a variável resposta é a curva do tempo total em teste (TTT), proposta por Aarset (1987). Essa curva é definida por:

$$G\left(\frac{r}{n}\right) = \frac{\sum_{i=1}^{r} T_{i:n} + (n-r)T_{r:n}}{\sum_{i=1}^{n} T_i}, \tag{6}$$

em que $r = 1, ..., n$ e $T_{i:n}$,$i = 1, ..., n$ são as estatísticas de ordem da amostra.

Um gráfico de $G\left(\frac{r}{n}\right)$ versus r/n apresenta as formas definidas na Figura 2.

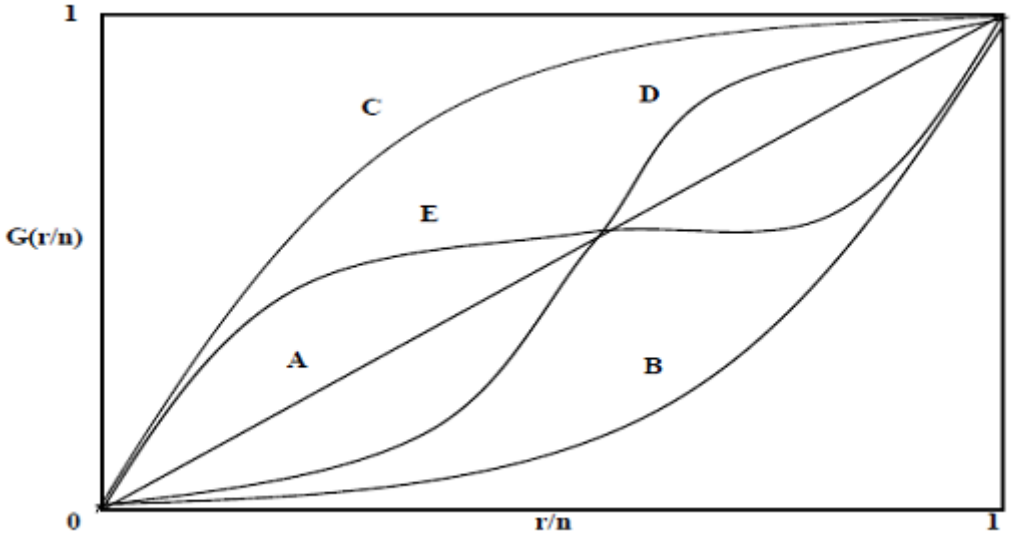

Figura 2: Curva do tempo total em teste.

Sua interpretação é feita da seguinte forma:

Se uma reta diagonal é observada **(A)**, uma função de risco constante é indicada, ou seja, o modelo Exponencial é indicado. Se a curva é convexa **(B)** a função de risco é monotonicamente decrescente e as distribuições candidatas são Weibull e a Weibull modificada generalizada. Se a curva é côncava **(C)**, ou seja, crescente,as distribuições candidatas são Weibull e Weibull modificada. Segundo Nakano (2017), se a curva é côncava e então convexa **(E)**, a função de risco é unimodal (Log-Normal e Log-Logística), e no caso reverso **(D)** a função de risco tem a forma de U. Várias regiões côncavas e convexas direcionam para o ajuste de uma função de risco multimodal. Essas curvas (riscos multimodais) podem ser ajustadas através de distribuições de múltiplos riscos ou distribuição de misturas.

2.2 Modelos Probabilísticos

Segundo Colosimo (2006), os modelos probabilísticos ou paramétricos, também conhecidos como distribuições de probabilidade são usados na análise estatística para dados de sobrevivência. Os parâmetros desses modelos são estimados a partir do método de verossimilhança. Algumas distribuições são comumente usadas para descrever os dados de estudos, como a gaussiana, por exemplo. Mas, quando o estudo envolve o tempo até a ocorrência de um evento de interesse, é necessário fazer uso de outras distribuições para a obtenção dos resultados buscados. Serão

apresentadas a seguir, algumas das distribuições mais comuns na análise de sobrevivência.

2.2.1 Distribuição Weibull

A Distribuição Weibull é muito usada para descrever o tempo de vida de produtos e indivíduos e possui função de risco monótona, ou seja, ela pode ser crescente, decrescente ou constante e suas funções de densidade, sobrevivência e risco são definidas por:

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

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

e

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

para $t \geq 0$, $\gamma > 0$ é o parâmetro de forma e $\alpha > 0$ é o parâmetro de escala.

Para $\gamma > 1$, a função será crescente e para $\gamma < 1$, decrescente. Um caso particular desta distribuição se dá para $\gamma = 1$. Quando isso ocorre, tem-se a distribuição Exponencial, que apresenta um único parâmetro e é uma das distribuições mais simples em análise de sobrevivência. Sua principal característica é a de ser a única distribuição com função de risco constante, o que significa que a chance de falhar tanto para uma observação nova como para uma antiga é a mesma.

2.2.2 Distribuição Log-Normal

Assim como a distribuição Weibull, a Log-Normal também é bastante utilizada para descrever o tempo de vida de produtos e indivíduos, porém sua função de risco não é monótona. Ela cresce, atinge um valor máximo e depois decresce, logo essa distribuição pode apresentar função de risco unimodal. Ela possui uma relação com a distribuição normal: se uma variável aleatória T segue uma distribuição Log-Normal, o logaritmo desta variável irá seguir uma normal com parâmetros μ e σ. A densidade de variável aleatória T com distribuição Log-Normal é:

$$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right\}, \qquad t > 0,$$

em que $-\infty < \mu < +\infty$ é a média e $\sigma > 0$ é o desvio-padrão do logaritmo do tempo de falha.

As respectivas funções de sobrevivência e risco são:

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

e

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

em que $\Phi(.)$ é a função de distribuição acumulada da normal padrão. As funções de sobrevivência e risco para a Log-Normal não possuem uma forma analítica explícita, pois dependem de $\Phi(.)$

2.2.3 Distribuição Log-Logística

Esta distribuição é uma alternativa à Log-Normal e Weibull. A função de densidade para uma variável aleatória T que tenha distribuição Log-Logística é:

$$f(t) = \frac{\gamma}{\alpha^\gamma}t^{\gamma-1}\left(1+\left(\frac{t}{\alpha}\right)^\gamma\right)^{-2}, \qquad t>0,$$

em que $\alpha > 0$ e $\gamma > 0$ são os o parâmetros de escala e forma, respectivamente.

As funções de sobrevivência e taxa de falha são representadas por:

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

e

$$h(t) = \frac{\gamma\left(\frac{t}{\alpha}\right)^{\gamma-1}}{\alpha\left[1+\left(\frac{t}{\alpha}\right)^\gamma\right]}.$$

Essa função possui uma vantagem em relação a distribuição Log-Normal, pois ela apresenta funções de sobrevivência e risco bem definidas. Outra característica dessa distribuição é que a função de risco apresenta, assim como na distribuição Log-Normal, formas unimodais. Assim, as funções com $\gamma = 1$ tem um comportamento apenas decrescente e com $\gamma > 1$ a densidade de probabilidade e de risco tem formas crescente até um determinado ponto de pico e depois decresce, ou seja, uma função unimodal (Damião,2017).

2.3 Estimação dos Parâmetros dos Modelos

Alguns dos métodos mais conhecidos quando se trata de estimação de parâmetros de modelos são o método de mínimos quadrados e o método de momentos.

Apesar de serem uns dos mais comumente usados, em análise de sobrevivência sua utilização é inviável, pois eles não englobam as censuras na estimação. Uma alternativa para isso é fazer uso do método de máxima verossimilhança, que, por sua vez, engloba as censuras e possui propriedades apropriadas quando se trata de grandes amostras.

2.3.1 O Método de Máxima Verossimilhança

Esse método busca encontrar o valor de $\boldsymbol{\theta}$, em que $\boldsymbol{\theta}$ pode ser um parâmetro ou conjunto de parâmetros, que maximizem a probabilidade de ocorrência da amostra observada. Os elementos da amostra auxiliam na construção da função de verossimilhança da seguinte forma: a função de sobrevivência será a contribuição de cada observação censurada, $S(t_i|\boldsymbol{\theta})$, e a função de densidade será a contribuição de cada falha, $f(t_i|\boldsymbol{\theta})$.

Assim, segundo Nakano (2017), com base na amostra (t_1, δ_1), (t_2, δ_2), ..., (t_n, δ_n), e considerando que os t_i's são independentes e provenientes da mesma distribuição de probabilidades indexada pelo parâmetro $\boldsymbol{\theta}$, a função de verossimilhança é genericamente escrita na seguinte forma:

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

$$= \prod_{i=1}^{n} \Big[h(t_i|\boldsymbol{\theta})\Big]^{\delta_i} S(t_i|\boldsymbol{\theta}),$$

em que δ_i é a variável indicadora de censura.

Para encontrar os estimadores, resolve-se o seguinte sistema de equações:

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

Os estimadores de máxima verossimilhança possuem as seguintes propriedades:

- São assintoticamente não viesados;
- São consistentes;
- São eficientes, ou seja, possuem variância mínima, na classe de estimadores não viesados;
- Possuem distribuição assintótica Normal.

2.4 Modelos de Regressão

Nos estudos estatísticos é bastante comum a existência de relação entre duas ou mais variáveis. Em análise de sobrevivência, geralmente há variáveis que possuem associação com o tempo de vida em questão. Elas são chamadas de covariáveis e são usadas para representar a heterogeneidade dos dados que estão sob análise.

Para mostrar a relação, é importante que seja feita uma modelagem para os dados em estudo. Essa modelagem, conhecida como análise de regressão, busca apresentar a influência das variáveis explicativas sobre os tempos de sobrevivência.

No modelo de regressão linear a relação entre a variável resposta e as covariáveis é apresentada por meio de uma relação linear. No caso de uma única covariável, o gráfico desta versus a resposta deve mostrar evidências de uma relação linear, caso o modelo seja aceitável para esta situação. Ou seja, a nuvem de pontos deste gráfico deve dar indicações de que uma reta é uma boa aproximação para a relação entre as variáveis.(Colosimo, 2006). Nesse modelo, a variação em torno da reta geralmente possui uma distribuição normal.

Ao fazer a modelagem em análise de sobrevivência, a utilização direta do modelo de regressão linear é inviável. A forma como é composta a variável resposta e o fato de que sua distribuição geralmente apresenta assimetria na direção dos maiores tempos de sobrevivência são fatores que inviabilizam a utilização do modelo de regressão linear, pois a variável resposta precisa incluir as censuras, caso elas existam e a distribuição da resposta de forma assimétrica inviabiliza o pressuposto de normalidade na variação em torno da reta.

Uma das formas de fazer um modelo de regressão em análise de sobrevivência é reparametrizar a distribuição de probabilidade dos tempos ou usar modelos de locação e escala, o que irá gerar um modelo paramétrico, ou ainda, podem ser usados os modelos de riscos proporcionais, que irão gerar um modelo semi-paramétrico, conhecido como modelo de regressão de Cox.

Neste estudo, será utilizada a reparametrização das distribuições de probabilidade citadas na seção 2.3.

A reparametrização pode ser feita de diversas formas: ela pode acontecer no parâmetro de forma, escala ou em ambos. Sendo mais comum acontecer no parâmetro de escala. Considerando $\boldsymbol{x} = (x_0, x_1, ..., x_p)'$ um vetor formado de covariáveis, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ um vetor de parâmetros associado às covariáveis, $g(.)$ uma função positiva e contínua e sendo α o parâmetro de escala, a reparametrização pode ser representada da seguinte forma:

$$\alpha = g(\mathbf{x}'\boldsymbol{\beta})$$

Alguns exemplos de $g(.)$ são definidos por:

- $\alpha = \exp(\boldsymbol{x}'\boldsymbol{\beta})$;
- $\alpha = (\boldsymbol{x}'\boldsymbol{\beta})$;
- $\alpha = -(\boldsymbol{x}'\boldsymbol{\beta})$, entre outros.

Neste estudo será utilizada a função na forma $\alpha = \exp(\mathbf{x}'\boldsymbol{\beta})$, pois assim, será garantido que o parâmetro continuará sendo positivo.

2.4.1 Modelo de regressão Weibull

Ao considerar que uma variável aleatória T segue uma distribuição Weibull,(α, γ), com função de densidade definido por:

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

ao reparametrizar, relacionando as covariáveis ao parâmetro α, e utilizando $\alpha = \exp(\boldsymbol{x}'\boldsymbol{\beta})$, o modelo será representado por:

$$f(t|\boldsymbol{x}) = \frac{\gamma}{\boldsymbol{x}'\boldsymbol{\beta}}^{\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\boldsymbol{x}'\boldsymbol{\beta}}\right)^{\gamma}\right\}),$$

e sua respectiva função de sobrevivência é definida por:

$$S(t|\boldsymbol{x}) = \exp\left\{-\left(\frac{t}{\exp(\boldsymbol{x}'\boldsymbol{\beta})}\right)^{\gamma}\right\},$$

em que $t > 0$, $\gamma > 0$ é o parâmetro de forma, $\alpha > 0$ é o parâmetro de escala, $\boldsymbol{x} = (x_0, x_1, ..., x_p)'$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$.

2.4.2 Modelo de regressão Log-Normal

Seja T uma variável aleatória com distribuição Log-Normal (μ, σ) com densidade de probabilidade dada por

$$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right\}, \qquad t > 0,$$

em que μ é a média e σ é o desvio-padrão do logaritmo do tempo de falha. Ao reparametrizar a média na forma $\mu = (\boldsymbol{x}'\boldsymbol{\beta})$, o modelo será dado por

$$f(t|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(t) - (\boldsymbol{x}'\boldsymbol{\beta})}{\sigma}\right)^2\right\}, \qquad t > 0,$$

e sua função de sobrevivência será:

$$S(t|\boldsymbol{x}) = \Phi\left(\frac{-\log(t) + (\boldsymbol{x}'\boldsymbol{\beta})}{\sigma}\right),$$

em que $\Phi(.)$ é a função de distribuição acumulada da normal padrão, $\boldsymbol{x} = (x_0, x_1, ..., x_p)'$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$.

2.4.3 Modelo de regressão Log-Logístico

Ao supor que uma variável aleatória T tenha distribuição Log-Logística (α, γ), com função de densidade de probabilidade definida por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}\left(1 + \left(\frac{t}{\alpha}\right)^\gamma\right)^{-2}, \qquad t > 0,$$

em que $\alpha > 0$ e $\gamma > 0$ são os parâmetros de escala e forma, respectivamente. Ao fazer a reparametrização no parâmetro de escala, utilizando $\alpha = \exp(\boldsymbol{x}'\boldsymbol{\beta})$, o modelo será dado por:

$$f(t|\boldsymbol{x}) = \frac{\gamma}{\exp(\boldsymbol{x}'\boldsymbol{\beta})^\gamma} t^{\gamma-1}\left(1 + \left(\frac{t}{\exp(\boldsymbol{x}'\boldsymbol{\beta})}\right)^\gamma\right)^{-2}, \qquad t > 0,$$

A função de sobrevivência é representada por:

$$S(t|\boldsymbol{x}) = \frac{1}{1 + \left(\frac{t}{\exp(\boldsymbol{x}'\boldsymbol{\beta})}\right)^\gamma},$$

em que $\alpha > 0$ e $\gamma > 0$ são os parâmetros de escala e forma, $\boldsymbol{x} = (x_0, x_1, ..., x_p)'$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$.

2.5 Análise de Resíduos

Uma parte de extrema importância na análise de dados é avaliar o ajuste do modelo proposto. Essa avaliação busca mostrar a adequação da distribuição definida para os dados, além de validar as suposições do modelo.

Técnicas gráficas são utilizadas para avaliar, por exemplo, a distribuição dos erros por meio dos resíduos. Vale ressaltar que elas são úteis também para rejeitar modelos que são inapropriados e não para mostrar que determinado modelo é o correto para o ajuste dos dados em questão.

Uma das maneiras de fazer a análise de resíduos é por meio do resíduo de

Cox-Snell.

2.5.1 Resíduos de Cox-Snell

Os resíduos de Cox-Snell servem como recurso para avaliar o ajuste global do modelo. Eles são definidos como o valor negativo do logaritmo natural da probabilidade de sobrevivência para cada observação:

$$\hat{e}_i = \hat{H}(t_i|x_i) = -\log(\hat{S}(t_i|x_i))$$

Os resíduos $\hat{e}_i$ vêm de uma população homogênea e devem seguir uma distribuição exponencial padrão se o modelo for adequado (Lawless 2003). Graficamente, a análise pode ser feita de duas formas:

- Observando se o gráficos da função de sobrevivência dos resíduos estimados por Kaplan-Meier e os ajustados pelo modelo exponencial geram aproximadamente uma reta com inclinação 1.
- Observando se as curvas dos resíduos, tanto a empírica ajustada por Kaplan-Meier como a ajustada pelo modelo exponencial, estão próximas.

A Figura 3 ilustra um ajuste ideal para os dados:

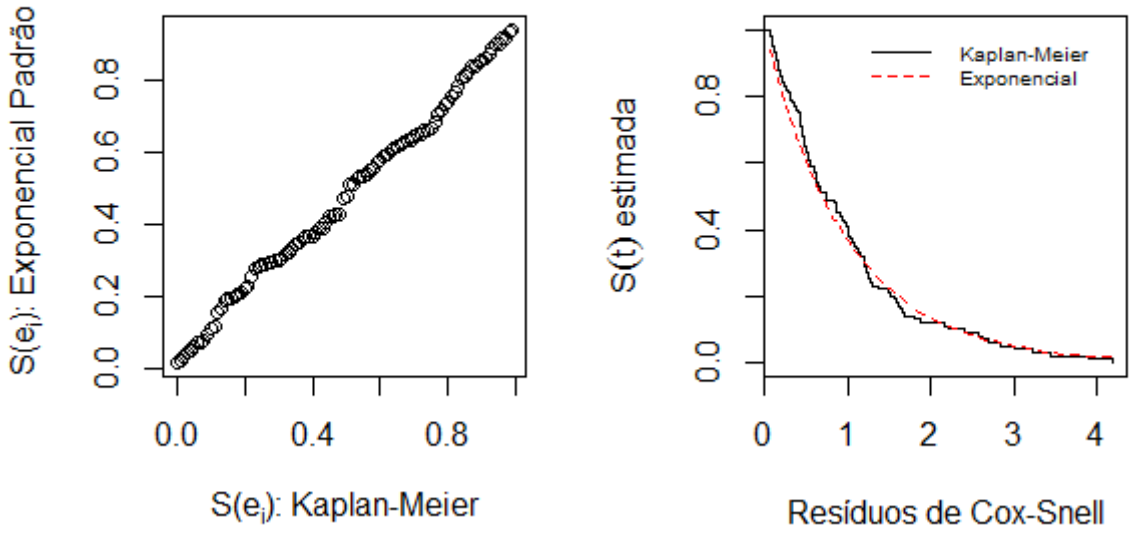

Figura 3: Exemplo do comportamento dos resíduos de Cox-Snell para um modelo ideal.

3 Metodologia

Para alcançar os objetivos propostos, serão buscados resultados com base no uso das técnicas apresentadas na revisão de literatura deste projeto.

3.1 Descrição dos dados

A base de dados que será utilizada neste trabalho diz respeito a difusão de políticas sociais no Brasil com foco no Programa Saúde da Família (PSF).

O banco de dados contém diversas informações a respeito de 4137 municípios, sendo elas de cunho político e social, que poderão, ao final desse estudo, justificar a adesão ao PSF para essas unidades de análise.

Os municípios serão estudados segundo dois tempos iniciais: 1991 e 1997. A utilização desses dois tempos tem por objetivo a apresentação de resultados segundo as ciências políticas e também sob o olhar estatístico do Programa Saúde da Família. Do ponto de vista político é interessante apresentar a análise a partir do corte feito no ano de 1997, pois assim, há informações completas para as eleições de 1996. Já do ponto de vista estatístico, é interessante apresentar os resultados considerando o tempo inicial em 1991, ano em que os municípios começaram a aderir ao Programa Saúde da Família.

Os resultados serão apresentados com o tempo em dias, sendo que, os tempos iniciais considerados serão 01/01/1991 e 01/01/1997 e o tempo final considerado para as duas análises será 31/12/2011. A variável resposta é o tempo até a adesão ao PSF. Como não há informações completas para as eleições de 1992, há o corte para 1997 para a análise sob o olhar político, pois existem informações completas a respeito das eleições de 1996. Logo, no ano seguinte, os resultados das eleições começam a influenciar a tomada de decisão dos governos.

Como o tempo de sobrevivência pode ser influenciado por alguns fatores, será apresentada a análise para 14 covariáveis. A Tabela 1 apresenta cada variável explicativa e sua descrição.

Tabela 1: Variáveis explicativas associadas aos municípios brasileiros.

Variável	Descrição da variável
NEP	Número efetivo de partidos
Margem de vitória	Margem de vitória do candidato que ficou em 1° lugar versus o 2° lugar
Percent_ganhador	Percentual de votos do candidato eleito
Alinhamento federal	Alinhamento partidário entre o governo municipal e o governo federal
Ano eleitoral	Se era ou não ano de eleições
Ideologia	Escala de 0 a 10 de posicionamento ideológico
População até 17 anos	Percentual de pessoas com idade abaixo de 17 anos
População acima de 65 anos	Percentual de pessoas com idade acima de 65 anos
Vizinho	Percentual de vizinhos de determinado município que adotaram ao PSF
Região Metropolitana	Se o município é ou não considerado região metropolitana
Taxa de urbanização	Percentagem da população urbana em relação à população total de determinado lugar
Capacidade fiscal média	Condição financeira que determinado governo possui
Região	Região de pertencimento do município
Porte populacional	Quantidade de habitantes do município

O sistema político é considerado competitivo, quando o candidato é eleito com menos de 45 por cento dos votos válidos, ou quando, por exemplo, a proporção da divisão de votos partidários é alta. Os municípios que tiveram a ocorrência de segundo turno também são considerados competitivos politicamente. As variáveis número efetivo de partidos, margem de vitória e percentual de votos do candidato eleito são variáveis de competição política.

A variável *número efetivo de partidos* faz o cálculo da parcela de votos que os partidos obtêm em uma eleição majoritária. Considera-se que uma eleição é muito competitiva quando a fragmentação de votos partidários é alta. Atribui-se que uma disputa eleitoral é pouco ou não competitiva quando a proporção da divisão de votos partidários é baixa (Coêlho et al.,2015)). A fórmula utilizada para o cálculo é a seguinte:

$$N_p = \sum_{i=1}^{x} \frac{1}{1 + \left(\frac{S_1^2}{S_i}\right) - S_i}. \tag{7}$$

Sendo :

- N_p, o indicador de competição político-partidária por ano eleitoral;
- S_1, a maior proporção na divisão de votos partidários;
- S_i, as demais componentes na divisão de votos partidários;
- x, a menor proporção na divisão de votos partidários.

Segundo Coêlho et al.(2015), alguns teóricos consideram que há indivíduos que tomam decisões segundo seus valores e crenças. Isso faz com que eles sejam mais receptivos às mudanças institucionais que tenham ênfase no social se forem comparados com outros indivíduos de perfil conservador.

Para a variável ideologia são utilizadas as medidas de posicionamento ideológico de Power e Zucco (2008), resultantes das estimativas das pesquisas de survey realizadas sobre a percepção dos congressistas acerca do posicionamento ideológico de seus partidos. A escala varia de 0 (extrema esquerda) a 10 (extrema direita).(Coêlho et al.,2015)

A variável *alinhamento federal* é definida da seguinte forma:

$$Alinhamento = \begin{cases} 0, & \text{se o partido não é alinhado politicamente com o governo federal,} \\ 1, & \text{se o partido é alinhado politicamente com o governo federal.} \end{cases}$$

A variável *ano eleitoral* é definida de forma semelhante, assumindo o valor 0 para os municípios que aderiram ao PSF em anos em que não ocorreram eleições municipais e 1 para os que adotaram o programa em anos de eleições municipais.

Região metropolitana também é definida de forma parecida, assumindo 0 se o município não é considerado região metropolitana e 1 se ele é região metropolitana.

A variável *vizinho* define o número de municípios por estado, que adotou o PSF por ano, e estima se adoções posteriores estão correlacionadas com o aumento percentual de adoções anteriores. Assim, é utilizada a proporção de cidades em cada estado que aderiu ao PSF em anos anteriores aos novos casos de adoção.

As variáveis NEP, margem de vitória, percentual de votos para o candidato eleito, ideologia, população até 17 anos, população acima de 65 anos e taxa de urbanização foram medidas nos anos de eleições: 1996,2000,2004 e 2008. Para definir qual informação usar em cada uma destas variáveis, foi levado em conta o ano em que o município aderiu ao PSF. Por exemplo, se um município falhou em 1998, a informação utilizada foi referente à última eleição antes da data de adesão, neste caso, 1996.

A seguir, serão apresentados os métodos utilizados para alcançar os objetivos propostos.

3.2 Métodos

Para iniciar o estudo será feita a análise descritiva da base de dados utilizada. O estimador mais usado para a análise inicial deste tipo de dados, que neste caso busca investigar o tempo até a adesão ao Programa Saúde da Família, é o de Kaplan-Meier, conforme exposto na Revisão de Literatura. Dessa forma, serão apresentados, primeiramente, os gráficos da função de sobrevivência estimada e função de risco acumulada, ambas segundo Kaplan-Meier e a curva do tempo total em teste (TTT) do tempo de sobrevivência. Em seguida, serão apresentadas as curvas de sobrevivência para as variáveis qualitativas.

O segundo passo será a definição de possíveis modelos de probabilidade. Após serem feitas as estimativas utilizando Kaplan-Meier e a partir, também, do comportamento observado da função de risco acumulada e TTT, é possível sugerir modelos que possam apresentar um bom ajuste para os dados.

Após a definição das possíveis distribuições de probabilidade, será feito o estudo conjunto do tempo até a adesão ao PSF e suas covariáveis. A análise que será realizada a partir de modelos de regressão definidos na seçãp 2.4 e a estimação dos parâmetros desses modelos será feita pelo método de máxima verossimilhança. Esses modelos irão mostrar a influência dessas covariáveis na variável resposta e então, dentre eles, será escolhido o mais adequado para apresentar os efeitos que os fatores políticos, institucionais e regionais apresentam na decisão dos governos municipais para a adesão ao Programa Saúde da Família.

Em seguida, será feita a validação do ajuste do(s) modelo(s) escolhido(s) por meio do resíduo de Cox-Snell definido na seção 2.5.

4 Resultados e Discussões

4.1 Análise Descritiva

A análise se inicia com os resultados exploratórios das variáveis em estudo.

A função de sobrevivência estimada segundo Kaplan-Meier para os dados com tempo incial em 01/01/1991 e 01/01/1997 é apresentada na Figura 3. Naturalmente, a probabilidade de sobrevida decai com o aumento do tempo de estudo. É importante ressaltar que nos dados desta análise não há censuras, ou seja, ao longo do tempo de estudo, todos os municípios sob observação aderiram ao Programa Saúde da Família em algum momento. É possível observar que, mesmo com escalas de tempo diferentes, o comportamento observado para a função de sobrevivência é semelhante em ambos os casos.

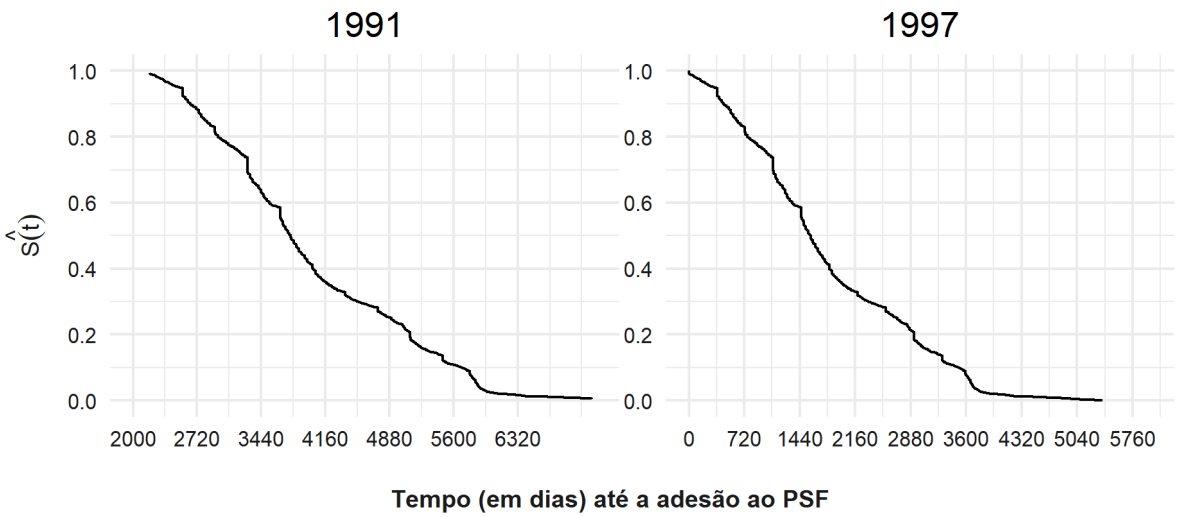

Figura 4: Função de sobrevivência estimada segundo a metodologia de Kaplan-Meier.

Como definido nas seções 2.1.5 e 2.1.6, as metodologias da função de risco acumulada e do Tempo Total em Teste serão utilizadas para auxiliar na escolha de possíveis distribuições de probabilidade que podem ser usadas para modelar o tempo até a adesão ao PSF.

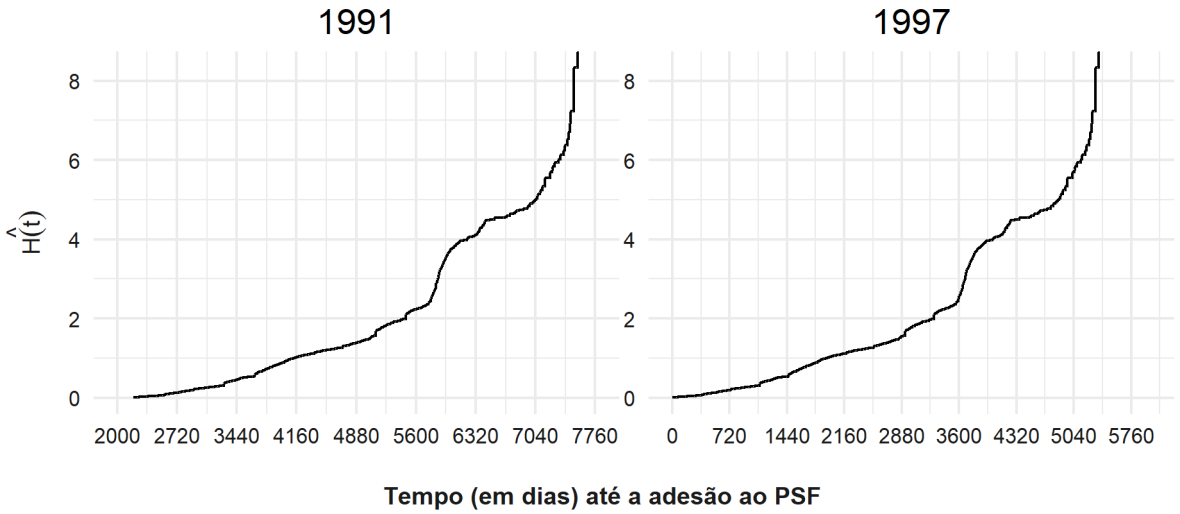

Figura 5: Função de risco acumulada segundo a metodologia de Kaplan-Meier.

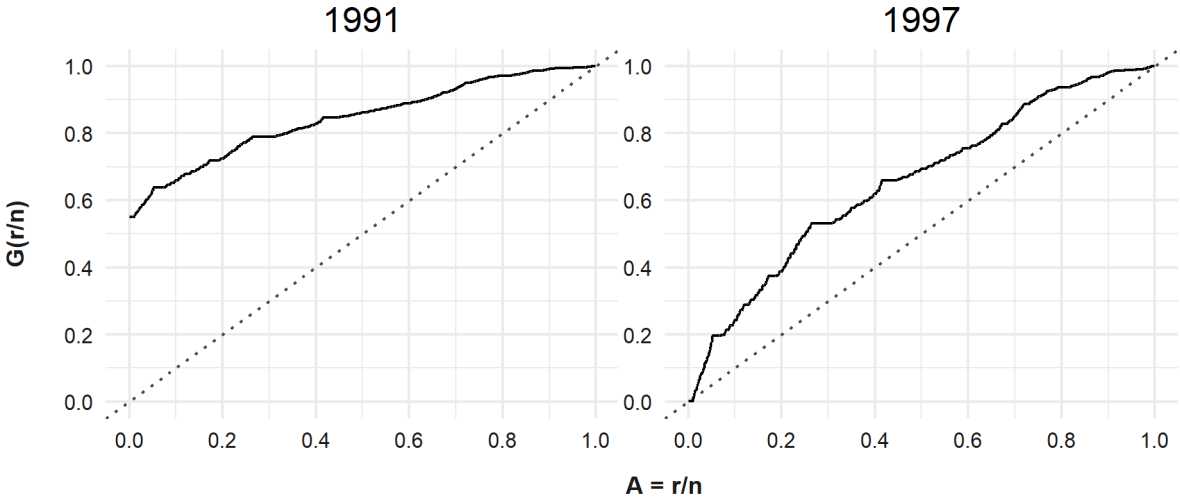

Figura 6: Curva do Tempo Total em Teste (TTT) do tempo até a adesão ao PSF.

Por meio dos resultados apresentados nas Figuras 5 e 6, observa-se que os dados em estudo apresentam comportamento da função de risco crescente. Dessa forma, dentre as distribuições definidas na seção 2.2, a distribuição weibull é a mais indicada.

Porém, como as distribuições Log-Normal e Log-Logística também apresentam bons resultados em análise de dados que a distribuição Weibull é candidata possível, este trabalho propõe utilizar essas três distribuições de probabilidade e fazer um estudo comparativo entre elas para definir qual distribuição é a mais indicada para a análise desses dados.

A próxima etapa da análise descritiva é o estudo individual de cada covariável com a variável resposta por meio da função de sobrevivência estimada por Kaplan-

Meier.

A primeira covariável a ser analisada é *região*,como pode ser visto na Figura 7. Inicialmente, todas as curvas apresentam um comportamento semelhante e, como esperado, com o passar do tempo, a probabilidade de adesão ao Programa Saúde da Família decaiu. Nota-se, ao longo do tempo analisado, que a região Norte tem uma curva mais distante das demais, o que é um indício de que os municípios pertencentes a esta região demoraram mais para adotarem o PSF. A região Centro-Oeste,tanto considerando o tempo inicial como 1991 ou 1997, após certo tempo, apresenta um decaimento considerável na sua curva em relação às demais, o que significa dizer que, provavelmente, na região Centro-Oeste, o tempo até a adesão ao PSF foi menor que nas outras regiões.

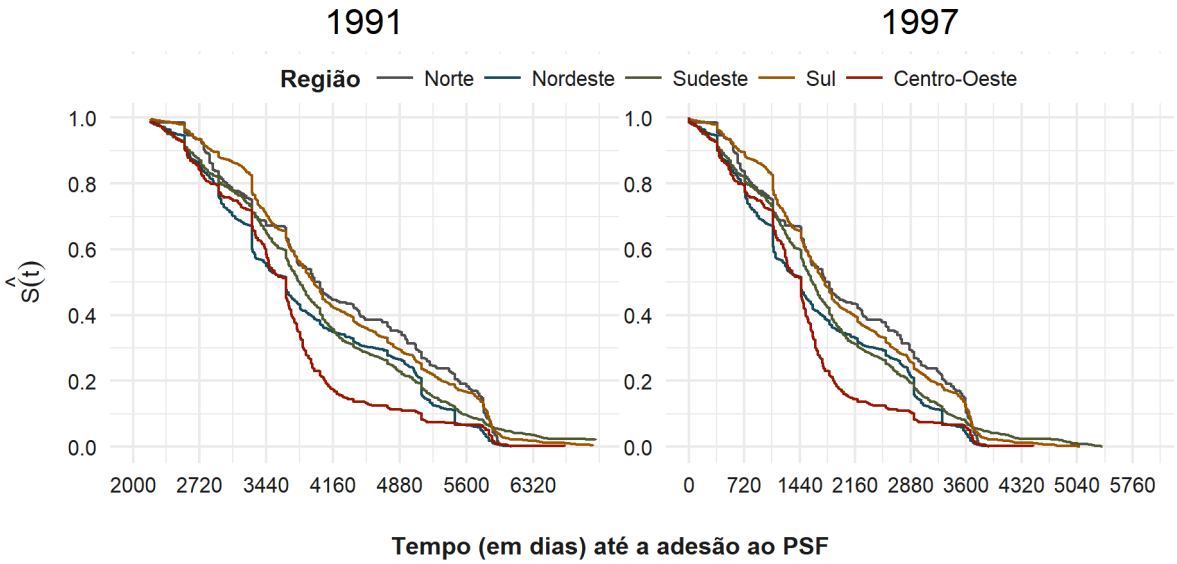

Figura 7: Função de sobrevivência estimada pela metodologia de Kaplan-Meier segundo a covariável Região.

Na variável *porte*, Figura 8, observa-se que a curva referente aos municípios com mais de 500 mil habitantes decai mais rapidamente que as demais. Consequentemente, ela chega a zero antes das outras. Dessa forma, a adesão ao PSF pelos municípios com mais de 500 mil habitantes acontece de forma mais acelerada em relação aos municípios de menor porte. Esse comportamento é observado tanto para o tempo inicial em 1991 quanto para 1997.

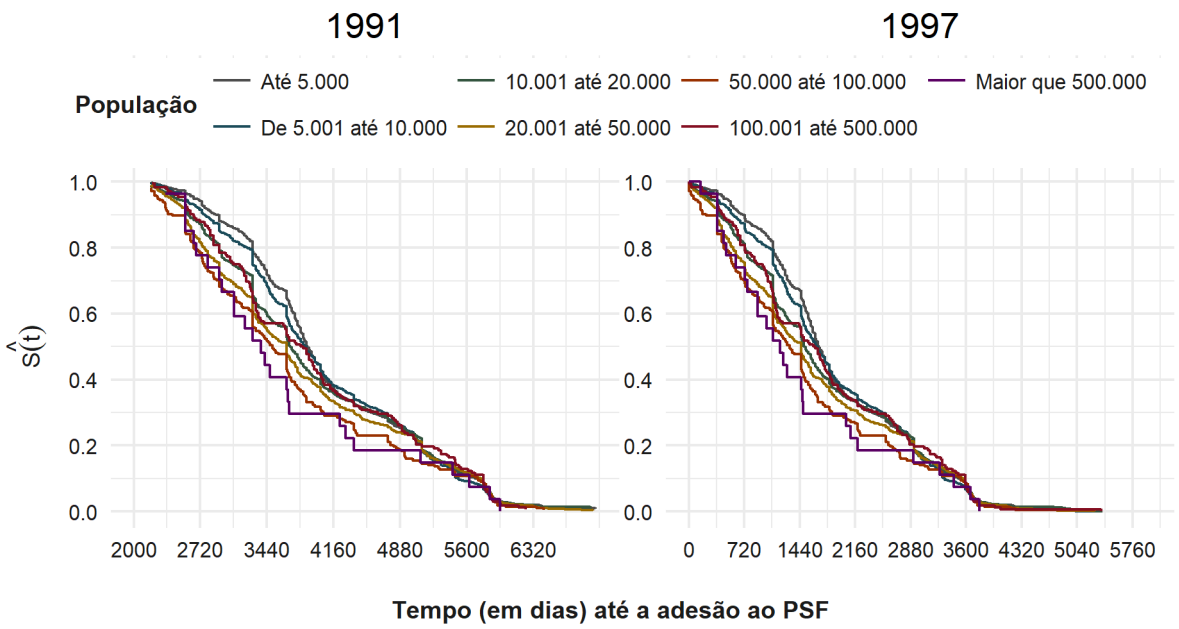

Figura 8: Função de sobrevivência estimada pela metodologia de Kaplan-Meier segundo a covariável Porte.

Segundo Cox & McCubins (1986), a convergência partidária entre o governo federal e os governos municipais tem por consequência uma maior cooperação para a implementação de políticas públicas. Pela Figura 9 é possível observar uma grande proximidade nas curvas para os municípios com e sem alinhamento. Porém, ainda assim, os municípios com alinhamento federal aderem mais rapidamente que os demais.

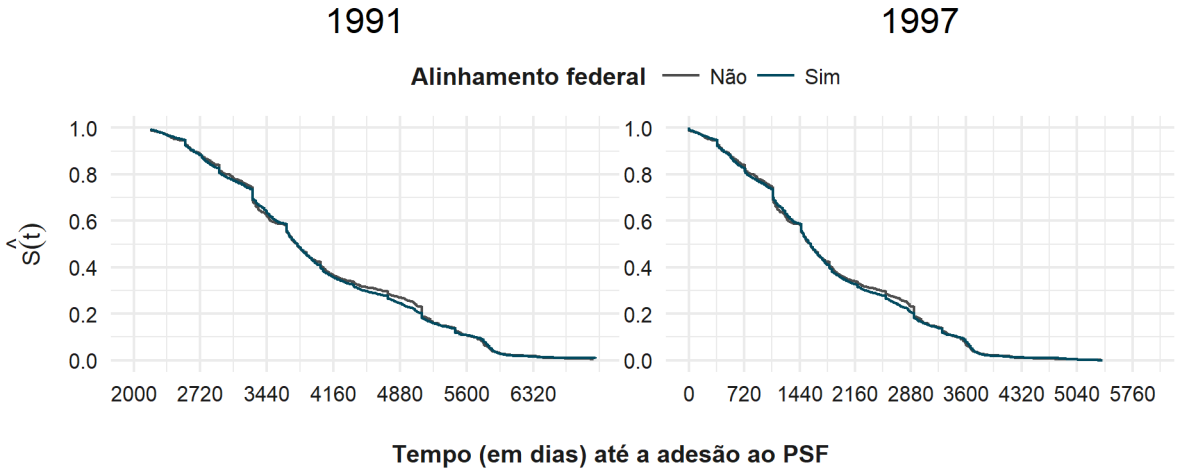

Figura 9: Função de sobrevivência estimada pela metodologia de Kaplan-Meier segundo a covariável Alinhamento Federal.

Como esperado, ao analisar a variável *ano eleitoral*, Figura 10, nota-se que, em anos eleitorais, tanto para o ano inicial 1991 quanto para 1997, a probabilidade de

sobrevivência decai bem mais rapidamente que para anos não eleitorais, o que pode ser um indício de que essa covariável é importante para a modelagem.

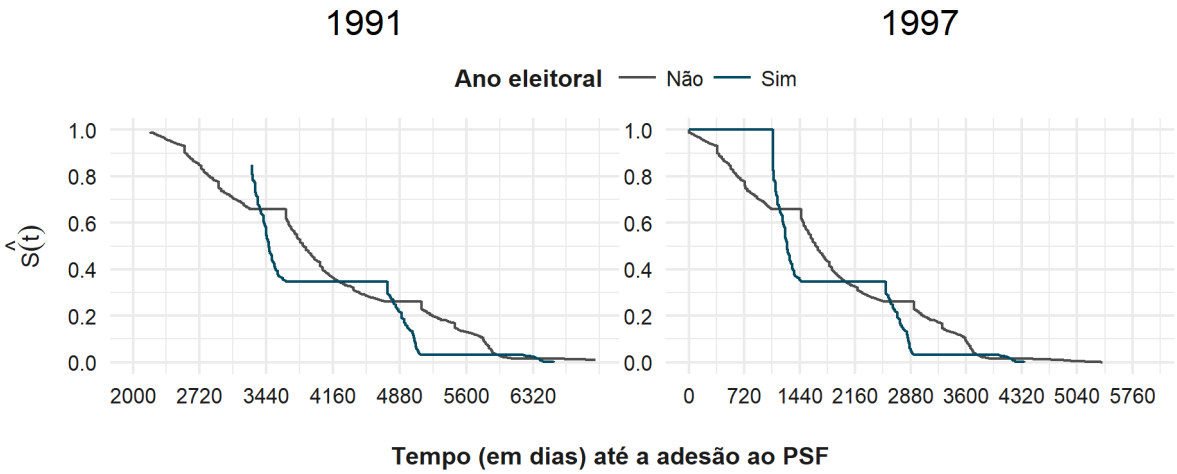

Figura 10: Função de sobrevivência estimada pela metodologia de Kaplan-Meier segundo a covariável Ano Eleitoral.

Ao observar a Figura 11, *região metropolitana*, nota-se que o decaimento das curvas para o tempo inicial 1991 e também para 1997 se apresenta de forma semelhante, porém esse decaimento para os municípios que não são considerados região metropolitana se dá de forma um pouco mais rápida que para os demais.

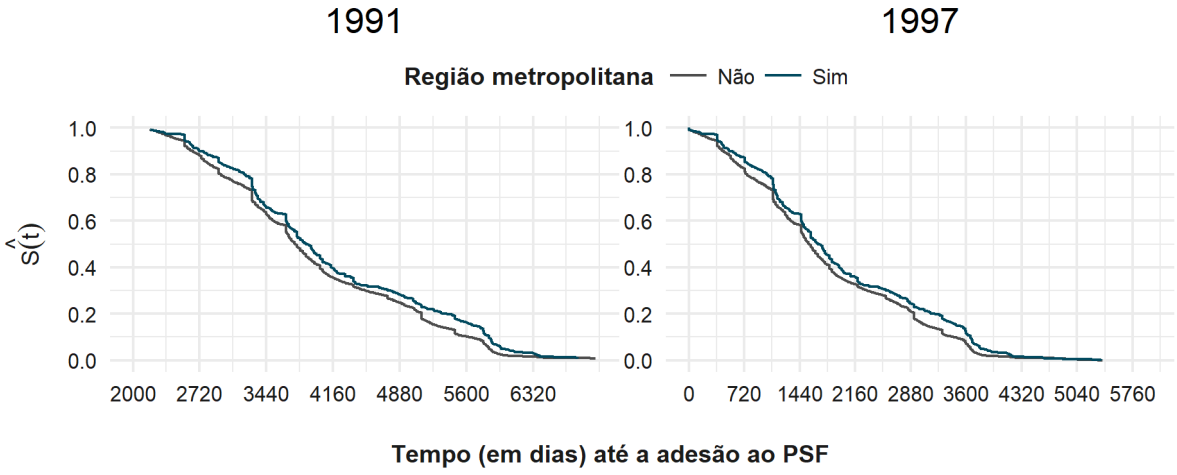

Figura 11: Função de sobrevivência estimada pela metodologia de Kaplan-Meier segundo a covariável Região Metropolitana.

4.2 Definição da distribuição de probabilidades

Como apresentado nas Figuras 5 e 6, será mostrada a comparação entre as curvas de sobrevivência para as distribuições Weibull, Log-Normal e Log-Logística

em relação à curva de sobrevivência estimada por Kaplain-Meier para os dados em estudo.

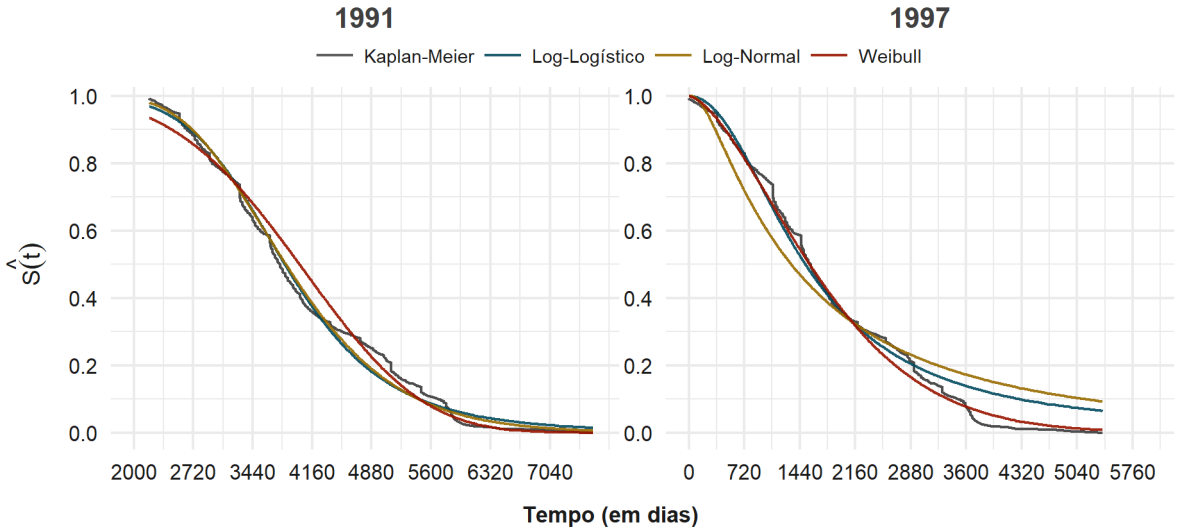

Figura 12: Comparação entre a curva estimada segundo a metodologia de Kaplan-Meier e as funções de sobrevivência estimadas das distribuições Weibull, Log-Normal e Log-Logística.

A Figura 12 mostra, para o tempo inicial 1991, que, no início da curva, a distribuições Log-Normal e Log-Logística se adequam melhor. Porém, no final, a distribuição Weibull é a que melhor se adapta. Para o ano inicial 1997, a distribuição Weibull se adequa à curva de Kaplain-Meier de forma quase perfeita e a distribuição Log-Logística se adequa em boa parte do tempo. Sendo assim, para finalidade de comparação, a modelagem será feita baseada nos modelos Weibull e Log-Logístico.

4.3 Análise de regressão

Pela análise anterior, as distribuições Weibull e Log-Logística foram definidas para a modelagem. Para a criação dos modelos de regressão, se fez necessário a inclusão de variáveis *dummy* para as covariáveis *região* e *porte*. A categoria de referência para as duas variáveis foi definida a partir do estudo exploratório das mesmas e as curvas mais distantes das demais, para as duas covariáveis, foram as categorias definidas como referência.

A variável *dummy* indica presença ou ausência de uma característica pré-definida na observação em análise. Se uma variável é formada por *k* categorias, então ela terá *k-1* variáveis *dummy*. A variável *região* se divide nas categorias Norte, Nordeste, Sudeste, Sul e Centro-Oeste. Definiu-se como referência a região *Norte* , classificando as demais da seguinte forma:

$$D_{i1} = \begin{cases} 1, & \text{se o município pertence à região Nordeste}, \\ 0, & \text{c.c.} \end{cases}$$

$$D_{i2} = \begin{cases} 1, & \text{se o município pertence à região Sudeste}, \\ 0, & \text{c.c.} \end{cases}$$

$$D_{i3} = \begin{cases} 1, & \text{se o município pertence à região Sul}, \\ 0, & \text{c.c.} \end{cases}$$

$$D_{i4} = \begin{cases} 1, & \text{se o município pertence à região Centro-Oeste}, \\ 0, & \text{c.c.} \end{cases}$$

De maneira análoga, tendo a categoria *porte populacional acima de 500 mil* como referencia, foram criadas variáveis *dummy* para as demais categorias da covariável *porte*.

Dessa forma, serão feitos os modelos completos considerando as duas datas iniciais e as duas distribuições propostas e o nível de significância considerado será de 10%.

4.4 Definição dos modelos de regressão para tempo inicial 01/01/1991.

A análise será iniciada com a modelagem considerando o tempo inicial como 1991. A Tabela 2 apresenta as estimativas de máxima verossimilhança, erro padrão e p-valor dos parâmetros do modelo de regressão Weibull completo. Ou seja, primeiramente foi feita a estimação do modelo de regressão Weibull considerando todas as variáveis explicativas disponíveis no estudo.

Tabela 2: Estimativas para os parâmetros do modelo de regressão completo Weibull.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	9,4787	0,0952	<0,0001
NEP	0,0110	0,0091	0,2230
Margem de Vitória	-0,0250	0,0262	0,3400
Percent_ganhador	0,0346	0,0382	0,3650
Alinhamento federal	-0,0301	0,0088	0,0007
Ano eleitoral	-0,0263	0,0088	0,0027
Ideologia	0,0025	0,0019	0,2060
População até 17 anos	-2,3752	0,1151	<0,0001
População acima de 65 anos	-0,4543	0,3208	0,1570
Vizinho	-0,0010	0,0003	<0,0001
Região Metropolitana	0,0611	0,0142	<0,0001
Taxa de Urbanização	0,0177	0,0203	0,3810
Capacidade fiscal média	-0,2227	0,0590	0,0002
Nordeste	-0,1779	0,0178	<0,0001
Sudeste	-0,3307	0,0193	<0,0001
Sul	-0,3358	0,0208	<0,0001
Centro-Oeste	-0,4097	0,0213	<0,0001
Porte populacional abaixo de 5 mil	0,1460	0,0531	0,0059
Porte populacional entre 5 mil e 10 mil	0,1769	0,0523	0,0007
Porte populacional entre 10 mil e 20 mil	0,1706	0,0516	0,0009
Porte populacional entre 20 mil e 50 mil	0,1456	0,0507	0,0041
Porte populacional entre 50 e 100 mil	0,1054	0,0513	0,0400
Porte populacional entre 100 mil e 500 mil	0,1116	0,0505	0,0272
Gamma (γ)	4,3	-	-

A Figura 13 apresenta a análise gráfica do resíduo de Cox-Snell para o modelo de regressão Weibull completo.

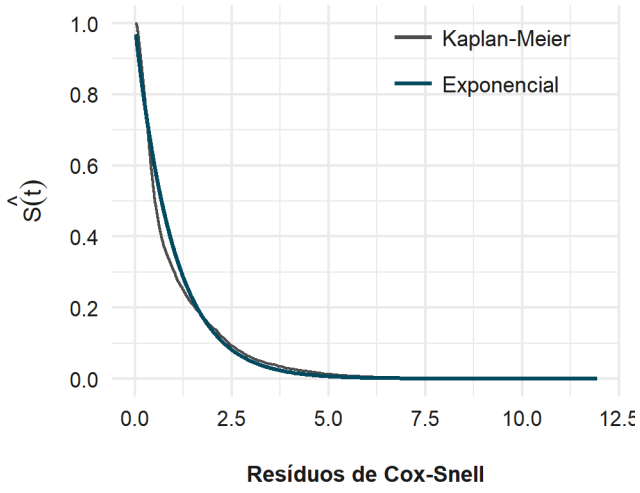

Figura 13: Resíduos de Cox-Snell para o modelo de regressão Weibull com todas as variáveis explicativas.

Por meio da Figura 13, observa-se que o modelo se ajusta bem aos dados. Dessa forma, pode-se interpretar os resultados apresentados na Tabela 2.

Ao considerar o nível de significância de 10%, nota-se que, as covariáveis significantes para o modelo completo são: *alinhamento federal, ano eleitoral, população até 17 anos, vizinho, região metropolitana, capacidade fiscal média, região e porte*. Em geral, os resultados das estimativas dos parâmetros apresentam valores baixos no modelo completo.

As categorias da variável *região*, em relação a região Norte, categoria de referência, possuem probabilidade de sobrevivência menor, o que confirma a análise descritiva: a região Norte demora mais tempo que as demais para aderir ao PSF. Nota-se também, que a menor estimativa observada é a da região Centro-Oeste, o que significa que ela é a região que adere mais rapidamente ao programa sob análise, resultado que também é observado na análise descritiva dessa covariável.

Já para porte, todos os municípios com menos de 500 mil habitantes possuem maior probabilidade de sobrevivência em relação à categoria de referência, o que significa dizer que, os municípios com mais de 500 mil habitantes falham mais rapidamente. Esse resultado confirma a análise descritiva feita para essa covariável apresentada na Figura 8.

Ser ano eleitoral, ter alinhamento federal, ter população abaixo de 17 anos, ter em sua vizinhança municípios que adotaram o PSF, não ser região metropolitana e ter alta capacidade fiscal neste modelo, significa dizer que o município irá aderir mais rapidamente ao PSF.

Após obter as estimativas para o modelo completo, é feita a seleção de variáveis por meio do método Backward, que consiste em retirar a covariável menos significante do modelo e fazê-lo novamente. Com as novas estimativas, retira-se novamente a covariável menos significante e o mesmo processo é repetido até que todas as covariáveis presentes sejam significativas.

Ao realizar a seleção de variáveis, o modelo de regressão Weibull final está exposto na Tabela 3.

Tabela 3: Estimativas para os parâmetros do modelo final de regressão Weibull.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	9,4536	0,0723	<0,0001
NEP	0,0125	0,0074	0,0910
Alinhamento federal	-0,0292	0,0088	0,0009
Ano eleitoral	-0,0270	0,0088	0,0021
População até 17 anos	-2,2983	0,0797	<0,0001
Vizinho	-0,0011	0,0003	<0,0001
Região Metropolitana	0,0645	0,0141	<0,0001
Capacidade fiscal média	-0,1990	0,0571	0,0005
Nordeste	-0,1799	0,0175	<0,0001
Sudeste	-0,3306	0,0193	<0,0001
Sul	-0,3358	0,0205	<0,0001
Centro-Oeste	-0,4017	0,0207	<0,0001
Porte populacional abaixo de 5 mil	0,1460	0,0523	0,0053
Porte populacional entre 5 mil e 10 mil	0,1786	0,0515	0,0005
Porte populacional entre 10 mil e 20 mil	0,1728	0,0508	0,0007
Porte populacional entre 20 mil e 50 mil	0,1494	0,0500	0,0028
Porte populacional entre 50 e 100 mil	0,1118	0,0507	0,0274
Porte populacional entre 100 mil e 500 mil	0,1185	0,0500	0,0179
Gamma (γ)	4,3	-	-

A Figura 14 apresenta a análise gráfica do resíduo de Cox-Snell para o modelo de regressão Weibull final.

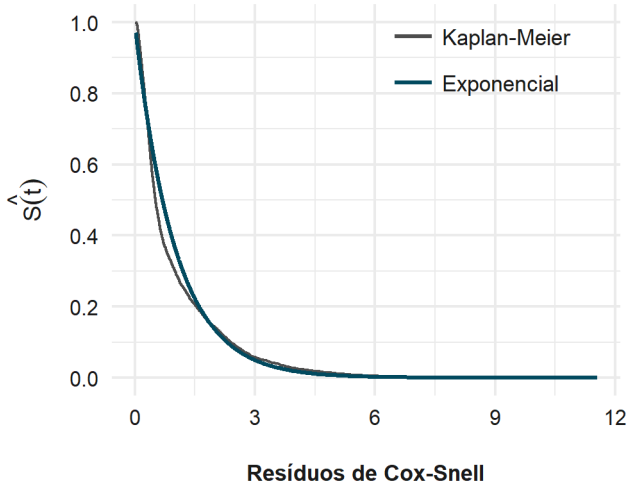

Figura 14: Resíduos de Cox-Snell para o modelo final de regressão Weibull.

Nota-se um bom ajuste do modelo aos dados. Assim, é possível analisar os resultados apresentados na Tabela 3.

Observa-se agora, como definido pelo método Backward, que todas as covariáveis são significativas ao nível de 10% de significância para o novo modelo. *NEP*

é a única covariável que não foi significante no modelo completo e que agora, com a nova modelagem, entra no grupo de covariáveis significativas. Quanto maior o número efetivo de partidos, mais demorada será a adesão ao PSF pelo município. Nota-se também que não houve muita mudança nos valores do erro padrão para ambos modelos, o que indica um bom ajuste do modelo aos dados. Isto é comprovado ao observar o gráfico dos resíduos de Cox-Snell que apresenta um bom ajuste aos dados, semelhante ao ajuste gerado pelo modelo completo. Vale ressaltar que, o valor do parâmetro de forma é maior que 1, $\gamma = 4,3$, o que confirma, para a distribuição Weibull, que a função de risco acumulada é estritamente crescente.

Agora, ao considerar que os tempos até a adesão ao PSF podem ser modelados pela distribuição Log-Logística, as estimativas dos parâmetros, erro padrão e p-valor do modelo de regressão Log-Logístico com todas as variáveis explicativas estão na Tabela 4 e seu resíduo de Cox-Snell, na Figura 15.

Tabela 4: Estimativas para os parâmetros do modelo de regressão completo Log-Logístico.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	9,5176	0,1026	<0,0001
NEP	0,0258	0,0097	0,0082
Margem de Vitória	-0,0520	0,0342	0,1280
Percent_ganhador	0,1187	0,0627	0,0585
Alinhamento federal	-0,0327	0,0092	0,0004
Ano eleitoral	0,0510	0,0085	<0,0001
Ideologia	0,0063	0,0020	0,0021
População até 17 anos	-2,9159	0,1147	<0,0001
População acima de 65 anos	-0,2100	0,3181	0,5090
Vizinho	-0,0009	0,0003	0,0007
Região Metropolitana	0,0661	0,0142	<0,0001
Taxa de Urbanização	-0,0278	0,0216	0,1980
Capacidade fiscal média	-0,3522	0,0669	<0,0001
Nordeste	-0,1701	0,0193	<0,0001
Sudeste	-0,3597	0,0196	<0,0001
Sul	-0,3509	0,0204	<0,0001
Centro-Oeste	-0,3716	0,0207	<0,0001
Porte populacional abaixo de 5 mil	0,0929	0,0551	0,0915
Porte populacional entre 5 mil e 10 mil	0,1415	0,0541	0,0089
Porte populacional entre 10 mil e 20 mil	0,1387	0,0533	0,0093
Porte populacional entre 20 mil e 50 mil	0,1172	0,0522	0,0248
Porte populacional entre 50 e 100 mil	0,0583	0,0523	0,2660
Porte populacional entre 100 mil e 500 mil	0,1207	0,0513	0,0186
Gamma (γ)	7,35	-	-

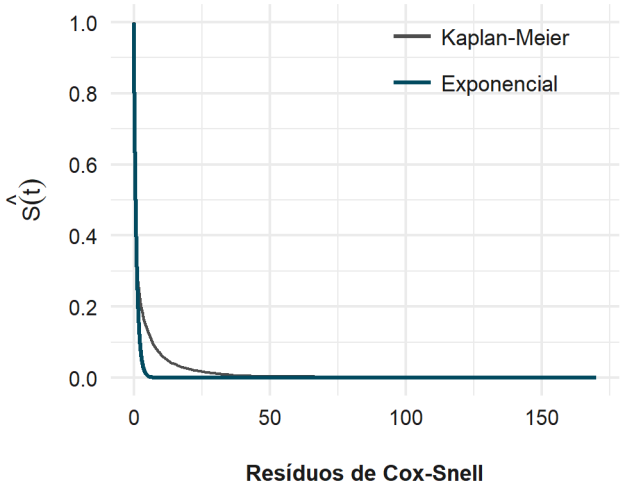

Figura 15: Resíduos de Cox-Snell para o modelo de regressão Log-Logístico com todas as variáveis explicativas.

É possível notar uma boa adequação do modelo aos dados. A seguir será apresentada a análise dos parâmetros estimados para a distribuição Log-Logística.

Observando os resultados gerados pela modelagem feita a partir da distribuição Log-Logística, tem-se que as variáveis significativas para o modelo a um nível de significância de 10% são: *NEP*, *percentual de votos do eleito*, *alinhamento federal*, *ano eleitoral*, *ideologia*, *população até 17 anos*, *vizinhos que aderiram ao PSF*, *região metropolitana*, *capacidade fiscal média*, *região* e *porte*. As variáveis *ano eleitoral* e *taxa de urbanização* obtiveram estimativas com o sinal invertido em relação ao modelo Weibull tendo o mesmo ano inicial de referência. Trocando assim, a forma de interpretação dessas covariáveis.

Aqui, o modelo final também é obtido pelo método de seleção de variáveis Backward. E as estimativas dos parâmetros, erro padrão e p-valor do modelo de regressão Log-Logístico final são apresentados na Tabela 5.

Tabela 5: Estimativas para os parâmetros do modelo final de regressão Log-Logístico.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	9,5337	0,0748	<0,0001
NEP	0,0175	0,0073	0,0160
Alinhamento federal	-0,0324	0,0092	0,0004
Ano eleitoral	0,0503	0,0085	<0,0001
Ideologia	0,0063	0,0020	0,0023
População até 17 anos	-2,8375	0,0784	<0,0001
Vizinho	-0,0009	0,0003	0,0010
Região Metropolitana	0,0659	0,0142	<0,0001
Capacidade fiscal média	-0,3678	0,0656	<0,0001
Nordeste	-0,1719	0,0187	<0,0001
Sudeste	-0,3608	0,0196	<0,0001
Sul	-0,3464	0,0202	<0,0001
Centro-Oeste	-0,3711	0,0203	<0,0001
Porte populacional abaixo de 5 mil	0,0926	0,0546	0,0900
Porte populacional entre 5 mil e 10 mil	0,1400	0,0536	0,0091
Porte populacional entre 10 mil e 20 mil	0,1360	0,0529	0,0102
Porte populacional entre 20 mil e 50 mil	0,1126	0,0519	0,0300
Porte populacional entre 50 e 100 mil	0,0515	0,0521	0,3230
Porte populacional entre 100 mil e 500 mil	0,1149	0,0512	0,0248
Gamma (γ)	7,35	-	-

O gráfico de resíduos de Cox-Snell apresenta uma adequação aos dados semelhante a que foi apresentada pelo modelo de regressão Log-Logístico completo, como mostra a Figura 16.

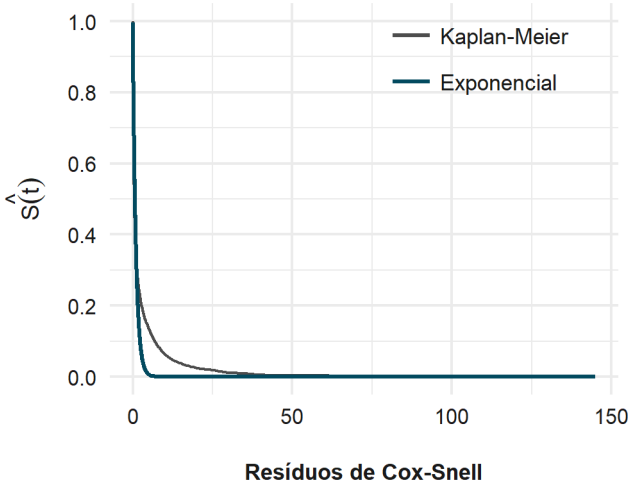

Figura 16: Resíduos de Cox-Snell para o modelo final de regressão Log-Logístico.

Nota-se que os municípios com mais de 500 mil habitantes aderem mais rapidamente ao Programa Saúde da Família. A região Norte adota o programa mais lentamente que as demais regiões. Municípios com maior capacidade fiscal têm mais

facilidade para aderir ao programa, bem como municípios que não são regiões metropolitanas, que possuem vizinhos que aderiram ao programa, os com alto percentual de população até 17 anos, com valores mais baixos de ideologia (esquerda), em anos não eleitorais, com alinhamento federal e com baixo número efetivo de partidos. Nota-se também que, o valor do parâmetro de forma é maior que 1, $\gamma = 7,35$. Isso confirma, para a distribuição Log-Logística, que a função de risco acumulada é unimodal.

4.5 Definição dos modelos de regressão para tempo inicial 01/01/1997.

Agora serão apresentados os modelos de regressão ao considerar a data inicial como 01/01/1997. A Tabela 6 apresenta o modelo de regressão Weibull completo.

Tabela 6: Estimativas para os parâmetros do modelo de regressão completo Weibull.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	10,1778	0,2407	<0,0001
NEP	0,0254	0,0229	0,2680
Margem de Vitória	-0,0590	0,0674	0,3810
Percent_ganhador	0,1136	0,1008	0,2600
Alinhamento federal	-0,0716	0,0223	0,0013
Ano eleitoral	0,0049	0,0222	0,8250
Ideologia	0,0090	0,0048	0,0628
População até 17 anos	-5,5974	0,2910	<0,0001
População acima de 65 anos	-0,9838	0,8051	0,2220
Vizinho	-0,0025	0,0007	0,0002
Região Metropolitana	0,1355	0,0357	0,0001
Taxa de Urbanização	-0,0026	0,0511	0,9600
Capacidade fiscal média	-0,6041	0,1485	<0,0001
Nordeste	-0,4013	0,0449	<0,0001
Sudeste	-0,7704	0,0484	<0,0001
Sul	-0,7746	0,0520	<0,0001
Centro-Oeste	-0,9151	0,0534	<0,0001
Porte populacional abaixo de 5 mil	0,2789	0,1337	0,0369
Porte populacional entre 5 mil e 10 mil	0,3593	0,1318	0,0064
Porte populacional entre 10 mil e 20 mil	0,3349	0,1300	0,0100
Porte populacional entre 20 mil e 50 mil	0,2816	0,1279	0,0276
Porte populacional entre 50 e 100 mil	0,1852	0,1293	0,1520
Porte populacional entre 100 mil e 500 mil	0,2424	0,1271	0,0565
Gamma (γ)	1,7	-	-

A adequação do modelo aos dados é apresenta na Figura 17 e nota-se um bom ajuste, como esperado.

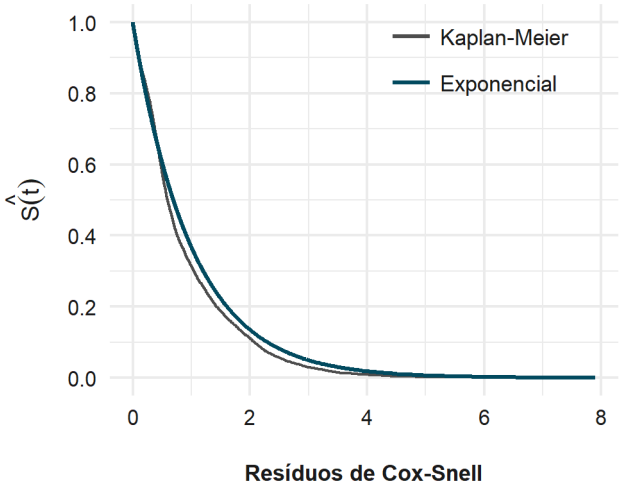

Figura 17: Resíduos de Cox-Snell para o modelo de regressão Weibull com todas as variáveis explicativas.

Aqui, bem como para o tempo inicial 1991, a região Norte adere ao PSF mais lentamente que as demais e os municípios com mais de 500 mil habitantes aderem mais rapidamente ao programa. Municípios com maior capacidade fiscal, que não são regiões metropolitanas, que possuem vizinhos que aderiram ao programa, os com alto percentual de população até 17 anos, com valores mais baixos de ideologia (esquerda) e com alinhamento federal têm mais facilidade para aderir ao programa.

Para definir o modelo de regressão Weibull final, a metodologia utilizada é a mesma apresentada anteriormente para a seleção de variáveis. Os resultados gerados pelo método Backward estão expostos na Tabela 7.

Tabela 7: Estimativas para os parâmetros do modelo final de regressão Weibull.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	10,1783	0,1690	<0,0001
Alinhamento federal	-0,0737	0,0222	0,0009
Ideologia	0,0084	0,0048	0,0813
População até 17 anos	-5,3656	0,2010	<0,0001
Vizinho	-0,0026	0,0006	<0,0001
Região Metropolitana	0,1425	0,0354	<0,0001
Capacidade fiscal média	-0,5690	0,1429	<0,0001
Nordeste	-0,4099	0,0439	<0,0001
Sudeste	-0,7690	0,0479	<0,0001
Sul	-0,7704	0,0510	<0,0001
Centro-Oeste	-0,9013	0,0519	<0,0001
Porte populacional abaixo de 5 mil	0,2300	0,1280	0,0724
Porte populacional entre 5 mil e 10 mil	0,3135	0,1262	0,0130
Porte populacional entre 10 mil e 20 mil	0,2891	0,1246	0,0203
Porte populacional entre 20 mil e 50 mil	0,2388	0,1227	0,0516
Porte populacional entre 50 e 100 mil	0,1480	0,1246	0,2350
Porte populacional entre 100 mil e 500 mil	0,2102	0,1234	0,0883
Gamma (γ)	1,7	-	-

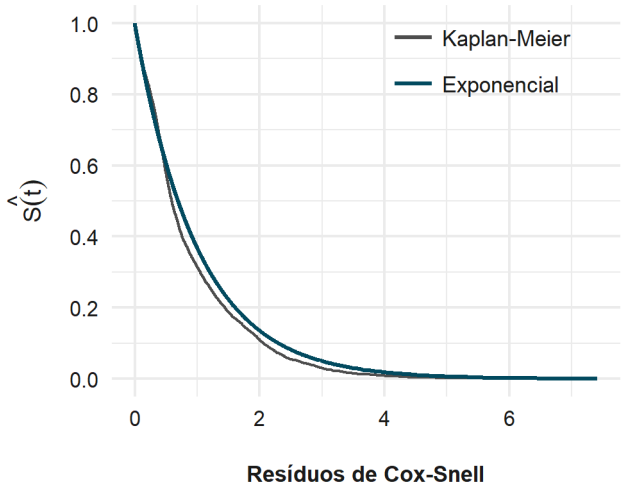

Figura 18: Resíduos de Cox-Snell para o modelo final de regressão Weibull.

Assim como no modelo de regressão Weibull completo, o modelo de regressão Weibull final se ajusta bem aos dados. Isso pode ser observado na Figura 18.

Nota-se aqui que as covariáveis significativas no modelo de regressão Weibull completo foram as mesmas selecionadas para o modelo de regressão Weibull final. É possível observar também que, o valor do parâmetro de forma maior que 1, $\gamma = 1,7$, mostra, para a distribuição Weibull, que a função de risco acumulada é estritamente crescente, como sugerido pelas Figuras 5 e 6.

Agora serão apresentados os resultados para a modelagem feita com base na distribuição Log-Logística para o ano inicial 1997. O modelo de regressão Log-Logístico completo é apresentado na tabela 8 e seu resíduo na Figura 19.

Tabela 8: Estimativas para os parâmetros do modelo de regressão completo Log-Logístico.

Variável	**Estimativa**	**Erro padrão**	**P-valor**
Intercepto	10,8084	0,2939	<0,0001
NEP	0,0510	0,0275	0,0644
Margem de Vitória	-0,1074	0,0927	0,2470
Percent_ganhador	0,2465	0,1648	0,1350
Alinhamento federal	-0,0921	0,0265	0,0005
Ano eleitoral	0,2034	0,0246	<0,0001
Ideologia	0,0190	0,0059	0,0013
População até 17 anos	-7,8644	0,3372	<0,0001
População acima de 65 anos	-1,0006	0,9187	0,2760
Vizinho	-0,0031	0,0008	<0,0001
Região Metropolitana	0,1801	0,0407	<0,0001
Taxa de Urbanização	-0,0695	0,0621	0,2630
Capacidade fiscal média	-0,9698	0,1959	<0,0001
Nordeste	-0,4258	0,0551	<0,0001
Sudeste	-0,9394	0,0562	<0,0001
Sul	-0,9247	0,0584	<0,0001
Centro-Oeste	-0,9752	0,0603	<0,0001
Porte populacional abaixo de 5 mil	0,2431	0,1625	0,1350
Porte populacional entre 5 mil e 10 mil	0,3600	0,1595	0,0240
Porte populacional entre 10 mil e 20 mil	0,3423	0,1574	0,0296
Porte populacional entre 20 mil e 50 mil	0,2772	0,1543	0,0723
Porte populacional entre 50 e 100 mil	0,1104	0,1552	0,4770
Porte populacional entre 100 mil e 500 mil	0,2963	0,1514	0,0503
Gamma (γ)	2,44	-	-

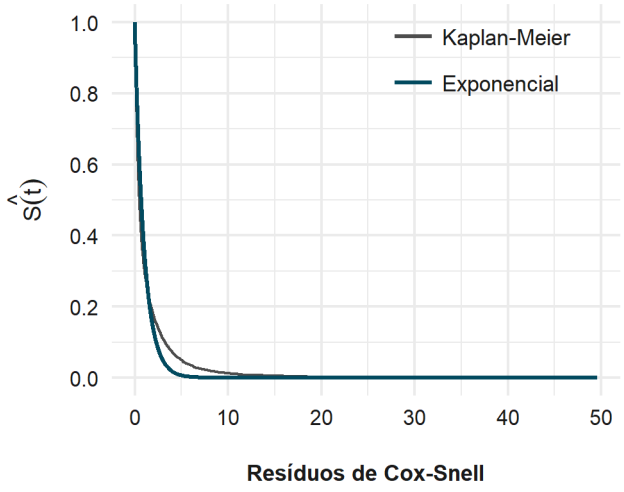

Figura 19: Resíduos de Cox-Snell para o modelo de regressão Log-Logístico com todas as variáveis explicativas.

É observada uma boa adequação do modelo aos dados, logo será feita a análise dos parâmetros estimados pelo modelo de regressão Log-Logístico completo.

Aqui observa-se que as variáveis *porte* e *região* possuem o mesmo comportamento observado nos demais modelos. Quanto maior a capacidade fiscal média, mais rapidamente o município adere ao PSF. Os municípios que não são considerados regiões metropolitanas, os que possuem vizinhos que adotaram o programa, que têm alta percentagem de habitantes com idades abaixo de 17 anos, com ideologia mais voltada para a esquerda, que possuem alinhamento federal e que tem número baixo de partidos efetivos e em anos não eleitorais têm maior probabilidade de falharem, ou seja, de adotarem o PSF.

O modelo final pela distribuição Log-Logística, com tempo inicial em 1997, é apresenta na tabela 9:

Tabela 9: Estimativas para os parâmetros do modelo final de regressão Log-Logístico.

Variável	Estimativa	Erro padrão	P-valor
Intercepto	10,7463	0,2190	<0,0001
NEP	0,0344	0,0211	0,1030
Alinhamento federal	-0,0916	0,0265	0,0005
Ano eleitoral	0,2019	0,0246	<0,0001
Ideologia	0,0189	0,0059	0,0014
População até 17 anos	-7,5475	0,2325	<0,0001
Vizinho	-0,0030	0,0008	<0,0001
Região Metropolitana	0,1816	0,0406	<0,0001
Capacidade fiscal média	-0,9968	0,1920	<0,0001
Nordeste	-0,4386	0,0531	<0,0001
Sudeste	-0,9423	0,0561	<0,0001
Sul	-0,9123	0,0579	<0,0001
Centro-Oeste	-0,9676	0,0590	<0,0001
Porte populacional abaixo de 5 mil	0,2350	0,1611	0,1450
Porte populacional entre 5 mil e 10 mil	0,3498	0,1582	0,0270
Porte populacional entre 10 mil e 20 mil	0,3295	0,1560	0,0347
Porte populacional entre 20 mil e 50 mil	0,2618	0,1532	0,0875
Porte populacional entre 50 e 100 mil	0,0928	0,1544	0,5480
Porte populacional entre 100 mil e 500 mil	0,2829	0,1510	0,0610
Gamma (γ)	2,44	-	-

A adequação do modelo aos dados é apresentada na figura a seguir:

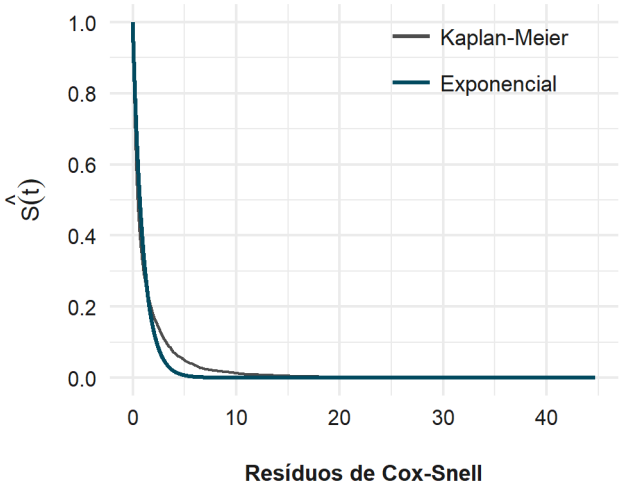

Figura 20: Resíduos de Cox-Snell para o modelo final de regressão Log-Logístico.

Observa-se um bom ajuste do modelo aos dados. Assim, será dado prosseguimento na análise dos parâmetros estimados.

Nota-se que o p-valor para a variável NEP está um pouco acima de 10%, porém optou-se por deixar esta variável no modelo, para fins de análise de competição

política. Já para as demais covariáveis o comportamento é o mesmo notado anteriormente para o modelo de regressão Log-Logístico completo. Vale ressaltar que, o valor do parâmetro de forma maior que 1, $\gamma = 2,44$. Isso confirma, para a distribuição Log-Logística, que a função de risco acumulada é unimodal.

5 Considerações Finais

Este trabalho teve por objetivo a contribuição para o desenvolvimento de modelos de regressão para dados em análise de sobrevivência, com foco no Programa Saúde da Família (PSF).

As distribuições foram escolhidas por meio da análise gráfica da função de risco acumulada segundo Kaplan-Meier e também pela curva do Tempo Total em Teste. A partir daí, foi feita uma extensão nas distribuições de probabilidade para inserir covariáveis por meio do parâmetro de escala, dando origem aos modelos de regressão Weibull e Log-Logístico. Pelo método de seleção de variáveis Backward, foram escolhidas as variáveis que estão expostas nos modelos de regressão finais, tendo seus ajustes analisados por meio dos resíduos de Cox-Snell.

Dessa forma, foram propostos 4 modelos de regressão para o tempo até a adesão do programa Saúde da Família, sendo 2 deles baseados na distribuição Weibull e os demais, na distribuição Log-Logística.

De maneira geral, houve coerência entre os modelos e os resultados esperados a partir da análise descritiva. Sob o ponto de vista estatístico, todos apresentaram um bom ajuste aos dados, com algumas particularidades. Por exemplo, a variável *ano eleitoral* apresentou resultados diferentes nos modelos de regressão finais. Dessa maneira, a escolha do modelo ideal irá variar de acordo com o objetivo buscado pelas ciências políticas.

Para estudos futuros seria interessante discutir melhor os resultados obtidos com estudiosos das Ciências Política e, talvez, propor a melhora desses modelos com termos de interação, por exemplo.

Referências

Coêlho, D. B., Cavalcante, P., and Turgeon, M. (2015). Mecanismos de difusão de políticas sociais no brasil: Uma análise do programa saúde da família. *Revista de Sociologia e Política*, 24(58):145–165.

Colosimo, E. A. and Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. ABE - Projeto Fisher, São Paulo, 1ª edição.

Cox, G. W. and McCubbins, M. D. (1986). Electoral politics as a redistributive game. *Journal of Politics*, 48:370–389.

de Almeida Godinho Rosa, W. and Labate, R. C. (2005). Programa saúde da família: A construção de um novo modelo de assistência. *Revista Latino-am Enfermagem*, pages 1027–1034.

dos Santos, D. F. (2017). Modelo de regressão log-logístico discreto com fração de cura para dados de sobrevivência. *Universidade de Brasília*.

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

Nakano, E. Y. (2017). Um curso de análise de sobrevivência.

Viana, A. L. D. and Poz, M. R. D. (2005). A reforma do sistema de saúde no brasil e o programa de saúde da família. *Revista Saúde Coletiva*, 15:225–264.