



Universidade de Brasília
Departamento de Estatística

**Modelo de Regressão Log-Logístico Discreto para Dados da Política de
Zoeneamento e Uso do Solo na Presença de Observações Censuradas**

Jady Goes Lima

Relatório Final de Monografia apresentado para o Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

**Brasília
2018**

Jady Goes Lima

**Modelo de Regressão Log-Logístico Discreto para Dados da Política de
Zoeneamento e Uso do Solo na Presença de Observações Censuradas**

Orientadora:

Profa. Dra. **Juliana Betini Fachini Gomes**

Relatório Final de Monografia apresentado para o Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

**Brasília
2018**

Agradecimentos

Agradeço primeiramente aos meus pais por todo o suporte dado ao longo de minha vida, por sempre estarem presentes, serem meus grandes incentivadores, acreditarem em meu potencial e serem meus maiores exemplos como pessoas e profissionais.

À minha família, por mesmo longe, sempre estarem torcendo por mim e pelas constantes demonstrações de amor em cada pequena conquista e, também, ao longo de todos os tropeços e quedas.

Aos meus amigos, e em especial à Paloma, por estarem sempre presentes nos mais variados momentos dessa jornada, oferecendo mais do que o necessário para que eu conseguisse avançar a caminhada de forma mais alegre, divertida e fácil.

Por fim, agradeço a Profa. Dra. Juliana Betini Fachini Gomes por ter me aceitado de braços abertos como sua orientanda, ser sempre muito compreensiva, atenciosa e preocupada. Não poderia ter escolhido uma pessoa melhor para me orientar ao final de mais um ciclo! Você é inspiração!

Sumário

1 Introdução	5
2 Revisão de Literatura	7
2.1 Introdução à Análise de Sobrevida	7
2.2 Caracterizando o Tempo de Sobrevida	8
2.2.1 Variável Aleatória Contínua	8
2.2.2 Variável Aleatória Discreta	10
2.3 Técnicas Não-Paramétricas	12
2.3.1 Estimção na Ausência de Censura	12
2.3.2 Estimção na Presença de Censura	13
2.4 Determinação Empírica	14
2.4.1 Curva do Tempo Total em Teste (TTT)	15
2.4.2 Gráfico da Função de Risco Acumulado	16
2.5 Distribuição Log-Logística	17
2.6 Discretização de Distribuições	17
2.7 Distribuição Log-Logística Discreta	18
2.8 Estimção via Método de Máxima Verossimilhança	18
3 Metodologia	21
3.1 Material	21
3.2 Métodos	23
3.2.1 Modelo de Regressão Log-Logístico	23
3.3 Resíduos de Cox-Snell	24
4 Resultados e Discussões	26
4.1 Plano Diretor	26
4.1.1 Análise Descritiva	26
4.1.2 Estimção da Distribuição Log-Logística Discreta	30
4.1.3 Estimção do Modelo de Regressão Log-Logístico Discreto	31
4.2 Política Específica	33
4.2.1 Análise Descritiva	33
4.2.2 Estimção do Modelo Log-Logístico Discreto	37
4.2.3 Estimção do Modelo de Regressão Log-Logístico Discreto	38
4.3 Considerando Ambas as Formas de Adesão	40
4.3.1 Análise Descritiva	40
4.3.2 Estimção do Modelo Log-Logístico Discreto	43
4.3.3 Estimção Modelo de Regressão Log-Logístico Discreto	44
4.4 Conclusão	46
Referências	49

Resumo

Modelo de Regressão Log-Logístico Discreto para Dados da Política de Zoneamento e Uso do Solo na Presença de Observações Censuradas

Neste trabalho são propostas três diferentes maneiras para realizar a contagem discreta do tempo até a adesão da política de Zoneamento e Uso do Solo (ZUS), sendo construído, para cada abordagem, um modelo de regressão Log-Logístico. A motivação para o desenvolvimento desses modelos surge quanto à raridade de estudos estatísticos na área de Ciências Políticas e à impossibilidade de aplicação de modelos contínuos em dados discretos, advindos de uma contagem de tempo em anos. Os três modelos construídos utilizaram a distribuição Log-Logística Discreta, sendo considerada a função de ligação exponencial para a inclusão de covariáveis no modelo. Para realizar a estimativa dos parâmetros do modelo, foi necessário construir um algoritmo computacional.

Palavras-chave: Dados discretos; Dados Censurados; Distribuição Log-Logística Discreta; Ciência Política; Modelo de Regressão

Lista de Figuras

1	Possíveis Formas da Curva TTT	15
2	Possíveis Formas da Função de Risco Acumulado	16
3	Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão via Plano Diretor	27
4	Determinações Empíricas para a Adesão via Plano Diretor	27
5	Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão via Plano Diretor	28
6	Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão via Plano Diretor	29
7	Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão via Plano Diretor	29
8	Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão via Plano Diretor	30
9	Ajustes da Distribuição Log-Logística Discreta para a Adesão via Plano Diretor	31
10	Análise de Resíduos de Cox-Snell para o Modelo de Plano Diretor	33
11	Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão via Política Específica	34
12	Determinações Empíricas para a Adesão via Política Específica	34
13	Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão via Política Específica	35
14	Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão via Política Específica	36
15	Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão via Política Específica	36
16	Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão via Política Específica	37
17	Ajustes da Distribuição Log-Logística Discreta para a Adesão via Política Específica	38
18	Análise de Resíduos de Cox-Snell para o Modelo de Política Específica	40
19	Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão Total	41

20	Determinações Empíricas para a Adesão Total	41
21	Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão Total	42
22	Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão Total	42
23	Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão Total	43
24	Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão Total	43
25	Ajustes da Distribuição Log-Logística Discreta para a Adesão Total	44
26	Análise de Resíduos de Cox-Snell para o Modelo Total para a Adesão Total	46

Lista de Tabelas

1	Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão via Plano Diretor	31
2	Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão via Plano Diretor	32
3	Estimativas do Modelo de Regressão Log-Logístico para a Adesão via Plano Diretor	32
4	Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão via Política Específica	38
5	Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão via Política Específica	39
6	Estimativas do Modelo de Regressão Log-Logístico para a Adesão via Política Específica	39
7	Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão Total	44
8	Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão Total	45
9	Estimativas do Modelo de Regressão Log-Logístico para a Adesão Total . .	45

1 Introdução

Amplamente utilizada em diferentes áreas do conhecimento, a Estatística é uma ferramenta que busca auxiliar na compreensão dos mais variados processos, trazendo respostas para inúmeras questões. Para isso, são utilizadas técnicas que auxiliam o estudo desde a criação de questionários e coleta de dados até a consolidação da informação.

Em alguns estudos, a variável que deseja-se estudar é o tempo até a ocorrência de um evento de interesse. Genericamente, tais eventos são ditos como falhas e podem ser definidos, por exemplo, como o tempo até a cura de um paciente em tratamento, o tempo até a ocorrência de um sinistro de um seguro ou o tempo até que uma lâmpada pare de funcionar.

Como esses estudos se interessam por uma variável temporal, a mesma não é observada instantaneamente, fazendo com que, em alguns casos, o evento de interesse não ocorra, fato este denominado censura. Dessa forma, a variável resposta é constituída então pelo tempo de falha (quando observa-se a ocorrência do evento de interesse) e pelo tempo de censura (quando não observa-se, até o final do estudo, a falha).

Apesar da informação adquirida nas censuras serem apenas parciais e incompletas, é de extrema importância incluí-las nas análises para que o resultado não apresente viés e mostre uma situação que difere da realidade. Buscando incluir a informação das censuras nas análises, a Análise de Sobrevivência foi desenvolvida, sendo esse o seu maior diferencial.

Este trabalho pretende utilizar as técnicas existentes em Análise de Sobrevivência para criar modelos de regressão log-logísticos discretos que permitam entender quais variáveis sociopolíticas influenciam na decisão de adesão de políticas públicas pelos governos municipais brasileiros, em certo período de tempo, focando especialmente na política de Zoneamento e Uso do Solo (ZUS).

Além da avaliação da influência de variáveis sociopolíticas na adesão da ZUS, o trabalho também pretende verificar o impacto da forma escolhida pelo município para aderir à política, podendo a adesão ser via Plano Diretor ou via Legislação Específica. Sendo assim, foi construído um modelo considerando apenas as adesões via Plano Diretor, logo em seguida, um modelo referente às adesões via Política Específica e, por fim, um modelo com todas as adesões, independentemente de como foram realizadas.

A política de Zoneamento e Uso do Solo busca a regulação do uso e da ocupação do solo urbano, que representa a materialização das relações socioeconômicas vigentes nas cidades, em função de condicionantes ambientais, legais e de características de infraestrutura instalada. Dessa forma, políticas urbanas de mobilidade (transporte urbano

e sistema viário), saneamento básico, aproveitamento dos recursos hídricos, preservação ambiental, habitação, rede de saúde, segurança, desenvolvimento socioeconômico, entre outras, produzem repercussões diretas no território e, por isso, tem na regulação do uso e da ocupação do solo um de seus principais instrumentos (Vaz,2006).

O zoneamento baseia-se em uma concepção da gestão do espaço urbano amparada na ideia de eleger os usos possíveis para determinadas áreas da cidade. Dessa forma, o que se pretende é evitar conflitos e repercussões negativas entre os usos. Sendo assim, o papel da legislação se restringe a direcionar a ocupação da cidade como forma de legitimação do espaço construído, desconsiderando o caráter de sua dinâmica, mesmo que perversa e excludente (Vaz,2006).

Para estudar os possíveis modelos que expliquem o comportamento da adesão à política de Zoneamento e Uso do Solo e selecionar o melhor entre eles será utilizado o *software* R.

2 Revisão de Literatura

2.1 Introdução à Análise de Sobrevida

A Análise de Sobrevida é uma área da Estatística que tem como objeto de estudo, variáveis que representam o tempo até a ocorrência do evento de interesse. Genericamente, tais eventos são ditos como falhas e podem ser definidos, por exemplo, como o tempo até a cura de um paciente em tratamento, o tempo até a ocorrência de um sinistro de um seguro ou o tempo até a adesão de um programa social pelo governo de um município.

Sabendo que os dados de sobrevida apresentam variáveis temporais, podem ocorrer situações que impossibilitem observar a ocorrência do evento de interesse, nesses casos, o estudo apresentará apenas uma informação parcial, definida como censura. Dessa forma, a variável resposta é constituída pelo tempo de falha (quando observa-se a ocorrência do evento de interesse) e pelo tempo de censura (quando não observa-se, até o final do estudo, a falha).

O tempo de falha nada mais é do que o tempo até a ocorrência da falha e consequentemente, só é considerado para as unidades de informação que sofreram o evento de interesse durante o estudo. Para que o mesmo seja bem definido, é necessária a fixação do tempo inicial do estudo, para tornar as observações comparáveis, a determinação da escala de medida que será utilizada, podendo o tempo ser considerado como discreto ou contínuo e a definição do que será considerado falha.

O diferencial da Análise de Sobrevida é conseguir utilizar as observações censuradas, pois apesar de serem informações incompletas, também são importantes e devem ser incorporadas nas análises. Caso os dados não apresentem censuras, todas as observações apresentarão o tempo de falha e técnicas estatísticas clássicas também podem ser usadas.

Para que seja possível incorporar a informação contida nas censuras e dessa forma, evitar que os resultados não sejam viciados e evidenciem a uma situação diferente da realidade é fundamental a criação de uma variável dicotômica que indique se a falha foi observada ou não. Tal variável é comumente chamada por variável indicadora de censura e é definida da seguinte forma:

Com $i = 1, 2, \dots, n$:

$$\delta_i = \begin{cases} 0, & \text{se o } i\text{-ésimo tempo foi censurado} \\ 1, & \text{se o } i\text{-ésimo tempo foi observado.} \end{cases}$$

As censuras podem ser subdivididas de acordo com a informação que possuem, quando a censura é à direita tem-se que o tempo de falha é sabidamente maior do que o

tempo registrado no estudo, quando é à esquerda sabe-se que o tempo registrado é maior que o tempo de falha e a censura intervalar indica que a falha ocorreu entre dois momentos observados.

As censuras à direita ainda podem ser classificadas a partir de três mecanismos, sendo eles:

- **Censura Tipo I:** Nesses casos, o que leva a ocorrência da censura é a determinação do tempo de duração do estudo. Sendo assim, todas as observações que não apresentarem o evento de interesse até o final do estudo serão consideradas observações censuradas.
- **Censura Tipo II:** Resultam de estudos em que a quantidade de indivíduos que falham é fixada. Dessa forma, após a obtenção da quantidade estabelecida, o estudo acaba e todos os participantes que não vivenciaram a falha tornam-se censuras.
- **Censura Aleatória:** Engloba as demais, acontece quando o acompanhamento do participante é interrompido ou quando a falha ocorre por um motivo diferente do estudado, ocorrendo naturalmente.

Como as análises de sobrevivências são realizadas a partir de resultados estatísticos assintóticos, quando o estudo possui censuras à direita, os resultados são robustos em relação às suas subdivisões e, dessa forma, as mesmas técnicas são utilizadas para qualquer mecanismo de censura à direita.

2.2 Caracterizando o Tempo de Sobrevivência

O tempo até a ocorrência do evento de interesse é uma variável não negativa que pode ser especificada tanto por uma função de sobrevivência quanto por uma função taxa de falha (função de risco). Como estas funções possuem uma relação matemática, quando uma delas é especificada a outra é definida consequentemente. (Nakano, 2017)

2.2.1 Variável Aleatória Contínua

Considerando que a variável de interesse seja contínua, tem-se uma série de funções, além das já citadas, que podem descrever seu comportamento. A seguir serão definidas as funções de densidade de probabilidades, sobrevivência, risco e risco acumulado.

I. Função Densidade de Probabilidades

A função densidade de probabilidades, $f(t)$, define a probabilidade de uma variável aleatória não negativa contínua T assumir um valor qualquer. Essa função satisfaz as seguintes propriedades:

i . $f(t) \geq 0$ para todo $t \geq 0$;

ii . $\int_0^{\infty} f(t)dt = 1$; e

iii . $P(a \leq T \leq b) = \int_a^b f(t)dt, \forall 0 \leq a < b$.

Também pode ser representada pelo limite da probabilidade de um indivíduo falhar no intervalo de tempo $[t, t + \Delta t]$, dividida pelo comprimento do intervalo (Nakano, 2017):

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

II. Função de Sobrevivência

A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um determinado momento t :

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u)du, t \geq 0.$$

Conseqüentemente, a função de distribuição acumulada pode ser definida como a probabilidade da observação não sobreviver até o tempo t , ou seja:

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

Sabendo que a função de sobrevivência é não crescente e absolutamente contínua, tem-se que:

$$\lim_{t \rightarrow 0} S(t) = 1 \quad \text{e} \quad \lim_{t \rightarrow \infty} S(t) = 0.$$

III. Função de Risco

A função de risco, $h(t)$, determina o risco instantâneo do indivíduo falhar e é definida como o limite da probabilidade do evento de interesse ocorrer no intervalo de tempo $[t, t + \Delta t]$ sabendo que antes de t o indivíduo não falhou dividida pelo intervalo de tempo Δt , sendo expressa da seguinte maneira:

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

A função de risco pode ser constante, situação a qual o risco de ocorrer o evento de interesse não muda com o passar do tempo, monótona crescente ou decrescente indicando, respectivamente, um crescimento ou decrescimento do risco ao longo do tempo, unimodal apresentando um crescimento no risco até certo momento e logo após um decrescimento, apresentar um formato de banheira ou "U" o qual em tempo inicial os riscos são maiores, nos momentos intermediários o risco decresce e nos instantes finais o risco torna a crescer. Ainda existem outros comportamentos da função de risco, sendo eles mais complexos. (Nakano, 2017)

Alguns autores (Colosimo, 2006) consideram a função de risco mais informativa do que a função de sobrevivência pelo fato de que, apesar de algumas funções de sobrevivência apresentarem comportamentos similares, suas funções de risco podem ser complementemente divergentes.

IV. Função de Risco Acumulado

Apesar da função de risco acumulado, $H(t)$, não apresentar uma interpretação direta, ela é utilizada na estimação não paramétrica por ser mais simples de trabalhar do que a função de risco e também contribui para a seleção do modelo mais adequado para ajustar um conjunto de dados. A função de risco acumulado é expressa por:

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

2.2.2 Variável Aleatória Discreta

Uma variável aleatória discreta assume um número finito ou infinito enumerável de valores. Como dito anteriormente, em análise de sobrevivência, a variável resposta costuma ser não negativa contínua pois o tempo normalmente é contado pelo tempo real. Porém, podem existir estudos em que o tempo seja contado por meio de ciclos o que torna a variável resposta discreta (Nakano, 2017).

Analogamente as variáveis contínuas, as variáveis discretas podem ser descritas por inúmeras funções e a seguir serão definidas as funções de distribuição de probabilidades, sobrevivência, risco e risco acumulado.

I. Função de Distribuição de Probabilidades

Seja uma variável aleatória discreta T que assume valores inteiros não negativos, a função de distribuição de probabilidades de T , $p(t)$, define a probabilidade da variável as-

sumir um valor possível, sendo a versão discreta da função de densidade de probabilidades. Tal função satisfaz as seguintes propriedades:

i. $p(t) \geq 0$, para $t = 0, 1, 2, \dots$;

ii. $\sum_{t=0}^{\infty} p(t) = 1$.

II. Função de Sobrevivência

Assim como dito anteriormente, a função de sobrevivência é definida pela probabilidade de um indivíduo sobreviver, ou seja, não experimentar o evento de interesse até um determinado tempo t . A forma como ela é expressa quando a variável reposta passa a ser discreta é dada por:

$$S(t) = P(T \geq t) = \sum_{k=t+1}^{\infty} p(k) = \sum_{k=t+1}^{\infty} P(T = k), t = 0, 1, 2, \dots$$

A função de sobrevivência de uma variável aleatória discreta possui um formato de escada pois decresce nos pontos em que t possui uma probabilidade positiva e é constante nos demais. Tem-se também que:

$$\lim_{t \rightarrow 0} S(t) = 1 - P(T = 0) \quad \text{e} \quad \lim_{t \rightarrow \infty} S(t) = 0.$$

III. Função de Risco

A função de risco, quando trata-se de uma variável aleatória discreta, é definida como a probabilidade condicional do indivíduo falhar no instante t , dado que o mesmo não tenha falhado antes de t . Sendo expressa por:

$$h(t) = P(T = t | T \geq t), t = 0, 1, 2, \dots$$

Assim como em variáveis aleatórias contínuas, a função de risco pode ser monótona crescente ou decrescente, unimodal, em formato de "U" e em formatos mais complexos. Porém, por ser definida a partir de uma probabilidade condicional, é uma função limitada no intervalo $[0,1]$ e nos valores negativos e não inteiros de t é igual a zero.

IV. Função de Risco Acumulado

Seguindo o mesmo raciocínio dado em variáveis aleatórias contínuas, a função de risco acumulado, $H(t)$, fornece o risco acumulado do indivíduo no tempo t e não possui uma interpretação direta. Tal função é definida da seguinte maneira:

$$H(t) = \sum_{k=0}^t h(k), t = 0, 1, 2, \dots$$

2.3 Técnicas Não-Paramétricas

O primeiro passo de qualquer análise estatística é buscar descrever o comportamento das variáveis, porém, quando trabalha-se com dados que possuem censuras, a análise descritiva convencional não funciona da melhor maneira. Para que a descrição de dados censurados seja feita adequadamente, o ponto de partida é buscar estimar a função de sobrevivência e, a partir dela, fazer estimações de medidas descritivas.

A seguir serão apresentados alguns métodos não paramétricos para estimação da principais funções utilizadas em análise de sobrevivência.

2.3.1 Estimação na Ausência de Censura

I. Estimação da Função Densidade de Probabilidades

Ao considerar a variável de estudo contínua, a estimação da função densidade de probabilidade será realizada a partir da seguinte expressão (Nakano, 2017):

$$f(\hat{t}) = \frac{\text{n}^\circ \text{ de indivíduos que falharam no intervalo } [t, t + \Delta t]}{n\Delta t},$$

em que:

- n : número total de indivíduos
- Δt : amplitude do intervalo

Caso a variável de interesse seja discreta, a função de distribuição de probabilidades é estimada por:

$$p(\hat{t}) = \frac{\text{n}^\circ \text{ de indivíduos que falharam no tempo } t}{\text{n}^\circ \text{ total de indivíduos}}, t = 0, 1, 2, \dots$$

II. Estimação da Função de Sobrevivência

No universo de dados não censurados, a função de sobrevivência é estimada simplesmente pela proporção dos indivíduos que não experimentaram o evento de interesse até o tempo t . A sua estimação é expressa por:

$$S(\hat{t}) = \frac{\text{n}^\circ \text{ de indivíduos que não falharam até o tempo } t}{\text{n}^\circ \text{ total de indivíduos}}, \text{ para todo } t \geq 0.$$

Como dito anteriormente, a função de sobrevivência apresenta o formato escada. Os degraus ocorrem em valores de tempo em que um indivíduo falhou e a altura do salto é de $1/n$. Em casos de empate, o tamanho do intervalo é multiplicado pela quantidade de empates.

2.3.2 Estimação na Presença de Censura

Os métodos de estimação mostrados na seção 2.3.1 não são adequados quando os dados apresentam censuras. Como, em análise de sobrevivência, dados censurados são bastante comuns, novos métodos de estimação são necessários. A seguir será apresentado o estimador não paramétrico de Kaplan - Meier e uma discussão dos estimadores de Nelson - Aalen e Tabela de vida para as funções de sobrevivência e risco acumulado, os quais são adequados para dados com censuras à direita.

Para a construção desses estimadores, considere $t_{(i)}$ o i -ésimo valor não censurado, distinto e ordenado da amostra, $i = 1, 2, \dots, k$, $t_{(0)} = 0$ e $t_{(k+1)} = \infty$. Dessa forma, tem-se que:

$$0 = t_{(0)} < t_{(1)} < t_{(2)} < \dots < t_{(k)} = \infty,$$

com $k \leq n$, sendo igual quando não forem observados valores de falha iguais, ou seja, não havendo empates.

I. Estimador de Kaplan - Meier

O estimador de Kaplan - Meier, conhecido também por estimador limite-produto, é uma adaptação da função de sobrevivência empírica e é definido pela expressão:

$$\hat{S}(t) = \prod_{i:t_{(i)} \leq t} \left[\frac{n_i - d_i}{n_i} \right] = \prod_{j:t_{(j)} \leq t} \left[1 - \frac{d_j}{n_j} \right] \quad (1)$$

em que:

- n_i : número de indivíduos que podem falhar (sob risco) no tempo t_i ;
- d_i : número de indivíduos que falharam no tempo t_i
- n : número total de indivíduos.

Para a construção desse estimador foi necessário seguir a intuição de que, em uma sequência de passos, o passo seguinte depende do antecedente, de modo que, para que um tempo superior seja alcançado, é obrigatória a passada por um tempo inferior ao mesmo. A partir desse raciocínio tem-se que, para que uma unidade de informação sobreviva até o tempo t_i , a mesma, necessariamente, sobreviveu até o tempo t_{i-1} , ou seja (Nakano, 2017):

$$\begin{aligned} S(t_{(i)}) &= P(T > t_{(i)}) \\ &= P(T > t_{(i-1)}, T > t_{(i)}) \\ &= P(T > t_{(i-1)})P(T > t_{(i)}|T > t_{(i-1)}) \\ &= S(t_{(i-1)})[1 - P(T \leq t_{(i)}|T > t_{(i-1)})] \\ &= S(t_{(i-1)})[1 - P(t_{(i-1)} < T \leq t_{(i)}|T > t_{(i-1)})] \\ &= S(t_{(i-1)})[1 - P(t_{(i-1)} < T \leq t_{(i)}|T \geq t_{(i)})], i = 1, 2, \dots, k \end{aligned}$$

A partir deste resultado, percebe-se que estimar a função de sobrevivência por Kaplan-Meier é equivalente a estimar a probabilidade condicional, q_i , de uma observação falhar no intervalo $[t_{(i-1)}, t_{(i)}]$, dado que ela estava sob risco em t_i , ou seja, sabendo que a mesma não havia experimentando o evento de interesse em $t_{(i-1)}$. A estimação de q_i é expressa por:

$$\hat{q}_i = \frac{\text{n}^\circ \text{ de indivíduos que falharam entre } (t_{(i-1)}, t_{(i)})}{\text{n}^\circ \text{ de indivíduos sob risco em } t_{(i)}} = \frac{d_i}{n_i}.$$

Sabendo que a relação entre a função de risco acumulado e a função de sobrevivência é dada por $S(t) = \exp\{-H(t)\}$, tem-se que o estimador de Kaplan-Meier da função de risco acumulado é:

$$H_{KM}(t) = -\log(S_{KM}(t)),$$

em que $S_{KM}(t)$ é a estimativa de Kaplan-Meier da função de sobrevivência.

Em casos em que o maior tempo observado é censurado, a função de sobrevivência estimada por Kaplan-Meier não atingirá o valor zero, isso acontece pois o tempo em que o último indivíduo falhou não é conhecido e somente após a falha de todos os indivíduos a probabilidade de sobrevivência será nula.

II. Outros Estimadores Não-Paramétricos

Além do estimador de Kaplan-Meier, existem outros estimadores que possuem a funcionalidade de estimar a função de sobrevivência, tais como o estimador de Nelson-Aalen e a tabela de vida. Porém, o estimador mais conhecido e utilizado é o de Kaplan-Meier, por ser um estimador de máxima verossimilhança de $S(t)$ e apresentar, conseqüentemente, as propriedades que este tipo de estimador traz consigo.

O estimador de Nelson-Aalen possui propriedades similares ao de Kaplan-Meier e, quando os resultados encontrados na estimação da função de sobrevivência desses dois estimadores são comparados, nota-se que os valores são bastante próximos, com as estimativas de Nelson-Aalen sempre maiores ou iguais as estimativas de Kaplan-Meier (Bohoris, 1994).

A tabela de vida foi proposta por demógrafos e atuários no século XIX e foi amplamente utilizada em censos demográficos buscando estudar as características que influenciam no tempo de vida dos seres humanos. Diferentemente dos estimadores de Kaplan-Meier e Nelson-Aalen que são construídos com um número de intervalos igual ao número de tempos de falha distintos, a tabela de vida agrupa os tempos de falha arbitrariamente.

2.4 Determinação Empírica

Para iniciar uma análise paramétrica em análise de sobrevivência é necessário, inicialmente, selecionar, dentre todos os possíveis modelos, o que melhor se ajusta no conjunto de dados. Pensando em reduzir o número de distribuições que possam ajustar o modelo,

podem ser feitos, anteriormente aos ajustes, estudos que busquem descrever a forma da função de risco ou a forma da função de risco acumulada.

2.4.1 Curva do Tempo Total em Teste (TTT)

Buscando descrever a forma da função de risco, foi desenvolvido o gráfico do Tempo Total em Teste (Aarset,1987), o qual é construído a partir de:

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

versus r/n , em que $i = 1, 2, \dots, n$ e $r = 1, 2, \dots, n$ são estatísticas de ordem da amostra.

Seguem exemplos de algumas curvas TTT, com os possíveis formatos que a função de risco pode assumir.

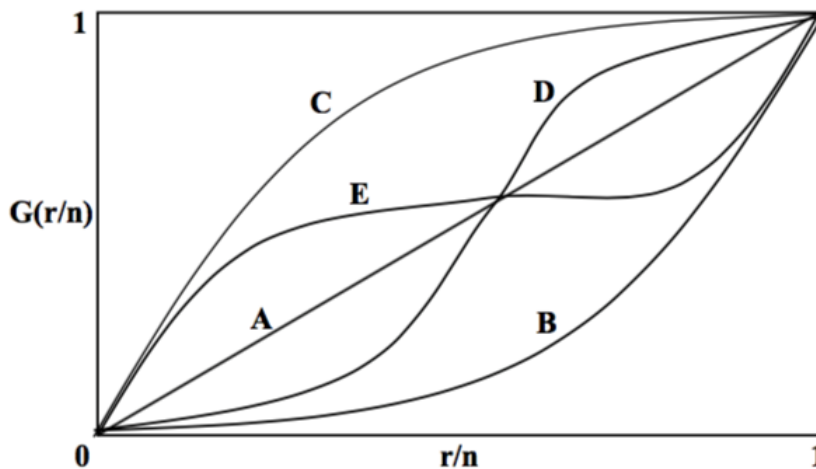


Figura 1: Possíveis Formas da Curva TTT

Considerando a Figura 1, a função de risco assumirá uma forma:

- constante quando o TTT apresentar uma reta diagonal (A);
- monótona crescente ou decrescente quando o TTT apresentar, respectivamente, uma curva côncava (C) ou convexa (B);
- unimodal quando o TTT apresentar uma curva inicialmente côncava e, em seguida, convexa (E);
- formato de "U" quando o TTT apresentar uma curva inicialmente convexa e, posteriormente, côncava (D)

Uma desvantagem na utilização do gráfico do Tempo Total em Teste é o fato de

que o mesmo não considera censuras, o que pode causar problemas na interpretação dos resultados, induzindo a uma escolha ruim do modelo que deve ser ajustado.

2.4.2 Gráfico da Função de Risco Acumulado

Uma alternativa a buscar descrever a forma da função de risco por meio das curvas TTT é investigar o comportamento da função de risco acumulado, plotando sua estimativa, $\hat{H}(t)$, que pode ser encontrada por meio do estimador de Kaplan-Meier.

Uma vantagem da utilização deste método é o fato das censuras serem levadas em consideração, evitando erros na escolha do modelo, principalmente em casos em que o banco de dados apresente uma grande quantidade de censuras. Em contrapartida, perde-se a referência de comparação, que no caso das curvas TTT, era fixada na diagonal ($X = Y$).

A interpretação do comportamento de $\hat{H}(t)$ é inversa à feita nas curvas TTT, e considerando a Figura 2 (Nakano, 2017), será:

- constante quando a curva apresentar uma reta diagonal (A);
- monótona crescente ou decrescente quando apresentar, respectivamente, uma curva convexa (B) ou côncava (C) ;
- formato de "U" quando apresentar uma curva inicialmente côncava e, em seguida, convexa (E);
- unimodal quando apresentar uma curva inicialmente convexa e, posteriormente, côncava (D)

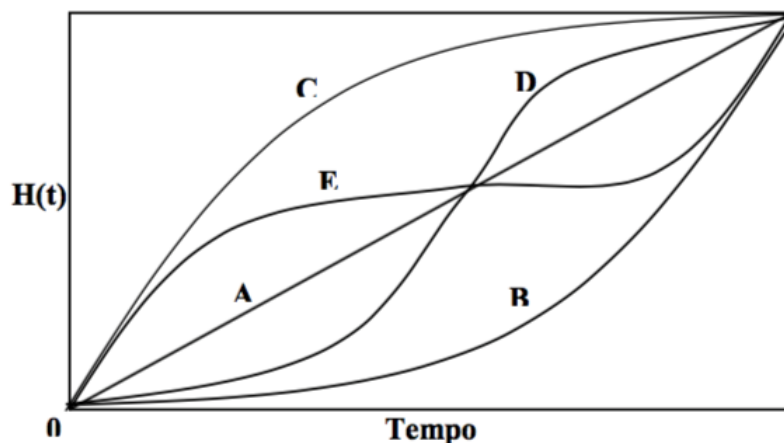


Figura 2: Possíveis Formas da Função de Risco Acumulado

2.5 Distribuição Log-Logística

Ao considerar uma variável aleatória T com distribuição Log-Logística. A função densidade de T é dada por:

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

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

Quando $t > 0$, as funções de sobrevivência e de risco da distribuição Log-Logística são dadas, respectivamente, pelas seguintes expressões:

$$S(t) = \frac{1}{1 + (t/\alpha)^\gamma} \quad \text{e} \quad h(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha [1 + (t/\alpha)^\gamma]}.$$

2.6 Discretização de Distribuições

Discretizar é transformar uma variável aleatória contínua X em uma variável aleatória discreta T com distribuição de probabilidades análoga a função densidade de X . Uma forma de discretizar uma variável é utilizar a transformação $T = [X]$, em que $[X]$ é a parte inteira de X , ou seja, o maior número inteiro menor ou igual a X (Nakano,2017). Sendo assim, a distribuição de probabilidades de T , e suas funções de sobrevivência, $S(t)$, e risco, $h(t)$ são dadas, respectivamente, por:

$$\begin{aligned} p(t) &= P(T = t) = P(t \leq X < t + 1) \\ &= F_X(t + 1) - F_X(t) \\ &= S_X(t) - S_X(t + 1), t = 0, 1, 2, \dots, \end{aligned}$$

$$\begin{aligned} S(t) &= P(T > t) \\ &= \sum_{k=t+1}^{\infty} p(k) \\ &= \sum_{k=t+1}^{\infty} S_X(k) - S_X(k + 1) \\ &= S_X(t + 1), t = 0, 1, 2, \dots, \end{aligned}$$

e,

$$\begin{aligned} h(t) &= \frac{p(t)}{S(t) + p(t)} \\ &= 1 - \frac{S_X(t + 1)}{S_X(t)}, t = 0, 1, 2, \dots \end{aligned}$$

2.7 Distribuição Log-Logística Discreta

Seja T a variável resultante da discretização de uma variável aleatória não negativa X contínua com distribuição Log-Logística de parâmetros α e γ , T segue distribuição Log-Logística discreta, sendo sua função de probabilidades, sobrevivência e risco, respectivamente, dadas por:

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

$$S(t) = \frac{1}{1 + \left(\frac{t+1}{\alpha}\right)^\gamma}, t = 0, 1, 2, \dots,$$

e,

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

em que, α e γ são os parâmetros de escala e de forma.

2.8 Estimação via Método de Máxima Verossimilhança

Após a definição do modelo de probabilidade que melhor se ajusta aos dados, deve-se buscar estimar os parâmetros que vão definir a forma e a escala da mesma. O método de máxima verossimilhança busca estimar tais parâmetros encontrando o valor da estimativa que maximiza a probabilidade de observar os dados obtidos na amostra.

Diferentemente de outras formas de estimação, o método de máxima verossimilhança consegue incorporar as informações parciais das censuras presentes na amostra, fazendo com que seja o método mais comumente utilizado em Análise de Sobrevivência.

Considerando que seja necessário estimar o vetor de parâmetros $\boldsymbol{\theta}$ de uma distribuição de probabilidades $p(t_i; \boldsymbol{\theta})$, que possua função de sobrevivência, $S(t; \boldsymbol{\theta})$, a função de verossimilhança, $L(\boldsymbol{\theta})$ é dada por:

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

sendo que δ_i a variável indicadora de falha e censura.

Desta forma, a contribuição na função de verossimilhança das falhas é a distribuição de probabilidades e a participação das censuras é a função de sobrevivência.

Definida a função de máxima verossimilhança, deve-se encontrar o valor de $\boldsymbol{\theta}$ que a maximize. No geral, por questões de simplicidade, costuma-se encontrar o máximo do

logaritmo de $L(t_i; \boldsymbol{\theta})$ buscando os valores de $\boldsymbol{\theta}$ em que as suas primeiras derivadas sejam iguais a zero.

A busca pelos estimadores de máxima verossimilhança pode ser um processo árduo e em muitas vezes são utilizados métodos numéricos para tal.

3 Metodologia

3.1 Material

O banco de dados estudado apresenta variáveis referentes ao tempo até a adesão da política de Zoneamento e Uso do Solo (ZUS), política essa que busca atingir a regulação do uso e da ocupação do solo urbano e a gestão do espaço urbano amparada na ideia de eleger os usos possíveis para determinadas áreas da cidade.

O município brasileiro pode aderir a uma política por meio da adesão de um Plano Diretor, que nada mais é do que um instrumento que regulamenta o desenvolvimento e a expansão urbana, ou por meio da criação de uma política específica. Dessa forma, o estudo será construído considerando três variáveis respostas diferentes:

- O tempo até a adesão da política ZUS via Plano Diretor;
- O tempo até a adesão da política ZUS via Política Específica;
- O tempo até a adesão da política ZUS desconsiderando a forma de adesão, podendo a mesma ser via Plano Diretor ou Política Específica.

Apesar de terem ocorrido adesões anteriores ao ano de 1997, o estudo considera as adesões à ZUS ocorridas a partir de 1997, visto que foi a partir do ano de 1996 que a variável competição política começou a ser mensurada.

Além da exclusão dos municípios que aderiram à ZUS anteriormente ao ano de 1997, também foram desconsiderados municípios em que os anos de adesão eram desconhecidos ou a informação não foi disponibilizada.

O último ano acompanhado foi 2015, e como o estudo não permite uma perda de supervisão das unidades observacionais (municípios), não ocorrem censuras (não adesão à ZUS) no meio do período analisado. Desta forma, as censuras só ocorrem caso os municípios não tenham aderido à ZUS até o ano de 2015, ou seja, todas as censuras existentes nos dados estão concentradas no último ano de análise e são consideradas censuras à direita do tipo I.

O tempo será tratado como uma variável discreta e sua contagem será realizada em anos completos. Sendo assim, o tempo até a adesão à ZUS (tempo de falha) varia entre 0 e 18 anos, sendo tratado como 0 quando o ano de adesão for 1997 e, analogamente, 18 quando a adesão tiver ocorrido em 2015.

Além das variáveis que indicam o tempo até a adesão à ZUS pelo Plano Diretor, pela política específica e pelo Plano Diretor ou pela Política Específica, o banco possui a variável identificadora de cada município e outras seis covariáveis.

As covariáveis podem ser referentes as eleições (Prefeito reeleito, Ano eleitoral, Margem de vitória e NEP), institucionais (Conselho) ou a características dos municípios (Região).

Para as variáveis eleitorais, as informações foram obtidas de 4 em 4 anos e, portanto, a construção das mesmas considerou a informação da eleição anterior mais próxima ao ano de falha ou censura. As variáveis conselho, região e ano eleitoral são as únicas que possuem informação para todos os anos de falha ou censura possíveis.

- (a) Conselho municipal de política urbana: descreve a existência de um conselho no município, sendo codificada com 1 para as unidades observacionais que o possuem e 0 caso contrário;
- (b) Prefeito reeleito: indica se o governante do município foi reeleito. Para esta variável, 1 indica que o prefeito é reeleito;
- (c) Ano de eleição: diz respeito a existência de processo eleitoral durante o ano de falha ou censura, 1 sendo ano eleitoral;
- (d) Região: retrata a região que o município está inserido, sendo a Sul a categoria de referência;
- (e) Margem de vitória: indica a diferença, em pontos percentuais, entre o primeiro e o segundo colocado nas eleições mais próximas ao ano de falha ou censura;
- (f) NEP: representa o número efetivo de partidos.

3.2 Métodos

3.2.1 Modelo de Regressão Log-Logístico

Assim como nas análises estatísticas com abordagens comuns, em estudos de Análise de Sobrevivência também pode ser interessante verificar se existem variáveis que influenciam o tempo até a ocorrência do evento de interesse. Para isto, é necessário construir um modelo de regressão para comparar o comportamento entre grupos com diferentes características.

Um modelo de regressão nada mais é do que uma fórmula matemática que expressa a relação existente entre a variável resposta e as variáveis explicativas, sendo possível, a partir dele, prever valores da variável de interesse.

Ao considerar o vetor de covariáveis $\mathbf{x} = (1, x_1, \dots, x_k)^T$, utiliza-se uma função de ligação $g(\cdot)$ para conectar a variável resposta com as variáveis explicativas. Sendo assim, o vetor de parâmetros $\boldsymbol{\theta}$ será reparametrizado por (Santos, 2017):

$$\boldsymbol{\theta} = g(\eta),$$

sendo $\eta = \mathbf{x}^T \boldsymbol{\beta}$ o preditor linear e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n)'$ o vetor de coeficientes de regressão.

Considerando a variável de interesse T uma variável aleatória com distribuição de probabilidades Log-Logística discreta e a função de ligação definida como a exponencial do preditor linear, ou seja, $g(\eta) = \exp(\mathbf{x}^T \boldsymbol{\beta})$, tem-se o modelo de regressão Log-Logístico discreto dado por:

$$p(t|\mathbf{x}) = \frac{1}{1 + [t/\exp(\mathbf{x}^T \boldsymbol{\beta})]^\gamma} - \frac{1}{1 + [(t+1)/\exp(\mathbf{x}^T \boldsymbol{\beta})]^\gamma}, t = 0, 1, 2, \dots$$

E, por consequência, as funções de sobrevivência e risco serão definidas, respectivamente, por:

$$S(t|\mathbf{x}) = \frac{1}{1 + [(t+1)/\exp(\mathbf{x}^T \boldsymbol{\beta})]^\gamma} \quad \text{e} \quad h(t|\mathbf{x}) = 1 - \frac{1 + [t/\exp(\mathbf{x}^T \boldsymbol{\beta})]^\gamma}{1 + [(t+1)/\exp(\mathbf{x}^T \boldsymbol{\beta})]^\gamma},$$

em que $t = 0, 1, 2, \dots$, γ é o parâmetro de forma da distribuição Log-Logística e $\boldsymbol{\beta}$ é o vetor de coeficientes da regressão, que será estimado a partir do método de máxima verossimilhança definido na Seção 2.8.

A interpretação dos coeficientes de regressão é definida a partir da função quantil do modelo Log-Logístico discreto. Uma proposta de interpretação é utilizar a razão de

tempos medianos, considerando a seguinte relação (Santos, 2017):

$$t_{0,5}(\hat{\alpha}; \hat{\gamma}) = \inf \left[t : \hat{\alpha} \left(\frac{0,5}{1-0,5} \right)^{\frac{1}{\hat{\gamma}}} - 1 \leq t \right] \cong \hat{\alpha} - 1,$$

e pensando que $\hat{\alpha} = \exp(\beta_0 + \beta_1 x)$, tem-se que $t_{0,5} + 1 = \exp(\beta_0 + \beta_1 x)$. Portanto, a razão do tempo mediano é composta pela seguinte expressão:

$$\frac{1 + t_{0,5}(x = 1, \hat{\gamma}, \hat{\beta})}{1 + t_{0,5}(x = 0, \hat{\gamma}, \hat{\beta})} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{\exp(\hat{\beta}_0)} = \exp(\hat{\beta}_1).$$

Portanto, se $\hat{\beta}_1$ for positivo, interpreta-se que o tempo mediano mais uma unidade ($t_{0,5} + 1$) do indivíduo pertencente ao grupo $x = 1$ é $e^{\hat{\beta}_1}$ vezes o tempo mediano mais uma unidade de um indivíduo do grupo $x = 0$. Para $\hat{\beta}_1$ negativo, conclui-se que o tempo mediano mais um de um indivíduo do grupo $x = 0$ é $e^{-\hat{\beta}_1}$ vezes o tempo $t_{0,5} + 1$ de um indivíduo do grupo 1 (Santos, 2017).

3.3 Resíduos de Cox-Snell

Após o ajuste de um modelo de regressão, é interessante que seja verificada a qualidade do ajuste do mesmo aos dados realizando uma análise de resíduos. Com esse intuito, foi utilizado o resíduo de Cox-Snell.

Os resíduos de Cox-Snell são valores determinados a partir da função de taxa falha acumulada estimada pelo modelo de regressão encontrado, sendo matematicamente escrito como:

$$\hat{e}_i = H(\hat{t}_i | \mathbf{x}).$$

Ao considerar a relação $H(t) = -\log S(t)$, tem-se que o resíduo de Cox-Snell para modelos de regressão Log-Logísticos é dado por:

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

Segundo Lawless (1982), caso o ajuste do modelo seja adequado, os resíduos de Cox-Snell devem seguir uma distribuição exponencial padrão. Sendo assim, a construção do gráfico que apresente a estimação via Kaplan-Meier da função de sobrevivência dos resíduos *versus* a distribuição exponencial padrão aplicada nos erros indicará bons ajustes quando as duas curvas estiverem próximas.

4 Resultados e Discussões

Como dito anteriormente, um município brasileiro pode aderir à políticas públicas via Plano Diretor ou via Política Específica, dessa forma, toda a análise de resultados foi baseada em averiguar o comportamento dessas formas de adesão. Por fim, o estudo desconsidera o meio utilizado pelo município para aderir à ZUS, podendo sua adesão ser realizada ou via Plano Diretor ou via Política Específica.

Visando compreender e explicar o comportamento da adesão à política ZUS para todos os cenários descritos, foram realizadas análises descritivas sobre o tempo de adesão desconsiderando quaisquer variáveis explicativas e posteriormente, englobando-as. Logo em seguida, foi ajustada a distribuição de probabilidades Log-Logística e, finalmente, construído um modelo de regressão Log-Logístico.

4.1 Plano Diretor

4.1.1 Análise Descritiva

Na Figura 3, pode-se observar a estimação não paramétrica da função de sobrevivência via Kaplan-Meier para os municípios que aderiram à ZUS via Plano Diretor. Neste caso, a função de sobrevivência decresce à medida que o tempo passa, porém não se aproxima de zero, visto que até 2015, dos 5.354 municípios considerados, apenas 2.573 haviam aderido à ZUS.

Como somente 48,05% dos municípios aderiram à ZUS, não foi possível calcular o tempo mediano, visto que esta medida propõe indicar em até que tempo metade das unidades observacionais irão sobreviver.

O maior tempo de sobrevivência observado, com 22 falhas, foi o tempo total de duração do estudo, sendo ele de 18 anos. Além disso, neste mesmo tempo são observadas todas as censuras, visto que, no contexto observado, não há possibilidade de ocorrência de censuras durante o período de análise.

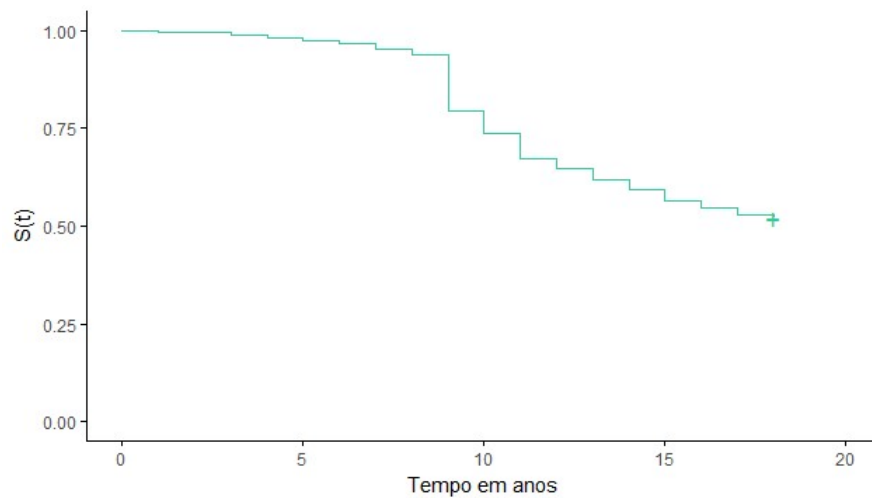


Figura 3: Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão via Plano Diretor

Buscando identificar possíveis distribuições de probabilidade para o tempo até a adesão da política ZUS, foram construídos a curva TTT e o gráfico empírico da função de risco acumulado, sendo apresentados na Figura 4.

O comportamento das curva TTT e da função de risco acumulado indicam que a função de risco estudada possui um padrão monótono crescente, já que a curva TTT apresenta uma forma côncava e o risco acumulado formato convexo.

Dado o comportamento descrito acima, uma possível distribuição candidata para ajustar o tempo até a adesão da ZUS seria uma distribuição que apresentasse uma função de risco monótona crescente, como a distribuição Weibull. Porém, em alguns estudos, a distribuição Log-Logística também apresenta bons ajustes apesar de possuir uma função de risco unimodal e, portanto, também será utilizada na seção a seguir para modelagem dos dados.

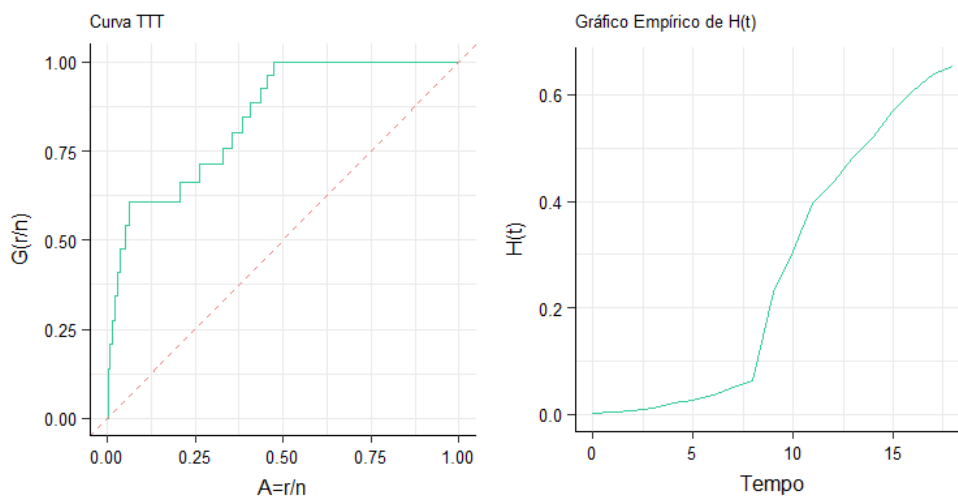


Figura 4: Determinações Empíricas para a Adesão via Plano Diretor

Partindo para a descrição das covariáveis, a primeira a ser descrita é referente à região do país em que o município está inserido. Ao observar a Figura 5, nota-se que até o tempo de 9 anos as regiões apresentam funções de sobrevivência estimadas bem próximas, a partir daí, a região Sul distancia-se das outras, visto que sua curva de sobrevivência possui uma queda maior do que as outras.

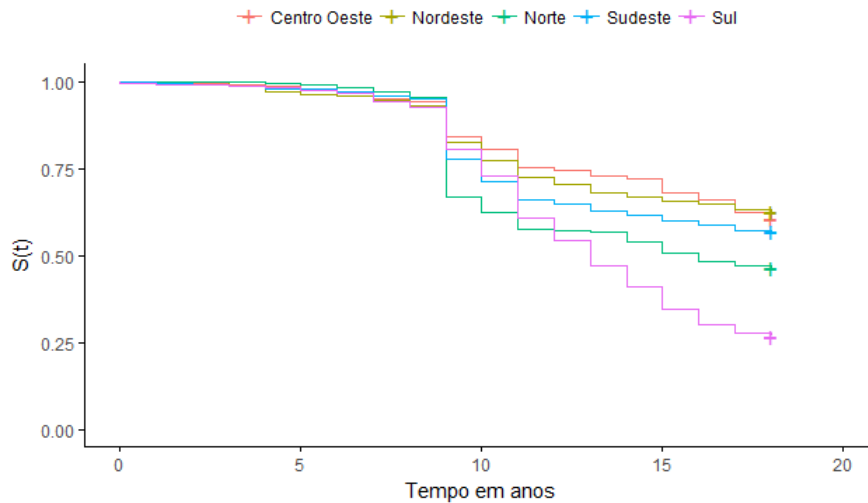


Figura 5: Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão via Plano Diretor

A partir da Figura 6, observa-se que a variável ano de eleição, que indica se o ano de adesão do Plano Diretor pelo município era eleitoral ou não, impacta na diminuição da sobrevivência dos objetos de estudo, já que a partir do tempo 4 a curva de sobrevivência estimada referente aos municípios que aderiram ao Plano Diretor em anos de eleição começa a se afastar da curva estimada para os municípios que aderiram em anos que não foram eleitorais.

É importante notar que, assim como explicado para a função de sobrevivência estimada sem covariáveis, menos da metade dos municípios que aderiram à ZUS via plano diretor falharam, e portanto, não é possível realizar o cálculo da mediana. Já para o grupo pertencente à falha em anos eleitorais, 100% dos casos falharam e o tempo mediano indica que metade dos municípios sobrevivem até o tempo de 11 anos.

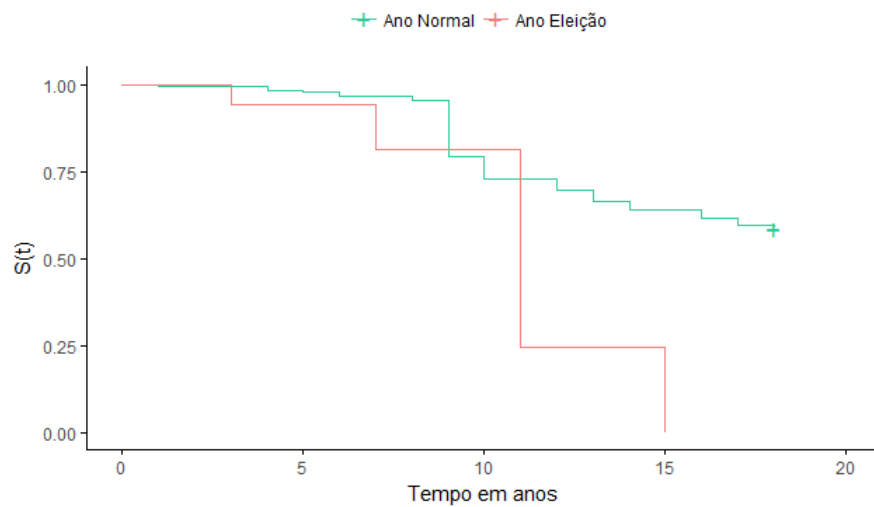


Figura 6: Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão via Plano Diretor

Ao estudar a variável que avalia se a adesão à ZUS por meio do Plano Diretor foi realizada em governo de um prefeito reeleito ou não, é percebido, a partir da Figura 7, que as curvas de sobrevivência estimadas são bem próximas até o tempo 13, após o mesmo ocorre um leve distanciamento entre os grupos, sendo visto que apenas para os tempos de 8 a 10 anos a sobrevivência de municípios governados por prefeitos reeleitos é maior.

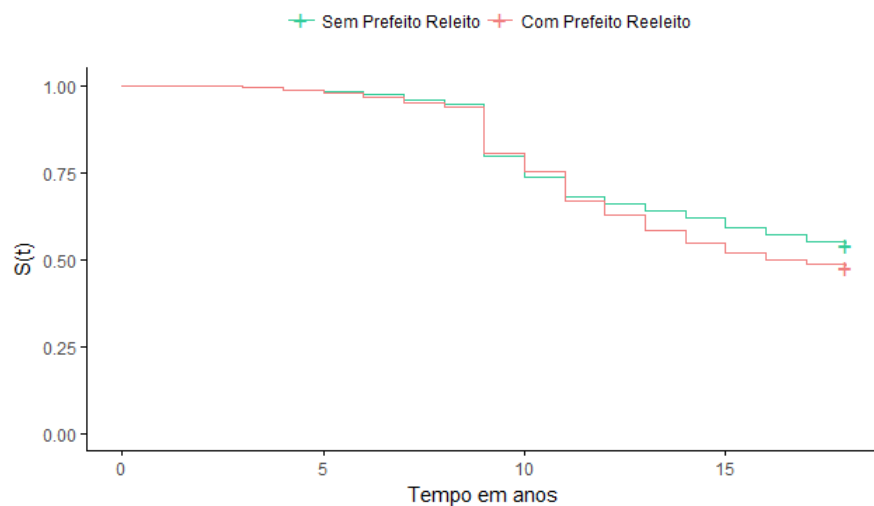


Figura 7: Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão via Plano Diretor

Ao explorar a variável explicativa Conselho, que indica se, no momento da adesão ao Plano Diretor, existia no município um Conselho Municipal, o comportamento notado é semelhante ao visto para a covariável região, visto que somente a partir do tempo 9 as curvas iniciam um afastamento, mostrando que a sobrevivência para municípios com conselhos municipais é inferior.

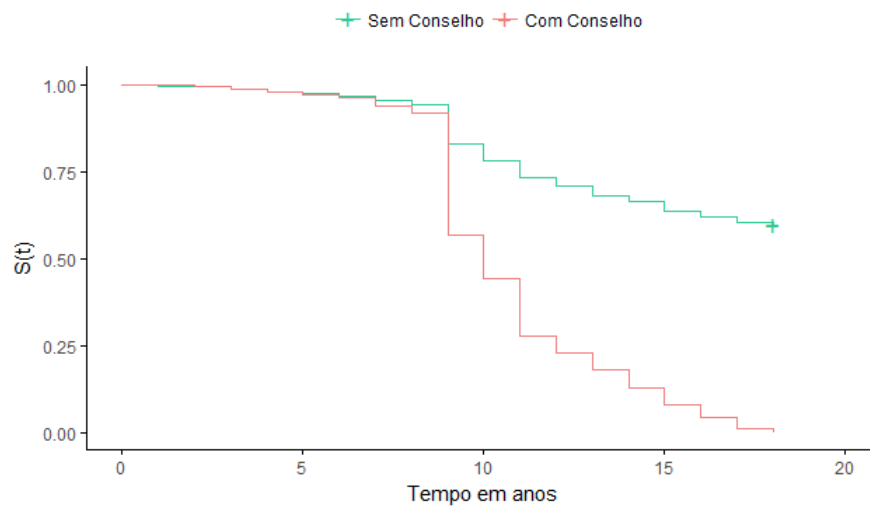


Figura 8: Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão via Plano Diretor

4.1.2 Estimação da Distribuição Log-Logística Discreta

De acordo com o informado na seção 4.1.1, a partir das análises realizadas pelas determinações empíricas, foram escolhidas as distribuições Weibull e Log-Logística para realizar o ajuste aos dados. Porém, ao tentar otimizar a Weibull, não houve convergência.

Logo, a distribuição de probabilidades Log-Logística foi escolhida para ajustar o tempo de sobrevivência até a adesão da ZUS via Plano Diretor.

A Figura 7 apresenta o ajuste aos dados e indica que a distribuição de probabilidades Log-Logística aparenta apresentar um bom ajuste para o conjunto de dados.

É interessante notar que, durante o 9º ano de estudo, ocorre uma maior quantidade de falhas, o que causa uma queda acentuada na função de sobrevivência no ano de 2006. Tal fato dificulta o ajuste de uma distribuição de probabilidades e diminui sua qualidade, fazendo com que até o tempo 9 as probabilidades sejam subestimadas e posteriormente superestimadas.

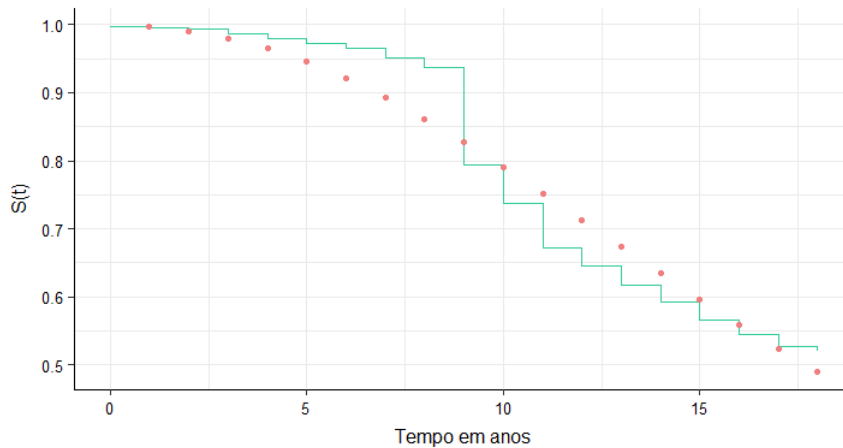


Figura 9: Ajustes da Distribuição Log-Logística Discreta para a Adesão via Plano Diretor

Tabela 1: Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão via Plano Diretor

Parâmetro	Estimativa	Erro Padrão	P-valor
α	18,6883	0,2057	<0,0001
γ	2,5022	0,0431	<0,0001

É interessante notar que o valor estimado do parâmetro α é alto, o que comprova a heterogeneidade dos dados e reafirma a necessidade de criar um modelo de regressão.

4.1.3 Estimação do Modelo de Regressão Log-Logístico Discreto

As análises descritivas realizadas na Seção 4.1.1 mostraram quais das covariáveis categóricas aparentam influenciar o tempo de sobrevivência até a adesão do Plano Diretor e foram capazes de direcionar quais variáveis provavelmente irão compor o modelo.

Para todas as variáveis categorizadas foi necessária a criação de variáveis *dummy* para que a informação qualitativa conseguisse ser inserida no modelo. No caso da variável região, foi necessária a criação de 4 *dummys*, visto que para variáveis com c categorias, são necessárias $c - 1$ *dummys*.

Como apenas a variável referente à reeleição do prefeito aparentou não causar grandes mudanças no tempo até a falha, buscou-se realizar a otimização do modelo generalizado, incluindo também as variáveis quantitativas Margem de Vitória e NEP. Porém, ocorreram dificuldades na convergência da otimização, o que impossibilitou a continuação do estudo a partir desse modelo.

Dessa forma, o modelo encontrado foi construído a partir de modelos com apenas uma covariável, sendo incluídas ou não, de acordo com os resultados de significância, novas variáveis.

As estimativas encontradas para o β_1 nos modelos com apenas uma variável estão presentes na Tabela 2, as estimativas dos outros dois parâmetros, β_0 e γ , em todos os modelos, possuem valores próximos de 3 e 2,6, respectivamente.

Tabela 2: Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão via Plano Diretor

Parâmetro	Estimativa	Erro Padrão	P-valor
Somente Conselho	-0,6350	0,0234	<0,0001
Somente Prefeito	-0,0841	0,0208	<0,0001
Somente Eleição	-0,5906	0,0258	<0,0001
Somente Margem de Vitória	0,0824	0,0485	0,0893
Somente NEP	-0,2746	0,0185	<0,0001
Somente Região			
Norte	0,1445	0,0390	<0,0001
Centro Oeste	0,4047	0,0407	<0,0001
Nordeste	0,3977	0,0270	<0,0001
Sudeste	0,2978	0,0270	<0,0001

A partir das estimativas encontradas para os modelos, nota-se que, para todas as variáveis dicotômicas, os valores estimados são negativos, indicando redução no tempo mediano mais um para os municípios que possuam essas características.

Para a covariável região, a categoria de referência foi a região sul, visto que sua curva foi a mais distante das demais. Confirmando o indicado pela análise descritiva realizada na Seção 4.1.1, todas as estimativas apresentaram valores positivos, caracterizando aumento no tempo mediano mais um dessas regiões quando comparadas ao sul do país.

Baseando-se no procedimento descrito acima e na análise de resíduos que será mostrada a seguir, o modelo escolhido englobou todas as covariáveis, exceto região. Dessa forma, o modelo considerou a seguinte função de ligação:

$$g(\eta) = \exp(\beta_0 + \beta_1 \text{Conselho} + \beta_2 \text{Prefeito} + \beta_3 \text{Eleição} + \beta_4 \text{MV} + \beta_5 \text{NEP}),$$

resultando nas estimativas apresentadas a seguir para os parâmetros do modelo escolhido:

Tabela 3: Estimativas do Modelo de Regressão Log-Logístico para a Adesão via Plano Diretor

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	3,5542	0,0418	<0,0001
β_1	-0,5135	0,0216	<0,0001
β_2	-0,0760	0,0181	<0,0001
β_3	-0,4904	0,0229	<0,0001
β_4	-0,1305	0,0442	0,0031
β_5	-0,2295	0,0180	<0,0001
γ	3,1909	0,0557	<0,0001

A partir dos resultados obtidos acima, foi realizada a análise de resíduos com Cox-Snell, apresentada na Figura 10. Como as curvas estão bem próximas, a qualidade de ajuste do modelo pode ser considerada razoável.

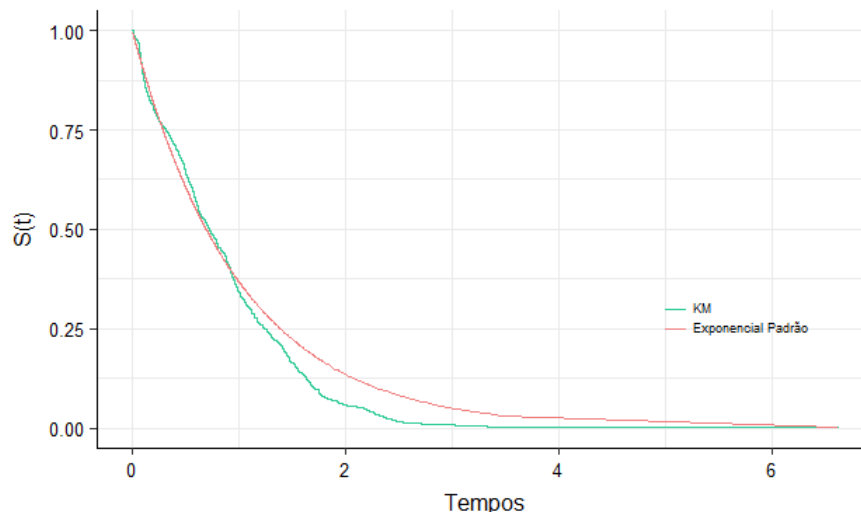


Figura 10: Análise de Resíduos de Cox-Snell para o Modelo de Plano Diretor

Por fim, a partir das estimativas encontradas, percebe-se que o tempo mediano de sobrevivência mais um ano até a adesão da política ZUS via Plano Diretor pelos municípios é cerca de $1/\exp(-0,5135) = 1,67$ vezes maior para os municípios que não possuem um conselho municipal. Logo, municípios que possuem um conselho municipal apresentam maior risco de aderir a política ZUS via Plano Diretor do que aqueles que não possuem.

De forma análoga, municípios com prefeito reeleito possuem maior risco de aderir a política ZUS via Plano Diretor do que municípios que não tem prefeito reeleito. E, em anos eleitorais, os municípios têm maior risco de aderir a política ZUS via Plano Diretor do que em anos sem eleições.

Para as variáveis quantitativas Margem de Vitória e NEP, pode-se dizer que a cada aumento em uma unidade dessas variáveis, o tempo mediano de sobrevivência mais um ano é reduzido em $\exp(-0,1305) = 0,88$ e $\exp(-0,2295) = 0,79$ vezes, respectivamente e, em contrapartida, com o aumento de Margem de Vitória e NEP o risco de falha aumenta.

4.2 Política Específica

4.2.1 Análise Descritiva

Assim como realizado para o Plano Diretor, utilizou-se a estimação não paramétrica de Kaplan-Meier para estimar a função de sobrevivência dos municípios que aderiram à ZUS via Política Específica.

Pela Figura 11, nota-se que o comportamento da probabilidade de sobrevivência pela Política Específica é bastante similar ao observado para o Plano Diretor, visto que ambas as curvas não se aproximam de zero.

Dentre os 5.167 municípios considerados, apenas 29,3% aderiram à ZUS via Política Específica até 2015, valor bem inferior ao encontrado para adesões via Plano Diretor, o que pode indicar preferência pela adesão via Plano.

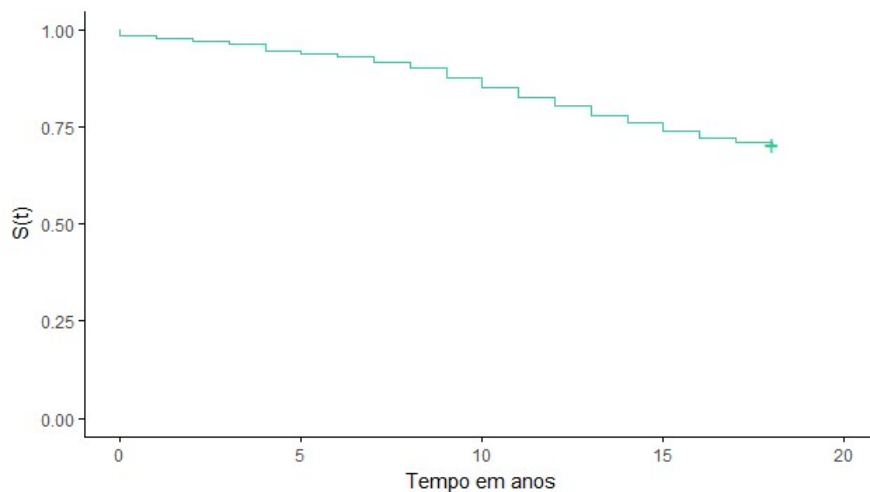


Figura 11: Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão via Política Específica

Iniciando o processo de investigação para identificar alternativas de distribuições de probabilidade para o tempo até a adesão da política ZUS, foram observados os gráficos de determinação empírica. E, assim como as indicações observadas para os dados de Plano Diretor, ambos os gráficos da Figura 12 apontam para uma função de risco acumulado com comportamento monótono crescente.

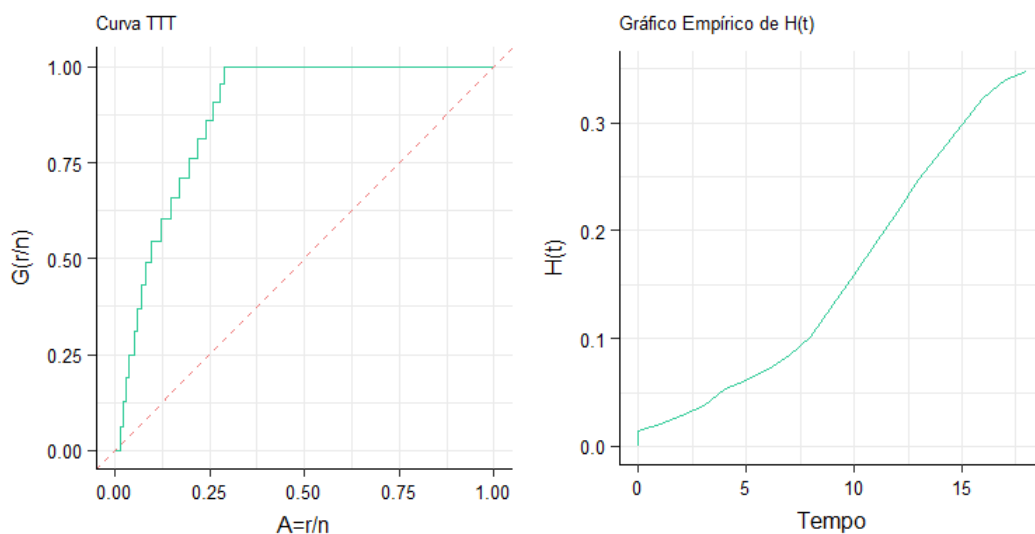


Figura 12: Determinações Empíricas para a Adesão via Política Específica

Para finalizar a descrição do tempo até a adesão da ZUS via Política Específica, foram inseridas, ao gráfico da função de sobrevivência, as informações sobre as quatro variáveis categóricas consideradas no estudo.

A partir da Figura 13 observa-se a probabilidade de sobrevivência dos municípios separada pelas regiões do país. As regiões Centro-Oeste, Norte e Sudeste apresentam comportamento de sobrevivência bastante similar em todo o período de estudo.

O Sul do país possui a maior queda na probabilidade de sobrevivência, o que indica uma maior adesão via Política Específica para os municípios sulistas, enquanto o Nordeste se destaca pela não adesão.

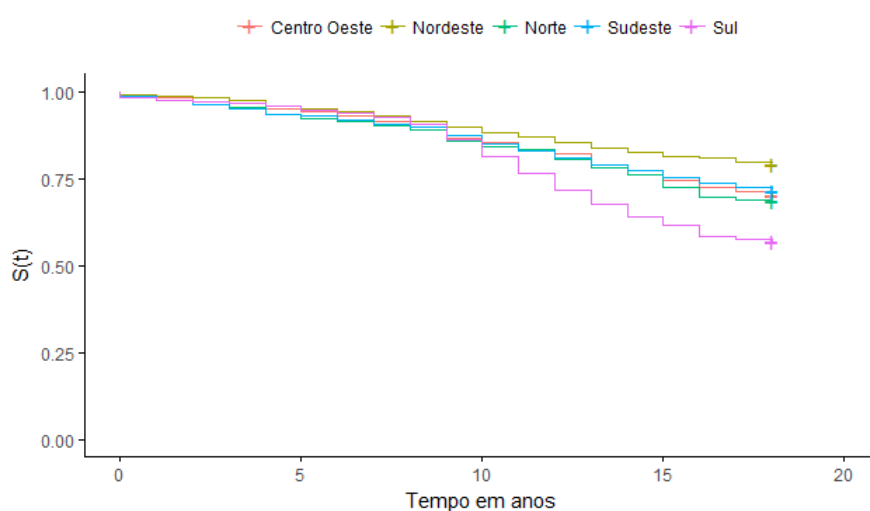


Figura 13: Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão via Política Específica

A Figura 14 apresenta duas curvas de sobrevivência completamente destoantes, mostrando que, em anos de eleição, a adesão da ZUS é mais intensa e, portanto, a probabilidade de sobrevivência decresce mais rapidamente.

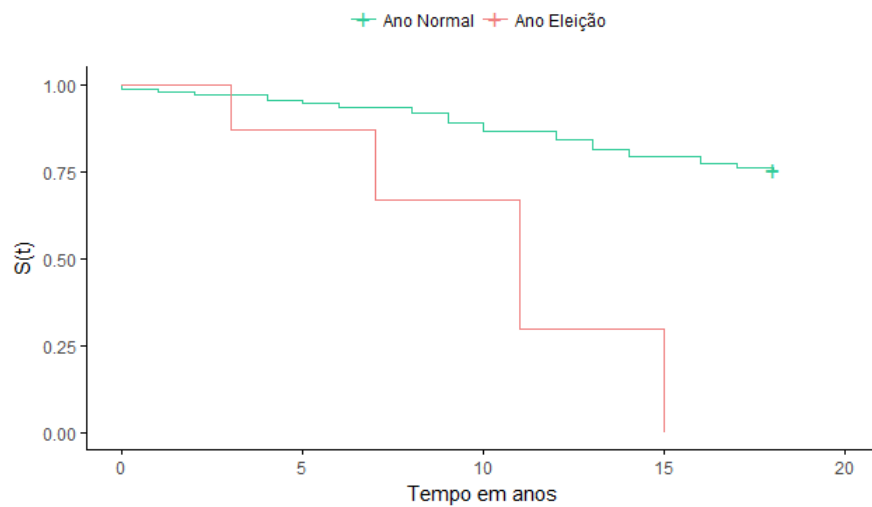


Figura 14: Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão via Política Específica

Diferentemente do observado para a variável ano de eleição, nota-se, na Figura 15, que a covariável referente ao prefeito do município ser reeleito não impacta fortemente a probabilidade de sobrevivência. Ainda assim, os municípios que possuem prefeitos reeleitos apresentam probabilidades de sobrevivência inferiores a partir do tempo 3.

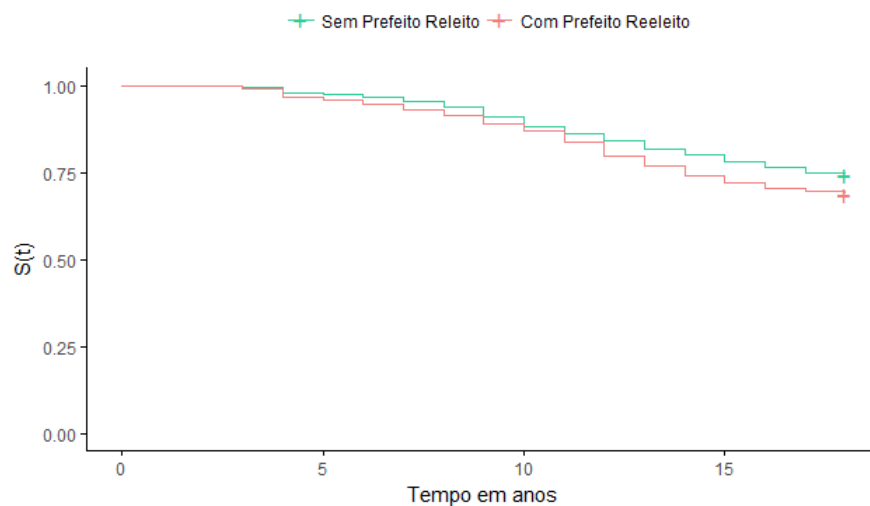


Figura 15: Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão via Política Específica

Estudando a variável referente à existência de um conselho municipal de política urbana, percebe-se, diferentemente de todas as outras, uma mudança de comportamento ao longo do tempo.

Até o tempo 11, municípios com conselhos apresentam probabilidade de sobrevivência superior, porém a partir do tempo 8, estes municípios começam a apresentar

uma queda na curva de sobrevivência mais acentuada, fazendo com que, nos anos finais do estudo, municípios sem conselho municipais tenham aderido menos à ZUS via Política Específica.

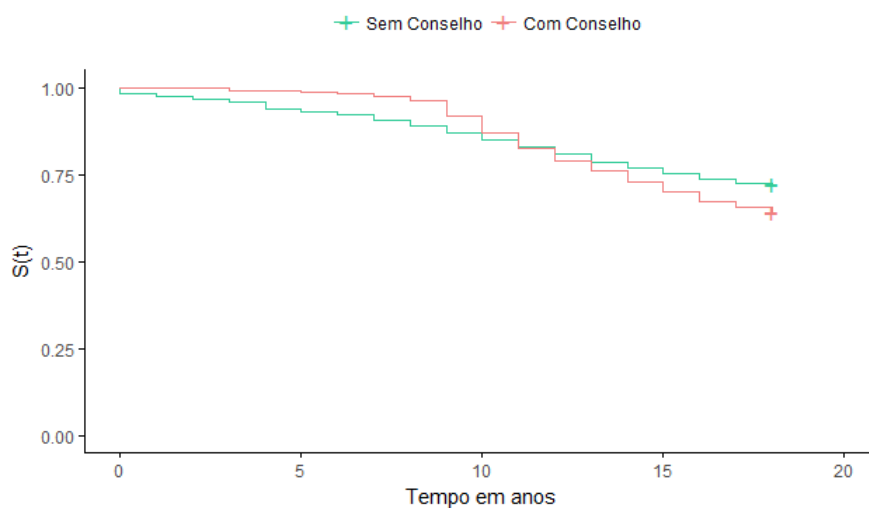


Figura 16: Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão via Política Específica

4.2.2 Estimação do Modelo Log-Logístico Discreto

Assim como feito para o Plano Diretor, a distribuição escolhida para ajustar os dados, apesar de possuir função de risco unimodal, foi a distribuição Log-Logística, visto que a distribuição Weibull apresentou dificuldades para atingir a convergência na otimização.

Para verificar a qualidade do ajuste dessa distribuição aos dados, foram plotadas as estimativas encontradas via Kaplan-Meier para a função de sobrevivência e os valores encontrados para a distribuição Log-Logística com os parâmetros determinados via otimização.

De acordo com os resultados demonstrados pela Figura 17, a distribuição Log-Logística aparenta ajustar-se bem aos dados considerados.

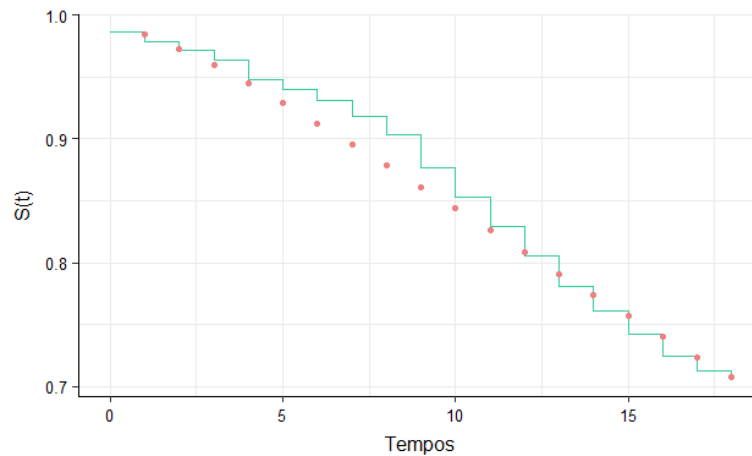


Figura 17: Ajustes da Distribuição Log-Logística Discreta para a Adesão via Política Específica

Tabela 4: Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão via Política Específica

Parâmetro	Estimativa	Erro Padrão	P-valor
α	34,7277	0,9560	<0,0001
γ	1,4658	0,0360	<0,0001

Bastante similar aos resultados obtidos para a estimação via Plano Diretor, o parâmetro α foi estimado para um valor alto, o que caracteriza a heterogeneidade dos dados de adesão via Plano Diretor.

4.2.3 Estimação do Modelo de Regressão Log-Logístico Discreto

Os mesmos procedimentos descritos na Seção 4.1.3 foram realizados para a construção de um modelo de regressão que conseguisse ajustar os dados de adesão da ZUS via Política Específica.

Sendo assim, a Tabela 5 apresenta as estimativas encontradas para o β_1 nos modelos com apenas uma cováriavel, sendo as estimativas de β_0 e γ sempre próximas de 3,7 e 1,7, respectivamente.

Tabela 5: Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão via Política Específica

Parâmetro	Estimativa	Erro Padrão	P-valor
Somente Conselho	-0,1508	0,0493	0,0022
Somente Prefeito	-0,1400	0,0326	<0,0001
Somente Eleição	-1,3211	0,0595	<0,0001
Somente Margem de Vitória	-0,1609	0,0903	0,0747
Somente NEP	-0,2541	0,0398	<0,0001
Somente Região			
Norte	0,2668	0,0754	<0,0001
Centro Oeste	0,4412	0,0790	<0,0001
Nordeste	0,6790	0,0551	<0,0001
Sudeste	0,4528	0,0534	<0,0001

A partir das estimativas encontradas para os modelos, nota-se que, para todas as variáveis dicotômicas, os valores estimados são negativos, indicando redução no tempo mediano mais um para os municípios que possuam essas características.

Assim como o realizado para o Plano Diretor, a região sul foi definida como categoria de referência para a variável região, devido ao distanciamento da sua curva de sobrevivência das demais. Confirmando o indicado pela análise descritiva realizada na Seção 4.2.1, todas as estimativas apresentaram valores positivos, caracterizando aumento no tempo mediano mais um dessas regiões quando comparadas ao sul do país, sendo o nordeste a região com a maior estimativa.

Considerando o modelo de regressão com todas as variáveis, exceto região, a função de ligação considerada foi dada por:

$$g(\eta) = \exp(\beta_0 + \beta_1 \text{Conselho} + \beta_2 \text{Prefeito} + \beta_3 \text{Eleição} + \beta_4 \text{MV} + \beta_5 \text{NEP})$$

e as estimativas obtidas para os parâmetros do modelo pela otimização estão presentes na Tabela 6 a seguir.

Tabela 6: Estimativas do Modelo de Regressão Log-Logístico para a Adesão via Política Específica

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	3,9207	0,0702	<0,0001
β_1	-0,1177	0,0330	0,0003
β_2	-0,1199	0,0298	<0,0001
β_3	-1,0993	0,0426	<0,0001
β_4	-0,2183	0,0666	0,0001
β_5	-0,1899	0,0295	<0,0001
γ	2,3693	0,0588	<0,0001

Avaliando a qualidade do ajuste do modelo encontrado, foi realizada a análise gráfica dos resíduos de Cox-Snell, presente na Figura 18.

Considerando os resíduos observados para o modelo de Plano Diretor, os resíduos para o modelo de Política Específica indicam um ajuste de pior qualidade, ainda assim, pode-ser considerado razoável.

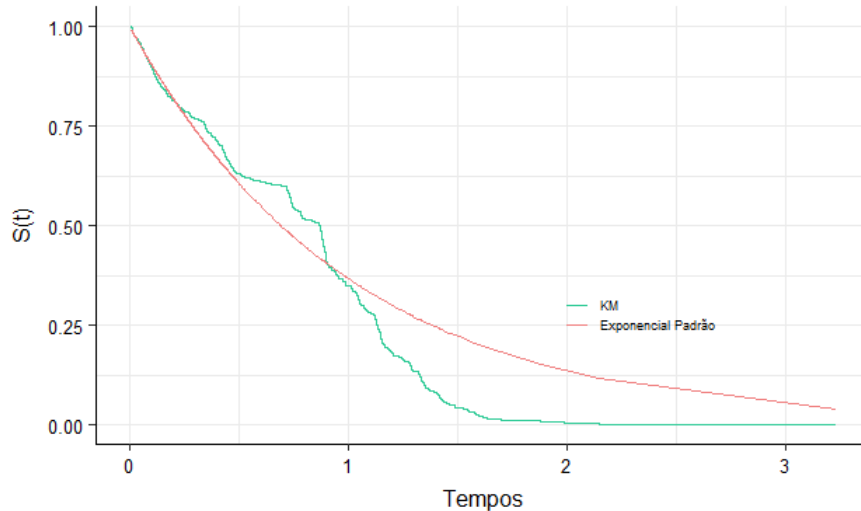


Figura 18: Análise de Resíduos de Cox-Snell para o Modelo de Política Específica

Verificada a qualidade do ajuste, pode-se interpretar os valores estimados encontrados e assim como visto para os dados de Plano Diretor, todas os parâmetros possuem valores de estimativas inferiores a zero.

No caso das variáveis quantitativas Margem de Vitória e NEP, as estimativas negativas indicam que o aumento em uma unidade das mesmas reduz o tempo mediano de sobrevivência mais um ano em $\exp(-0,2183) = 0,8038$ e $\exp(-0,1899) = 0,8270$ vezes, respectivamente.

Para as variáveis dicotômicas, percebe-se que a existência de um conselho, o prefeito ser reeleito e o ano ser eleitoral afetam na redução do tempo mediano mais um ano, visto que suas estimativas também são negativas, ou seja, municípios que apresentam essas características possuem maior risco de aderir a política ZUS via legislação específica.

4.3 Considerando Ambas as Formas de Adesão

4.3.1 Análise Descritiva

Para finalizar as análises feitas, foi considerado um terceiro banco de dados que desconsidera a forma como os municípios aderiram à ZUS, unificando as informações de falha presentes nas Seções 4.1 e 4.2.

A partir dessa unificação, nota-se que o comportamento da curva de sobrevivência total possui características observadas nas curvas encontradas para o Plano Diretor e

Política Específica, o que já era esperado.

Características marcantes vistas no Plano Diretor, como a queda mais acentuada no tempo 9, tornaram-se mais suaves devido ao comportamento observado para a curva de Política Específica, que apresenta quedas mais tênues.

Por volta de 45,21% dos municípios estudados ainda não haviam aderido à ZUS em 2015, o que indica a baixa taxa de adesão da ZUS pelos governos brasileiros.

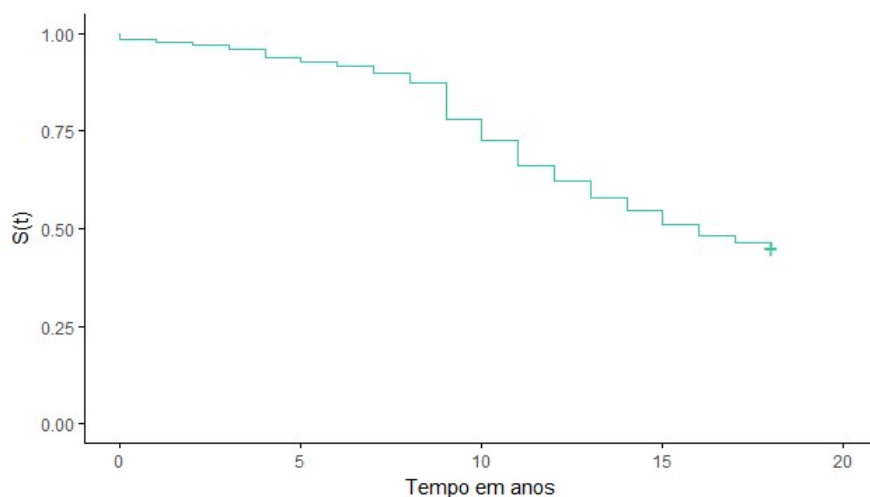


Figura 19: Função de Sobrevivência Estimada via Kaplan-Meier para a Adesão Total

Assim como os resultados obtidos para os dois bancos anteriores, os gráficos de determinação empírica, apresentados na Figura 20, indicam que distribuições que possam apresentar bons ajustes devem possuir função de risco acumulado monótono crescente.

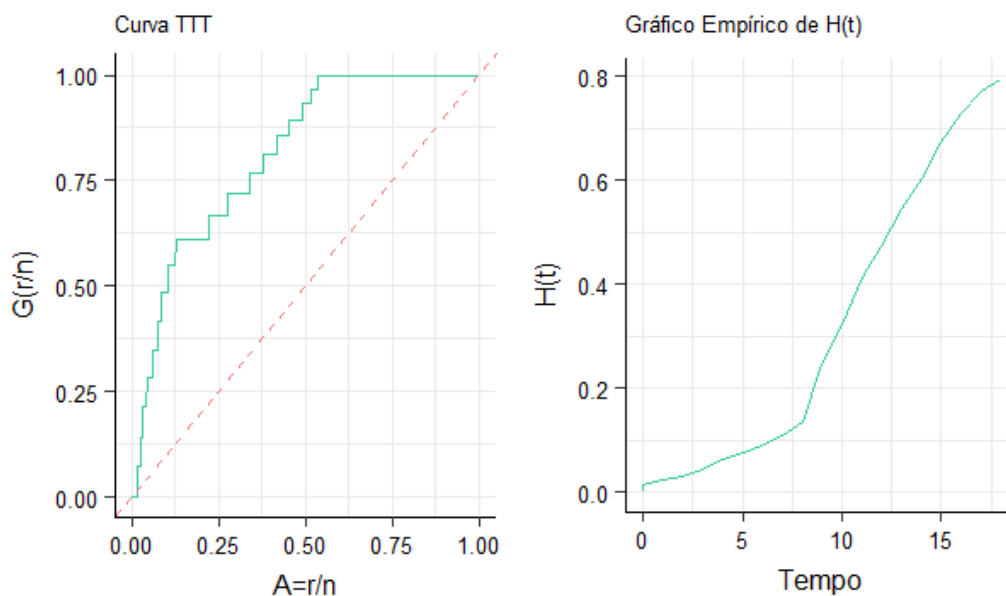


Figura 20: Determinações Empíricas para a Adesão Total

Dentre as curvas de sobrevivência estimadas por região, o Nordeste apresenta a menor quantidade de falhas, sendo, ao final do estudo, a região com as maiores probabilidades de sobrevivência. Em contrapartida, a região Sul se destaca das demais a partir do tempo 10, sendo a região com a maior porcentagem de municípios com a ZUS.

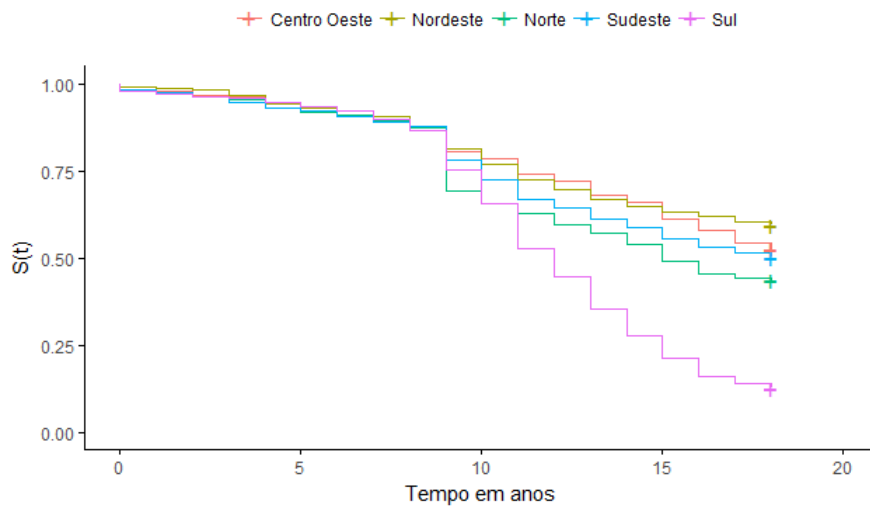


Figura 21: Função de Sobrevivência Estimada via Kaplan-Meier por Região para a Adesão Total

Apresentando os mesmos resultados encontrados para os bancos de dados anteriores, a curva de sobrevivência para municípios que aderiram em anos eleitorais possui quedas mais bruscas e passa a possuir probabilidades menores de sobrevivência a partir do tempo 12.

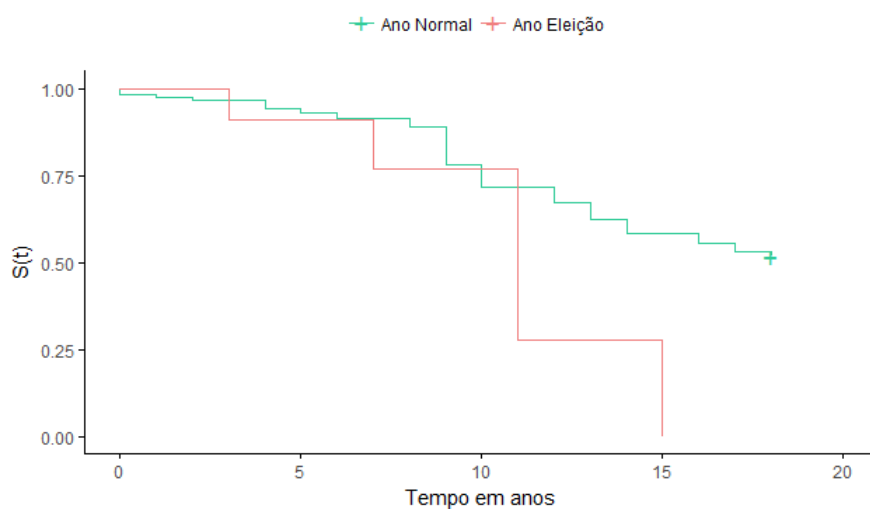


Figura 22: Função de Sobrevivência Estimada via Kaplan-Meier por Ano de Eleição para a Adesão Total

A covariável referente ao prefeito ser reeleito apresenta a menor distância entre as

curvas, sendo a probabilidade de sobrevivência para municípios com governantes reeleitos sempre menor ou igual à probabilidade dos municípios com novos prefeitos.

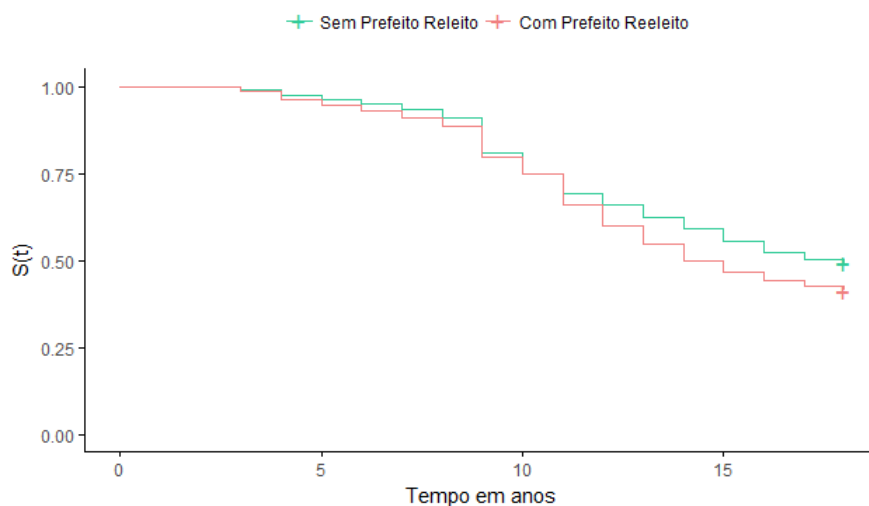


Figura 23: Função de Sobrevivência Estimada via Kaplan-Meier por Prefeito Reeleito para a Adesão Total

Por fim, sobre a variável referente a existência de conselho municipal de política urbana, a Figura 25 mostra que, a partir do tempo 9, municípios com conselho aderem mais à ZUS.

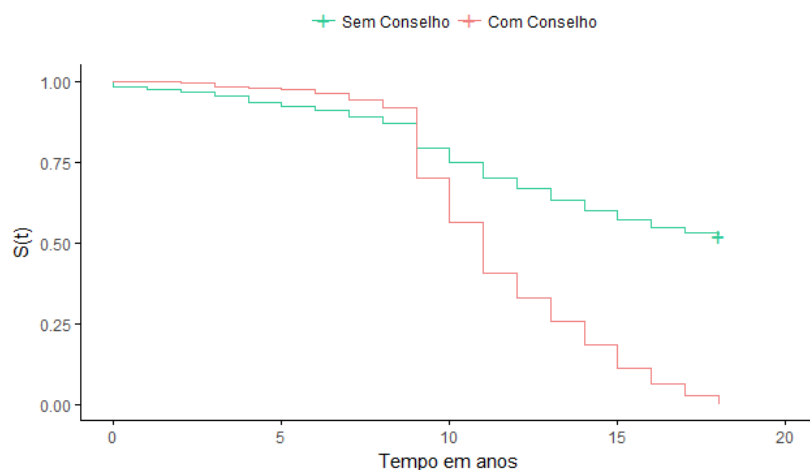


Figura 24: Função de Sobrevivência Estimada via Kaplan-Meier por Conselho para a Adesão Total

4.3.2 Estimação do Modelo Log-Logístico Discreto

Assim como feito para os bancos de dados anteriores, a distribuição escolhida para ajuste foi a Log-Logística, visto que ocorreram problemas de convergência na otimização

para a distribuição Weibull.

A Figura 24 demonstra que o ajuste da distribuição Log-Logística discreta pode ser considerado bom, apesar das estimativas das probabilidades de sobrevivência serem subestimadas entre os tempos 4 e 9.

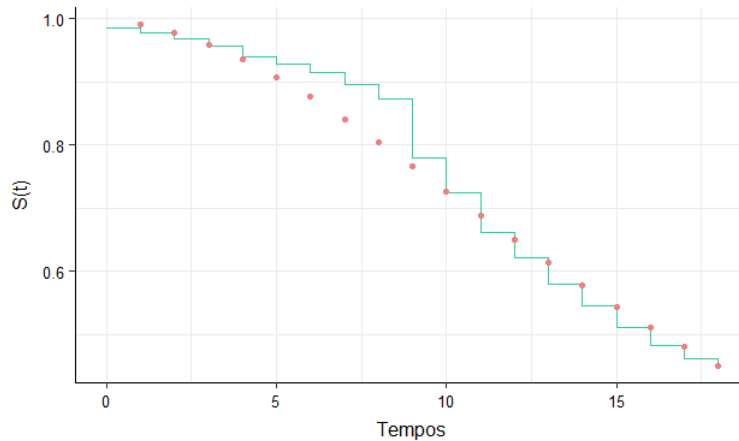


Figura 25: Ajustes da Distribuição Log-Logística Discreta para a Adesão Total

Tabela 7: Estimativas dos Parâmetros da Distribuição Log-Logística para a Adesão Total

Parâmetro	Estimativa	Erro Padrão	P-valor
α	17,3446	0,2118	<0,0001
γ	2,1535	0,0367	<0,0001

Reforçando mais uma vez o que foi visto para os dois bancos de dados, o valor da estimativa encontrada para o parâmetro α também foi grande.

4.3.3 Estimação Modelo de Regressão Log-Logístico Discreto

Assim como feito nas duas Seções anteriores, primeiramente foram construídos os modelos de regressão com apenas uma variável e, a partir deles, chegou-se no modelo final.

A Tabela 8 apresenta as estimativas encontradas para o β_1 nos modelos individuais, sendo as estimativas de β_0 e γ por volta de 2,9 e 2,5 em todos os modelos com apenas uma variável.

Tabela 8: Estimativas do Modelos de Regressão Log-Logístico com Uma Covariável para a Adesão Total

Parâmetro	Estimativa	Erro Padrão	P-valor
Somente Conselho	-0,4944	0,0286	<0,0001
Somente Prefeito	-0,1044	0,0204	<0,0001
Somente Eleição	-0,5548	0,0297	<0,0001
Somente Margem de Vitória	-0,0183	0,0547	0,7378
Somente NEP	-0,2335	0,0226	<0,0001
Somente Região			
Norte	0,2736	0,0452	<0,0001
Centro Oeste	0,4579	0,0462	<0,0001
Nordeste	0,5329	0,0307	<0,0001
Sudeste	0,3818	0,0307	<0,0001

Assim como visto para os bancos de dados anteriores, todas as covariáveis apresentam valores estimados para os parâmetros do modelo negativos. Para região, todas as regiões possuem taxa de falha menor do que a sul, visto que suas estimativas são maiores do que zero.

Partindo desses modelos iniciais e assim como as regressões anteriores, o modelo final encontrado engloba todas as covariáveis, exceto região. Portanto, a função de ligação utilizada foi dada por:

$$g(\eta) = \exp(\beta_0 + \beta_1 \text{Conselho} + \beta_2 \text{Prefeito} + \beta_3 \text{Eleição} + \beta_4 \text{MV} + \beta_5 \text{NEP}).$$

As estimativas obtidas para os parâmetros do modelo estão presentes na Tabela 9 a seguir.

Tabela 9: Estimativas do Modelo de Regressão Log-Logístico para a Adesão Total

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	3,4623	0,0452	<0,0001
β_1	-0,3931	0,0223	<0,0001
β_2	-0,0883	0,0187	<0,0001
β_3	-0,4775	0,0230	<0,0001
β_4	-0,2029	0,0460	<0,0001
β_5	-0,2097	0,0197	<0,0001
γ	3,0854	0,0527	<0,0001

A partir dos resultados obtidos acima, foi realizada a análise de resíduos com Cox-Snell, apresentada na Figura 26. Como as curvas estão bem próximas, a qualidade de ajuste do modelo pode ser considerada boa.

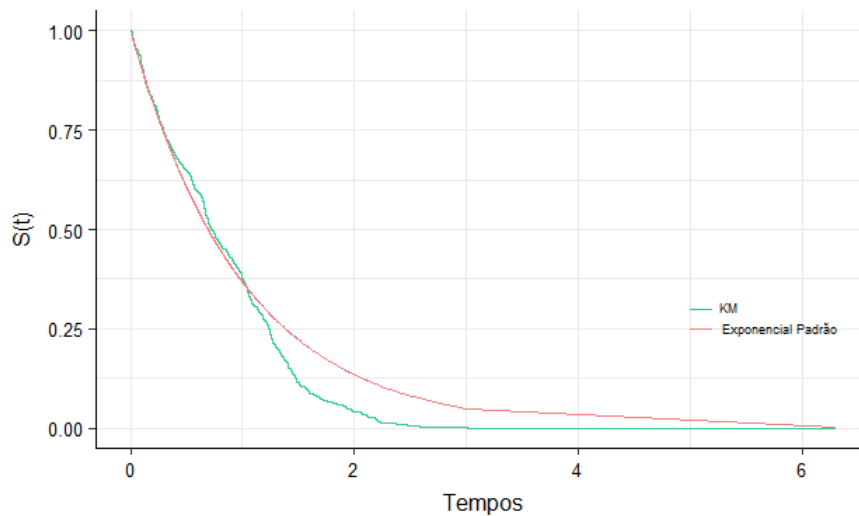


Figura 26: Análise de Resíduos de Cox-Snell para o Modelo Total para a Adesão Total

Por fim, após verificada a qualidade do ajuste do modelo, percebe-se pelos valores negativos encontrados para as estimativas dos parâmetros do modelo que todas as variáveis influenciam a redução do tempo mediano de sobrevivência mais um ano dos municípios.

A existência de um conselho municipal, o fato do prefeito em governo ser reeleito e o ano ser eleitoral reduzem o tempo mediano mais um ano $\exp(-0,3931) = 0,6749$, $\exp(-0,0883) = 0,9154$ e $\exp(-0,4775) = 0,6203$ vezes.

4.4 Conclusão

Comparando os resultados obtidos para os três bancos de dados, todos os modelos foram formados pelas mesmas variáveis e apresentaram ajustes razoáveis confirmados a partir dos resíduos de Cox-Snell.

De acordo com as estimativas encontradas para os três modelos, todas as variáveis impactam a probabilidade de sobrevivência reduzindo-a, e, conseqüentemente, o risco de falha aumenta. Além disso, todas as covariáveis presentes nos modelos finais foram significativas considerando os níveis de significância de 5% e 1%.

Para as variáveis dicotômicas conselho, prefeito reeleito e eleição, percebe-se que, em todos os três modelos, as estimativas dos parâmetros apresentaram valores negativos, indicando a redução do tempo mediano mais um ano para os municípios que apresentavam essas características.

Para as variáveis quantitativas margem de vitória e NEP, as estimativas também apresentaram valores menores que zero para todos os modelos construídos, mostrando que o aumento em uma unidade dessas variáveis causa diminuição no tempo mediano mais um ano, aumentando o risco de falha de adesão da política ZUS.

Em relação à variável região, a partir das análises descritivas e dos resultados encontrados para o modelo individual que a considera, nota-se que a probabilidade de sobrevivência dos municípios muda sim de acordo com a região em que o município está inserido, sendo o risco de falha maior para a região Sul do Brasil, entretanto, devido à problemas de otimização não foi possível inserir a mesma junto às outras covariáveis.

Por fim, a forma de tratar o tempo até a adesão da política ZUS não causa impactos no que influencia o tempo e não modifica os resultados, porém, comparando os resíduos de Cox-Snell, percebe-se que o banco que considerou as informações de falha provindas tanto do Plano Diretor quanto da Política Específica apresentou o melhor ajuste, fato esse que pode ser justificado pelo maior volume de informações utilizadas.

Referências

- Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, v. 36:p. 106–108.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley, Florida.
- Bohoris, G. (1994). Comparison of the cumulative-hazard and kaplan-meier estimators of the survivor function. *IEEE Transactions on Reliability*, v. 43:p. 230–232.
- Colosimo, E. A. and Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. Edgard Blucher, São Paulo. ABE - Projeto Fisher.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley.
- Nakano, E. Y. (2017). Um curso de análise de sobrevivência. Departamento de Estatística, Universidade de Brasília, Brasília.
- Nobre, L. P. (2016). Modelo de regressão weibull pada dados discretos em análise de sobrevivência. Trabalho de Conclusão de Curso - Departamento de Estatística, Universidade de Brasília, Brasília.
- Santos, D. F. (2017). Modelo de regressão log-logístico discreto com fração de cura para dados de sobrevivência. Dissertação (Mestrado em Estatística) - Departamento de Estatística, Universidade de Brasília, Brasília.
- Vaz, J. C. (2006). *Legislação de Uso e Ocupação do Solo*. Instituto Pólis.