



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE CIÊNCIA POLÍTICA
PROGRAMA DE GRADUAÇÃO EM CIÊNCIA POLÍTICA

FIDEL PARENTE DE ALENCAR PIMENTEL

**ANÁLISE DE SOBREVIVÊNCIA EM R:
GUIA APLICADO PARA DIFUSÃO DE POLÍTICAS PÚBLICAS**

Recife
2025

FIDEL PARENTE DE ALENCAR PIMENTEL

**ANÁLISE DE SOBREVIVÊNCIA EM R:
GUIA APLICADO PARA DIFUSÃO DE POLÍTICAS PÚBLICAS**

Trabalho de Conclusão de Curso
apresentado à Coordenação do Curso de
Bacharelado em Ciência Política da
Universidade Federal de Pernambuco,
como parte dos requisitos para a obtenção
do grau de Bacharel em Ciência Política.

Área de Concentração: Políticas Públicas,
Metodologia

Orientador: Prof. Dr. Dalson Britto Figueiredo Filho

Recife
2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Pimentel, Fidel Parente de Alencar.

Análise De Sobrevivência Em R: Guia Aplicado Para Difusão De Políticas Públicas / Fidel Parente de Alencar Pimentel. - Recife, 2025.

57p. : il.

Orientador(a): Dalson Britto Figueiredo Filho Figueiredo

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Filosofia e Ciências Humanas, Ciência Política, 2025.

Inclui referências.

1. Difusão de Políticas públicas. 2. Descentralização de Políticas públicas. 3. Análise de Sobrevivência. 4. Event History Analysis. I. Figueiredo, Dalson Britto Figueiredo Filho. (Orientação). II. Título.

320 CDD (22.ed.)

FIDEL PARENTE DE ALENCAR PIMENTEL

**ANÁLISE DE SOBREVIVÊNCIA EM R:
GUIA APLICADO PARA DIFUSÃO DE POLÍTICAS PÚBLICAS**

Trabalho de Conclusão de Curso
apresentado ao Curso de Ciência Política
da Universidade Federal de Pernambuco,
como parte dos requisitos para a obtenção
do grau de Bacharel em Ciência Política.

Aprovação: 14/04/2025

BANCA EXAMINADORA

Prof. Dr. Dalson Britto Figueiredo Filho (Orientador)
Universidade Federal de Pernambuco

Profa. Dra. Maria do Carmo Soares de Lima (Examinadora Interna)
Universidade Federal de Pernambuco

Prof. Dr. Rodrigo Galvão Pinho Lins (Examinador Externo)
Universidade Federal de Alagoas

Prof. Jorge Alexandre Barbosa Neves (Examinador Externo)
Universidade Federal de Minas Gerais

RESUMO

A literatura sobre difusão de políticas frequentemente se ocupa de compreender padrões de adoção de determinadas políticas públicas pelos municípios brasileiros. Particularmente, muitos trabalhos buscam identificar os fatores determinantes desses processos de difusão. O presente trabalho tem como objetivo sistematizar, em forma de tutorial, o passo a passo da implementação de análises de sobrevivência para investigação de fenômenos de difusão. O foco está nos aspectos técnicos da implementação computacional da metodologia, em linguagem R. Embora a área esteja repleta de avanços teóricos e metodológicos, a literatura disponível costuma ter uma abordagem pouco didática para quem não possui familiaridade com estatística e deseja realizar uma análise deste tipo. O trabalho apresenta o referencial conceitual para a aplicação da análise de sobrevivência e propõe o caso prático de análise dos planos diretores municipais. O levantamento cobre a demonstração das análises descritivas, da implementação do modelo de riscos proporcionais de Cox e da verificação dos seus pressupostos. Com o estudo, espera-se facilitar o processo de pesquisa quantitativa para investigadores iniciantes e intermediários da área. Códigos e dados estão disponíveis na internet.

Palavras chave: Políticas públicas. Difusão. Descentralização. Event history analysis. Análise de Sobrevivência

ABSTRACT

The literature on policy diffusion often focuses on understanding patterns in the adoption of certain public policies by Brazilian municipalities. In particular, many studies aim to identify the determinants of these diffusion processes. The present study aims to systematize, in the form of a tutorial, the step-by-step implementation of survival analysis to investigate the phenomenon of diffusion. The focus lies on the technical aspects of the computational implementation of the methodology using the R programming language. Although the field has seen significant theoretical and methodological advancements, the available literature tends to lack a didactic approach for those unfamiliar with statistics who wish to conduct this type of analysis. This work presents the conceptual framework for applying survival analysis and proposes a practical case study involving the analysis of municipal master plans. The study includes demonstrations of descriptive analyses, the implementation of the Cox proportional hazards model, and the verification of its assumptions. The aim is to facilitate the process of quantitative research for beginner and intermediate researchers in the field. All code and data are available online.

Keywords: Public policy. Diffusion. Decentralization. Event history analysis. Survival analysis.

SUMÁRIO

SUMÁRIO.....	4
INTRODUÇÃO.....	5
2 COMPREENDENDO OS DADOS DE SOBREVIVÊNCIA.....	8
2.1. Conceitos de análise de sobrevivência.....	8
2.2. Planos Diretores Municipais.....	10
2.3. Estrutura dos dados.....	14
2.4. Variável Dependente.....	15
2.5. Tabela de Vida.....	17
2.6. Estimador não-paramétrico Kaplan-Meier.....	18
2.7. Resultados.....	23
3 EVENT HISTORY ANALYSIS COMO ANÁLISE DE SOBREVIVÊNCIA.....	24
3.1. Introdução.....	24
3.2. Teoria e determinantes da difusão de políticas.....	24
3.3. Regressão de Cox.....	28
3.4. Implementando o modelo de Cox.....	29
3.4.1. <i>Variáveis Independentes</i>	29
3.4.2. <i>Variável dependente</i>	33
3.4.3. <i>Modelo e Interpretação</i>	34
3.5. Resultados.....	37
4 VERIFICAÇÃO DE PRESSUPOSTOS E AJUSTES DO MODELO.....	38
4.1. Introdução.....	38
4.2. Verificando pressupostos.....	38
4.2.1. <i>Independência entre censura e sobrevivência</i>	39
4.2.2. <i>Distribuição homogênea da variável dependente</i>	40
4.2.3. <i>Ausência de multicolinearidade</i>	41
4.2.4. <i>Linearidade das variáveis contínuas</i>	43
4.2.5. <i>Proporcionalidade dos Riscos</i>	44
4.2.6. <i>Independência dos tempos de sobrevivência</i>	45
4.3. Propondo um novo modelo e Resultados.....	47
5 CONCLUSÃO.....	51
REFERÊNCIAS.....	53

INTRODUÇÃO

Este trabalho apresenta um tutorial de análise de difusão de políticas públicas em R. Nosso leitor é o pesquisador, tomador de decisões ou estudante que deseja investigar empiricamente os padrões de adoção de políticas públicas entre os municípios brasileiros. O material contribui para preencher uma lacuna acadêmica de caráter pedagógico. Enquanto a literatura acadêmica, teórica e metodológica, nas ciências sociais se avoluma, faltam materiais com abordagem prática para a implementação de atividades técnicas de pesquisa.

A principal referência com caráter pedagógico em português para implementação de análises de sobrevivência consiste em obra sobre aplicações em saúde (CARVALHO e outros, 2011). Entre as revistas mais bem classificadas pelo *Qualis Referência* é possível encontrar apenas códigos replicáveis, nenhum material com foco pedagógico. Por esse motivo, um material especialmente desenvolvido para facilitar a compreensão sobre análise de sobrevivência tem o potencial de fomentar os estudos de difusão de políticas públicas em geral e auxiliar pesquisas empíricas na área de Ciência Política, em particular.

A ciência política brasileira vem se diversificando teórica e metodologicamente. Apenas nos anos 2000 veio-se estabelecer diálogo entre a tradicional literatura sobre descentralização e a literatura sobre *difusão de políticas públicas*, trazida de outras comunidades acadêmicas. O primeiro trabalho com esta abordagem foi “*Political Competition and the diffusion of conditional cash transfers in Brazil*”, de 2012. Para Coelho, as pesquisas sobre descentralização “minimizam fatores como competição política, desenvolvimento econômico, efeitos regionais”. (COELHO, p. 58 2012) E a abordagem sobre difusão no Brasil surge para acrescentar perspectivas à consolidada pesquisa sobre descentralização de políticas públicas.

Desde então diversas políticas públicas brasileiras foram investigadas sob essa ótica. Para listar algumas, temos o Programa Bolsa Família/bolsa escola federal, por Coelho (2012) e por Bastos (2015); o Bolsa escola, por Coelho (2012) e Sugiyama (2008); as Carreira de gestor governamental, por Palotti e outras (2016); a Lei de Acesso à Informação, por Batista (2018) e Tatemoto (2016); o Programa Saúde da família, por Coelho, Cavalcante e Turgeon (2016) e a estratégia de saúde da família por Passos (2023); os Planos municipais de saneamento básico, por

Carvalho (2016); o Programa Caminhos da Escola, por Maeda (2016); o Orçamento Participativo, por Wampler (2008) e Spada (2010 e 2014).

É nesse contexto de abordagens empíricas e metodologicamente mais estruturadas sobre o federalismo brasileiro que o presente trabalho se insere. A *análise de sobrevivência*, por vezes chamada, nas ciências humanas, de *Event History Analysis* (EHA) é uma estratégia metodológica quantitativa que se destaca entre os pesquisadores da difusão de políticas públicas.

Do ponto de vista estatístico, a análise de sobrevivência nos ajuda a responder perguntas quanto ao tempo necessário para que um determinado evento de interesse ocorra ou quanto à chance (risco) que ocorra a alguém em determinado momento. Do ponto de vista da difusão de políticas públicas, a EHA nos ajuda a investigar os determinantes que contribuem para uma inovação política ou política pública se espalhe através de um sistema político.

De acordo com Berry e Berry (1990), nos anos 1990 modelos capazes de conciliar fatores internos e externos da difusão de políticas teriam chegado na ciência política¹. As abordagens de EHA são “capazes de lidar com diferentes estruturas de dados sociais, partições populacionais, matrizes de proximidade, processos temporais e observações censuradas” (STRANG, 1991, p. 326).

Para o jovem pesquisador em ciências sociais aplicadas, no entanto, é significativo o desafio de encarar a prática de análise de dados, os métodos estatístico-computacionais e os *microdados*. Ainda mais com a existência de uma lacuna entre as literaturas das disciplinas envolvidas: *descentralização federativa, difusão de políticas, análise de sobrevivência*.

Assim sendo, este trabalho apresenta um tutorial² de estatística-computacional com uso de linguagem R para a análise da descentralização de políticas públicas no Brasil sob a perspectiva acadêmica da literatura de difusão de políticas públicas. Com esse objetivo, propomos utilizar modelos de sobrevivência, ou EHA, para analisar uma política pública ainda não investigada pela literatura da área: a difusão dos Planos Diretores Municipais.

O foco do trabalho não está nos resultados encontrados para a política pública, mas em apresentar um material equilibrado, que reflita a complexidade

¹ Ciência Política norte-americana.

² Este tutorial requer alguma familiaridade com a linguagem R, que pode ser obtida através do estudo de manuais introdutórios, como Aquino (2014)

técnica da análise, que acompanhe minimamente idas e vindas do processo de ajustes de um modelo estatístico e que não perca de vista o sentido didático do documento. Assim sendo, seu objetivo é servir de referência para pesquisadores que necessitem realizar uma pesquisa como a aqui descrita³.

O presente trabalho se justifica, portanto, ao apresentar uma abordagem didática e instrucional que reúne conhecimentos teóricos, metodológicos e empíricos em formato tecnicamente reproduzível, em área do conhecimento carente deste tipo de abordagem. O produto final tem potencial para impactar a pesquisa acadêmica na área, no Brasil, por ser de interesse empírico da ciência política e de outros ramos de pesquisa. Além disso, pode impactar o trabalho de tomadores de decisão, setores sociais e da gestão pública. Justifica-se assim por apresentar instrumentos para a compreensão da realidade sócio-política brasileira.

O trabalho está organizado em 3 seções. O *primeiro capítulo* apresenta o caso dos *Planos Diretores Municipais* e recupera os dados dos municípios, com destaque para os dados de adoção da política. Na sequência, apresenta como operacionalizar as análises descritivas e a função de Kaplan-Meier. O *segundo capítulo* apresenta a regressão de *Riscos Proporcionais de Cox*. Suge os determinantes para a modelagem, a operacionalização de seus dados e, finalmente, implementa o modelo sugerido. O *terceiro capítulo* apresenta os pressupostos metodológicos do modelo de Cox e verifica cada um deles através de técnicas estatísticas específicas. Então, propomos ajustes ao modelo inicialmente apresentado.

³ O código e os dados se encontram disponíveis no projeto Open Science Framework do autor, <https://osf.io/kfy24/>, e podem ser reproduzidos isoladamente para cada capítulo.

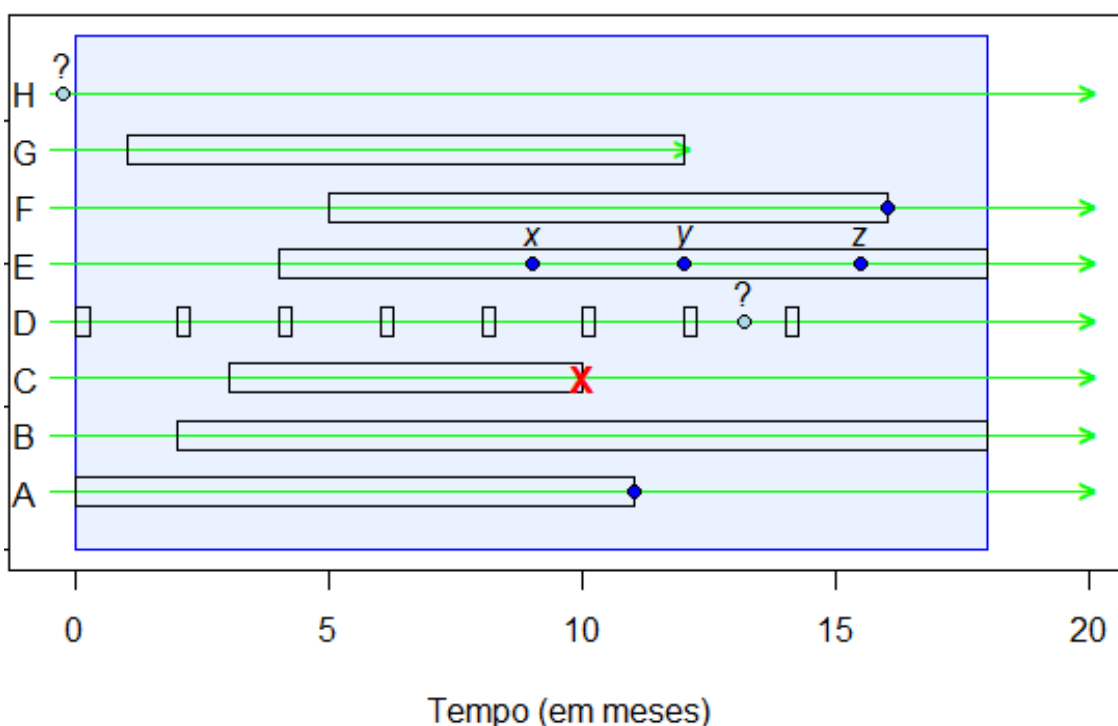
2 COMPREENDENDO OS DADOS DE SOBREVIVÊNCIA

Neste primeiro capítulo, introduziremos conceitos e ferramentas básicas de análise de sobrevivência a partir do caso prático de adoção dos *Planos Diretores* pelos municípios brasileiros. Abordaremos os dados e sua estrutura, a variável dependente na análise de sobrevivência, a tabela de vida e a função kaplan–Meier.

2.1. Conceitos de análise de sobrevivência.

Ao lidarmos com análises de sobrevivência, é importante ter em mente alguns conceitos básicos que dizem respeito aos cenários e aos dados de pesquisa. Nessa seção, trazemos os conceitos ilustrados pela *figura 1*, representando diferentes cenários de pesquisa e situações que podem surgir durante o acompanhamento das unidades de análise. Ainda que alguns dos cenários apresentados não sejam comuns em nosso tipo de pesquisa especificamente, compreendê-los é importante para certas análises de sobrevivência.

Figura 1. Representação de Cenários de pesquisa.



FONTE: O autor (2025)

Na *figura 1* as linhas horizontais (verde) representam as **unidades analisadas** (“indivíduos” da análise) e o eixo horizontal (eixo x) indica a passagem do tempo. Os círculos azuis indicam os **eventos de interesse** do estudo, isto é, o

momento em que ocorre o desfecho que está sendo analisado (como a adoção de uma política pública ou o diagnóstico de uma doença). O “X” vermelho representa uma **falha no acompanhamento**, sinalizando que houve algum tipo de perda de informação sobre aquele indivíduo antes que o evento pudesse ser observado.

O grande retângulo azul marca o recorte temporal da pesquisa, o intervalo para o qual os dados foram coletados. Ele se distingue dos retângulos pretos, que representam o período efetivo de observação de cada unidade (período de **acompanhamento** ou **follow up time**). Idealmente, o período de observação coincide com o período em que a unidade esteve sob risco efetivo de sofrer o evento de interesse. Percebe-se que os períodos de observação não necessariamente coincidem nem entre as unidades nem com a extensão do recorte da pesquisa, característica que dá flexibilidade às análises de sobrevivência.

O indivíduo A representa o caso ideal para o modelo, aquele em que o evento de interesse ocorre dentro do período da pesquisa, sendo um **evento observado**. O indivíduo B exemplifica a **censura à direita**⁴ mais comum, quando o evento não acontece até o fim do acompanhamento. O indivíduo B, assim como os indivíduos C e G, estão todos *censurados à direita*. No caso de C, a censura à direita se dá por **perda de acompanhamento** ou **descontinuidade** na coleta de dados. No indivíduo G ela ocorre por uma impossibilidade completa de continuidade: no caso de um município, ele poderia ter se fundido a outro, por exemplo. Apesar de representarem descontinuidades, tanto C quanto G contribuem com a análise, pois indicam que, durante certo período, o evento não ocorreu — o que é uma informação valiosa. Distinguir essas formas de censura pode contribuir, em certos casos, para compreender as limitações da pesquisa.

Outro conceito importante é a **censura intervalar** (indivíduo D). Aqui, o evento acontece entre dois pontos de observação, mas não se sabe exatamente quando (?). Isso pode ocorrer, por exemplo, quando os dados são registrados por semana ou por mês, e não em dias exatos. Esse tipo de censura pode representar uma perda de informação ou não, pois dependerá da granularidade da unidade de tempo requerida pela pesquisa. Assim, a **censura intervalar** não se confunde com o registro do tempo em uma unidade de medida “agregada”, como em *meses*.

⁴ Os tipos de censura são conceitos da análise de sobrevivência e estão descritos nos manuais teóricos sobre o assunto (BOX-STEFFENSMEIER e outros, 2004).

O indivíduo E representa pesquisas mais complexas, que tratam da ocorrência de **múltiplos eventos** (x, y, z). Esses eventos podem ser **estágios sucessivos** (como diferentes etapas da adoção da política), eventos concorrentes, quando **desfechos múltiplos** estão em jogo (ou adota-se a política x ou a y) ou **eventos recorrentes** (o mesmo desfecho se repete). O indivíduo F é similar ao A, no sentido de que representam eventos observados pela pesquisa. Importante notar que ainda que tenham sido acompanhados por recortes temporais distintos, ambos os eventos de interesse ocorreram no mesmo intervalo de observação (no caso, aos 11 meses), sendo equivalentes para fins da análise estatística.

O indivíduo H ilustra a **censura à esquerda**, que ocorre quando se sabe que o evento aconteceu antes do início do estudo, mas, como em D, não se conhece o momento exato (?). Embora esse tipo de censura não possa ser tratado diretamente pelos modelos utilizados nesse trabalho, ela pode ser útil para outras abordagens, ajudando a reduzir possíveis vieses de análise.

Por fim, é relevante mencionar o conceito de **truncamento**. Ele se refere à inclusão na base apenas de indivíduos que satisfazem determinadas condições — por exemplo, aqueles que já passaram por certo marco. O truncamento, em sentido estrito, difere da filtragem dos dados. Ele implica que os indivíduos que não atendem aos critérios nem chegam a entrar na análise, o que pode gerar **viés de seleção** se não for cuidadosamente considerado.

2.2. Planos Diretores Municipais

Uma das responsabilidades conferidas pela Constituição Federal à gestão municipal consiste em ordenar o território municipal através de políticas de planejamento e desenvolvimento da cidade. O *Plano Diretor* é o instrumento técnico para essa finalidade. O estatuto das cidades, Lei 10.257/01, determina que as Câmaras municipais devem institucionalizar os planos diretores das cidades que se enquadrem nas características descritas no artigo 41. Destaquemos as duas características mais relevantes: (i) as cidades com mais de vinte mil habitantes; (ii) cidades integrantes de regiões metropolitanas e aglomerações urbanas.

Para investigarmos empiricamente a adoção dos planos diretores municipais encontramos dados relevantes na pesquisa anual dos Municípios Brasileiros feita pelo IBGE para o ano de 2021 (IBGE, 2021). Os dados compartilhados pelo IBGE contém o índice ou dicionário de variáveis constantes naquela edição da pesquisa.

No caso em análise, os dados do nosso interesse estão codificados como segue. Primeiramente temos a *chave primária* dos dados, que consiste na variável que garante a unicidade das observações, no caso, o *código identificador dos municípios*. Junto a ela, o IBGE compartilha outras *variáveis externas*, que dizem respeito a informações básicas sobre os municípios, como sua população e região.

Na sequência identificamos dados de interesse assim como disponibilizados na base da pesquisa dos municípios. Entre eles: a existência ou não de planos diretores, a realização ou não de uma revisão, os anos em que eles se deram.

Quadro 1. Dados de Interesse da Pesquisa anual de Municípios - IBGE.

Variáveis Externas		Legislação e Instrumento de Planejamento	
Código do Município	CodMun	Plano Diretor - Existência	MLEG01
Unidade da Federação	UF	Ano de Criação da Lei	MLEG011
Código da UF	CodUF	Plano Revisto	MLEG012
Nome do Município	Mun	Ano da última revisão	MLEG013
População Estimada 2022	Pop	Plano diretor em elaboração	MLEG014
Faixa de População	Faixa_pop	-	-
Grande Região	Regiao	-	-

FONTE: IBGE (2022)

Identificados os dados de interesse, o primeiro passo em uma análise consiste em carregá-los e selecioná-los. Para isso, utilizamos a função *read.csv()* e as ferramentas do pacote *tidyverse*.

```
if(require(tidyverse) == F) install.packages("tidyverse"); require(tidyverse)
if(require(openxlsx) == F) install.packages("openxlsx"); require(openxlsx)
if(require(survival) == F) install.packages("survival"); require(survival)
if(require(survminer) == F) install.packages("survminer"); require(survminer)

# Carrega os dados no R
caminho_arquivo <- "C:/direcione/para/Base_MUNIC_2021.xlsx"
dados <- read.xlsx(caminho_arquivo, startRow = 1, sheet = 4)
dados <- dados %>%
  select(CodMun, UF, Regiao, Pop, Faixa_pop, Mleg01, Mleg011, Mleg012,
         Mleg013, Mleg014)
```

O passo seguinte é compreender a estrutura dos dados e tratá-los. Esses dois movimentos se dão simultaneamente em ciclos de compreensão (de cada uma das variáveis) e ajustes (transformações de classe, limpeza, etc.). No que se refere ao tratamento dos dados, os termos *wrangling* e *tidying* são geralmente usados para referir-se aos ajustes de limpeza, organização e adequação dos dados aos formatos necessários às análises (WICKHAM, 2014). Estes procedimentos podem exigir quantidades de esforço e conhecimento que não estão no escopo deste trabalho. Nossa abordagem, no entanto, lida com dados pré processados pelo IBGE, o que, no geral, garante que a fase de *wrangling* não seja tão complexa.

```
# Para entender os dados
glimpse(dados)
summary(dados)
table(dados$Mleg01)

# Caracteres viram categorias ou fatores
dados$Faixa_pop <- as.factor(dados$Faixa_pop)
dados$UF <- as.factor(dados$UF)
dados$Regiao <- as.factor(dados$Regiao)
```

No nosso caso, fazemos transformações para padronizar os dados ausentes com o valor apropriado *NA*. Consideramos suficiente tratar dados de *recusa a responder* como dados ausentes⁵. Além disso, transformamos dados dos anos para o formato lógico de data. Os *sims* e *nãos* são transformados em valores lógicos *TRUE* e *FALSE*. Por fim, tendo em vista que a pesquisa registra adoções até o ano de 2021, acrescentamos o valor 2022 para todos os municípios que informaram não terem adotado a política até o final do período. São os dados *censurados à direita*.

```
dados <- dados %>%
  mutate(Mleg01 = case_when(
    Mleg01 == "Sim" ~ TRUE,
    Mleg01 == "Não" ~ FALSE,
    TRUE ~ NA),

  Mleg012 = case_when(
    Mleg012 == "Sim" ~ TRUE,
    Mleg012 == "Não" ~ FALSE,
    TRUE ~ NA),
```

⁵ Decisões de pesquisa devem ser preferencialmente justificadas.

```

Mleg014 = case_when(
  Mleg014 == "Sim" ~ TRUE,
  Mleg014 == "Não" ~ FALSE,
  TRUE ~ NA))

dados$Mleg011 <- gsub("-", 2022, dados$Mleg011)
dados$Mleg011 <- gsub("Não informou", NA, dados$Mleg011)
dados$Mleg011 <- gsub("Recusa", NA, dados$Mleg011)
dados$Mleg013 <- gsub("-", 2022, dados$Mleg013)
dados$Mleg013 <- gsub("Não informou", NA, dados$Mleg013)
dados$Mleg013 <- gsub("Recusa", NA, dados$Mleg013)

```

A função *glimpse()* nos permite ter uma visão simplificada de como as variáveis estão dispostas: a classe dos dados e seus primeiros valores.

```

> glimpse(dados)
Rows: 5,570
Columns: 10
$ CodMun      <int> 1100015, 1100023, 1100031, 1100049, 1100056, 1100064,
$ UF          <fct> RO, RO, RO, RO, RO, RO, RO, RO, RO, RO, RO, RO, RO,
$ Regiao      <fct> 1 - Norte, 1 - Norte, 1 - Norte, 1 - Norte, 1 - Norte,
$ Pop         <int> 22516, 111148, 5067, 86416, 16088, 15213, 7052, 19255,
$ Faixa_pop   <fct> 4 - 20001 até 50000, 6 - 100001 até 500000, 2 - 5001
$ Mleg01      <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, FALSE, TRUE
$ Mleg011     <chr> "2006", "2019", "2022", "2006", "2014", "2006", "2016"
$ Mleg012     <lgl> TRUE, FALSE, NA, TRUE, FALSE, TRUE, FALSE, NA, TRUE,
$ Mleg013     <chr> "2010", "2022", "2022", "2021", "2022", "2017",
$ Mleg014     <lgl> NA, NA, TRUE, NA, NA, NA, NA, FALSE, NA, TRUE,

```

A função *summary()* nos dá uma visão geral mais específica sobre o estado de cada uma das variáveis. A seguir, vemos, a título de exemplo, como os municípios estão distribuídos entre adotantes ou não e dados indisponíveis (\$Mleg01) ; e dados de dispersão de suas populações (\$Pop).

```

> summary(dados$Mleg01)
  Mode   FALSE   TRUE   NA's
logical  2602  2960     8

> summary(dados$Pop)
  Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
  771     5454    11732    38298   25765   12396372

```


2.3. Estrutura dos dados.

A *análise de sobrevivência* é um conjunto de técnicas estatísticas utilizadas para se analisar dados de tempo decorrido até que determinado evento de interesse ocorra nos sujeitos observados (GUJARATI, p. 577, 2011) (BOX-STEFFENSMEIER e outros, p.2, 2004). No estudo da difusão de políticas públicas, ela pode ser usada para responder perguntas como: Quais fatores determinam a maior probabilidade de adoção de determinada política pública por municípios? Qual a probabilidade de que a adoção de uma política venha a ocorrer em certo município ou grupo de municípios? Qual a proporção de municípios terão adotado a política até determinado momento? Qual a probabilidade de que a adoção ocorra em determinado grupo de indivíduos em até um determinado momento?

Para que possamos entender melhor esse tipo de análise e tomar decisões acertadas ao modelá-la, é importante ter uma compreensão básica de seus fundamentos, assim como de alguns aspectos técnicos do procedimento de implementação.

Um importante aspecto consiste em avaliar a estrutura dos dados. Uma forma bastante comum de organização de dados é a forma *transversal* ou *cross-section*, em que, para cada unidade observada, temos não mais de uma linha. No seu formato transversal padrão, todas as observações (linhas do banco) referem-se a um único momento, o momento da análise, definido pelo pesquisador. Já ao analisar dados de sobrevivência, estamos interessados na sua dimensão temporal, e portanto, as observações registram diferentes momentos no tempo. A estrutura dos dados pode apresentar duas formas.

A primeira e mais simples delas é semelhante à estrutura transversal básica: cada observação retrata uma única unidade de análise. Diferencia-se apenas por conter ao menos um dado com dimensão temporal, retratando o evento de interesse. Este é o caso dos dados dispostos na MUNIC, como podemos observar na nossa base. Nela coincidem o número de municípios brasileiros (5.570), a quantidade de linhas e a quantidade de códigos únicos dos municípios.

```
> length(unique(dados$CodMun))  
[1] 5570  
> length(unique(dados$CodMun)) == dim(dados)[1]  
[1] TRUE
```

Este tipo de estrutura de dados é suficiente para análises mais simples. Ela pressupõe que lidaremos com variáveis relativamente estáveis (ou pelo menos, estáveis entre si) ao longo do tempo (ou pelo menos, ao longo do tempo observado) e o único fenômeno temporal relevante é a ocorrência do evento de interesse, no caso, a adoção ou não do Plano Diretor. Por sua vez, para observar o efeito de variáveis independentes menos estáveis, precisaremos de uma estrutura que capte a relação tempo-dependência dos dados. Para isso, usa-se um formato capaz de refletir os diferentes momentos das unidades de análise ao longo do tempo. Assim, a estrutura *cross-section-time-series* retrata cada unidade de análise repetidamente, em momentos distintos (observações). Ver quadro 2. Por ora, é suficiente observar que nossos dados estão dispostos de forma transversal convencional.

Quadro 2. Dados fictícios dispostos em *Cross-Section* e *Cross-Section Time-Series*.

Dados Cross-Section (ano 2020)		Dados Cross-Section time series		
Município	Partido Incumbente	Ano	Município	Partido Incumbente
Aracajú	PDT	2019	Belém	PSDB
Belém	PSDB	2020	Belém	PSDB
Cuiabá	MDB	2021	Belém	PSOL
Fortaleza	PDT	2019	Manaus	PSDB
Manaus	PSDB	2020	Manaus	PSDB
Recife	PSB	2021	Manaus	Avante

FONTE: O autor (2025)

2.4. Variável Dependente.

Outro ponto a ser observado é a estrutura da nossa variável dependente. Conceitualmente, a variável dependente com que lidamos em análises de sobrevivência é o momento do evento de interesse, ou seja, da *ocorrência das adoções*. Essa simples informação, no entanto, pode ser registrada sob diferentes formatos, que precisam ser levados em consideração.

Uma característica da variável dependente das análises de sobrevida é possuir uma *dimensão temporal*, refletindo o período até a ocorrência de um evento de interesse e uma *dimensão do evento de interesse*, refletindo sua ocorrência e/ou

características. Por possuir duas dimensões, essa informação geralmente é retratada por mais de uma variável em uma base de dados. Vejamos uma por vez:

A variável referente à *dimensão do evento*, onde se retrata o estado do evento, pode refletir (i) sua ocorrência / não ocorrência (nos formatos: T/F , 0/1, 1/2); ou ainda (ii) diferentes estágios ou tipos possíveis de desfecho da ocorrência (geralmente dados numéricos ou categóricos para cada desfecho: 0,1,2,3)⁶. Essa dimensão não informa o momento em que o evento ocorreu e por isso requer um dado complementar, a dimensão temporal.

A *dimensão temporal* pode vir representada nos seguintes formatos: (i) de “momentos específicos”, datas, anos (instantes mais precisos, como horas e minutos, tendem a não ser de interesse nas nossas aplicações) ; (ii) de “intervalos” ou “período transcorrido”, como: x dias, y meses, z anos, contados desde um momento previamente definido até o evento de interesse; (iii) de período transcorrido, mas definido por “*momentos de início e de fim da observação*”. Neste último caso, a dimensão temporal é retratada não em uma, mas duas variáveis⁷.

Dando continuidade à investigação dos nossos dados, a variável dependente está disposta em duas colunas (*features*): uma de classe lógica “\$Mleg01” (TRUE, para ocorrências da adoção; FALSE, para não ocorrências da adoção) e uma numérica “\$Mleg011” (os anos de aprovação do plano diretor municipal).

```
> summary(dados$Mleg01)
  Mode   FALSE    TRUE   NA's
logical  2602   2960
```

A variável da *dimensão temporal*, por sua vez, contém valores numéricos em anos. Ela pode ser visualizada com um gráfico de barras representando a quantidade de observações ano a ano. Para construí-lo, primeiro transformamos os dados em uma *tabela de contingência*, agregando a quantidade anual de registros.

```
# gera um data.frame com as ocorrências anuais de leis
dadosM <- dados %>%
  group_by(Mleg011) %>%
  dplyr::summarise(N = n()) %>%
```

⁶ Este formato é utilizado em análises mais complexas e não será trabalhado neste trabalho. Ver BOX-STEFFENSMEIER e outros (P. 155, 2004).

⁷ O Pacote Survival aceita diferentes formas para a variável dependente. Para detalhes, ver Therneau (2023).

```

ungroup()
# removemos os dados omitidos e os dados de follow-up time
dadosM <- na.omit(dadosM)
dadosM <- dadosM[dadosM$Mleg011 != '2022', ]

```

Em seguida, podemos gerar o gráfico de barras.

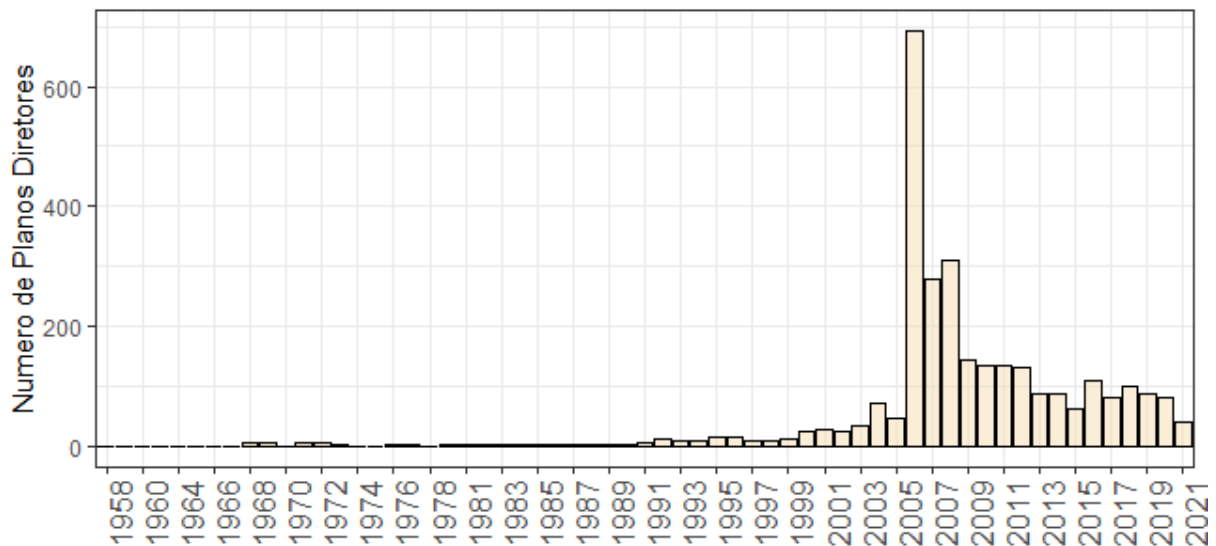
```

ggplot(data=dadosM, aes(x=Mleg011, y=N)) +
  geom_bar(stat = "identity", fill= "wheat", color= "black", alpha=0.5)+
  labs(title="Leis Municipais instituidoras de Planos diretores por ano",
       x = " ", y = "Número de Planos Diretores") + theme_bw() +
  theme(axis.text.x = element_text(size = 13, angle=90, hjust=1))

```

Neste ponto, é possível observar que a promulgação de *Planos Diretores* pelos municípios brasileiros é uma realidade de muitas décadas, porém, em números mínimos. Após a redemocratização, identificamos um pequeno crescimento e, finalmente, um impulso muito mais significativo a partir de meados da década de 2000. Este dado é condizente com o fato já discutido de que o estatuto das cidades é lei editada no ano de 2001.

Gráfico 1. Número de leis municipais instituidoras de planos diretores por ano.



FONTE: O autor (2025)

2.5. Tabela de Vida

A tabela de vida é a ferramenta mais básica de análise de sobrevivência. Com ela, podemos compreender como a variável dependente se comporta ao longo

do tempo no modelo. Para rodá-la utilizaremos 2 pacotes: *survival* e *survminer*. Usamos as funções *Surv()* para construir a nossa variável dependente (“*censura.anos*”). Então, rodamos um modelo simples utilizando a função *survfit()* para modelá-la “sem covariáveis” (~ 1).

```
# criando a variável dependente
dados$Mleg011_ <- as.numeric(dados$Mleg011)-1950
censura.anos_ <- Surv(time=as.numeric(dados$Mleg011_),
event=dados$Mleg01)

# tabela de vida
summary(survfit(censura.anos_ ~ 1), # tabela de vida.
times = c(10,20,30,40,50,60,70)) # seleciona linhas de interesse
```

Para facilitar a investigação da nossa variável, selecionamos o ano de 1950 como o marco inicial da contagem do passar do tempo em anos. Além disso, selecionamos os momentos da tabela de vida de maior interesse. Temos:

```
8 observations deleted due to missingness
time n.risk n.event survival
10 5560 3 0.999
20 5543 17 0.996
30 5519 23 0.992
40 5496 26 0.988
50 5396 123 0.965
60 3735 1770 0.647
70 2724 957 0.475
```

No *output*, é possível observar, para cada momento do estudo (*time*), quantas adoções ocorreram em cada intervalo representado pela linha da tábua de vida (*n. event*). Observar quantas unidades de análise ainda podem adotar (*n. risk*), deste valor estão de fora as unidades que já adotaram, os dados incompletos e as possíveis censuras. Também lemos a probabilidade acumulada de sobrevivência (de não-adoção) até aquele momento (*survival*)⁸.

⁸ O leitor pode encontrar disponíveis no projeto Open Science Framework do autor, <https://osf.io/kfy24/>, uma lista detalhada com 45 pacotes que citam na descrição “survival analysis”.

2.6. Estimador não-paramétrico Kaplan-Meier

Até aqui, discutimos a *estrutura dos dados* e a *variável dependente*. Outro aspecto fundamental para avançar na compreensão da análise de sobrevivência diz respeito ao caráter *não-paramétrico*, *semi-paramétrico* ou *paramétrico* dos modelos (CARVALHO e outros, 2011). Essa distinção refere-se, basicamente, à quantidade de suposições que podemos fazer sobre os nossos dados para que possamos operá-los estatisticamente.

As *modelagens paramétricas* têm maior precisão nas estimativas geradas. Esses modelos pressupõem fortes suposições teóricas sobre a forma da distribuição do dado utilizado, ou pelo menos sobre certas características (BOX-STEFFENSMEIER e outros, p. 21, 2004). São modelos mais adequados quando o objetivo é prever com precisão o tempo até a ocorrência da adoção. Naturalmente, certas perguntas somente poderão ser adequadamente respondidas com modelos deste tipo.

Por outro lado, o uso de modelos *semi e não-paramétricos* é mais factível para o tipo de pergunta que geralmente fazemos e para as características dos dados a que temos acesso nas análises básicas de difusão de políticas públicas. Primeiramente, os modelos semi e não-paramétricos são suficientes para nos responder sobre o efeito relativo das covariáveis no tempo até o evento. Eles também são mais flexíveis quanto ao tamanho da amostra, além de não demandarem fortes suposições sobre a forma da distribuição dos dados.

Uma das modelagens não-paramétricas mais simples e intuitivas para a análise de difusão das políticas nos municípios é o *estimador Kaplan-Meier*. O estimador representa a própria função de sobrevivência e, em uma leitura menos técnica, representa o vetor de probabilidades de o evento não ter ocorrido até certo momento entre as unidades. Assim, além de não fazer suposições sobre os dados, sua visualização é bastante intuitiva (CARVALHO e outros, p.100, 2011)

Na chamada curva *Kaplan-Meier*, o eixo *y* representa a proporção de municípios que não adotaram a política até o momento (eixo *x*). O gráfico, portanto, traça a sequência de multiplicação das probabilidades acumuladas de sobrevivência (não adoção). Para implementar curvas *Kaplan-Meier*, utilizamos os mesmos pacotes: *survival* e *survminer*. Construímos a variável dependente *surv()*, para todo o período dos dados, e montamos o modelo *survfit()*.

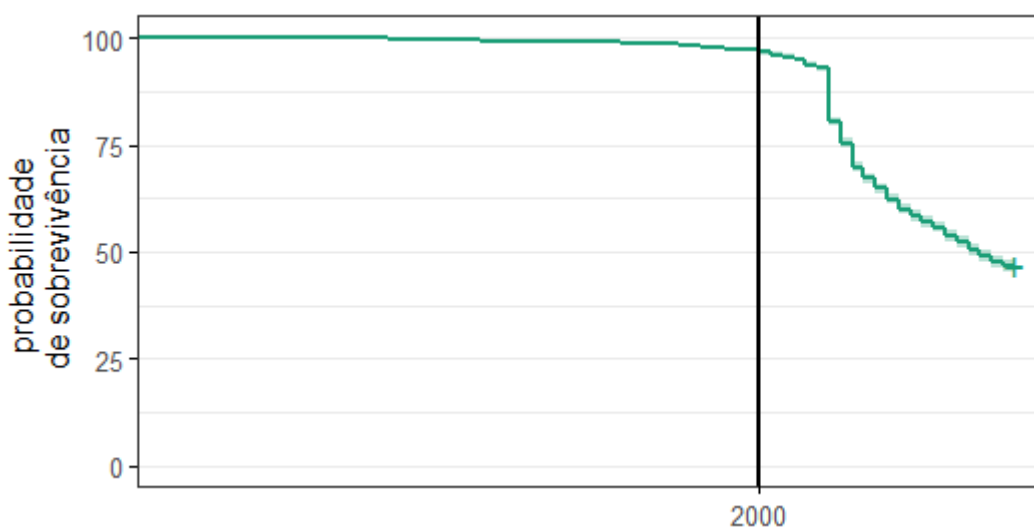
```
# fit kaplan-meier
censura.anos <- Surv(time=as.numeric(dados$Mleg011), event=dados$Mleg01)
fit_KM <- survfit(censura.anos ~ 1, ctype=2, stype=2, conf.type="log")

# gráfico: curva kaplan-meier
KM_plot <- ggsurvplot(fit_KM, data = dados, fun = "pct",
  xlim = c(1950, 2021), palette = "Dark2", pval = T,
  title = "Função Sobrevivência, Kaplan-Meier",
  font.title = c(20, "bold.italic"),
  xlab = " ", legend = "none",
  ylab = "probabilidade de sobrevivência",
  ggtheme = theme_bw(base_size = 13))

KM_plot$plot + geom_vline(xintercept = 2000,
  linetype = "solid", color = "black", size = 1)
```

No nosso caso, identificamos que a probabilidade dos municípios não criarem seus planos diretores caiu drasticamente após a adoção do *Estatuto das Cidades* em 2021.

Gráfico 2. Função de Sobrevivência (não adoção) - Kaplan-Meier.



FONTE: O autor (2025)

Outra função do estimador Kaplan-Meier é permitir comparar a função de sobrevivência entre grupos. Para isso, é preciso definir o modelo com uma *variável categórica*⁹ que funcionará como variável independente única. Com esse objetivo,

⁹ Para a melhor observação é desejável que a variável seja categorizada em poucos níveis.

vamos criar mais uma variável em nossa base de dados. Ela agrupa os municípios por população e será definida a partir dos dados que já possuímos.

Como já mencionado, o Estatuto das Cidades determinou que municípios com mais de 20 mil habitantes instituísem seus planos diretores próprios. Por essa razão, definiremos uma nova variável categórica (*\$populacao*) em grupos : (i) *de até 10 mil habitantes*, (ii) *entre 10 e 20 mil*, (iii) *entre 20 e 30 mil* e (iv) *com mais de 30 mil habitantes*. Além disso, limitaremos o gráfico ao intervalo de tempo em que o número de adoções ganhou corpo (às vésperas da redemocratização).

```
dados <- dados %>%
  mutate(populacao = case_when(
    Pop < 10000 ~ "até 10k hab.",
    Pop >= 10000 & Pop < 20000 ~ "10k-20k hab.",
    Pop >= 20000 & Pop < 30000 ~ "20k-30k hab.",
    Pop >= 30000 ~ "mais de 30k hab." ))

dados$populacao <- factor(dados$populacao,
  levels = c("até 10k hab.",
             "10k-20k hab.",
             "20k-30k hab.",
             "mais de 30k hab."))

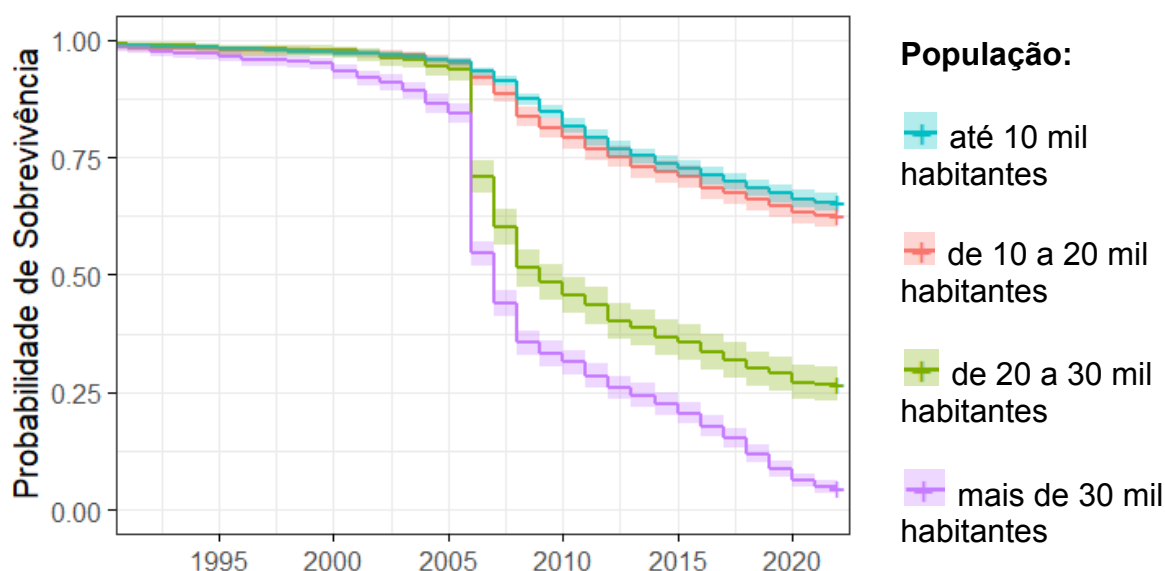
# criando o modelo estatístico
censura.anos <- Surv(time=as.numeric(dados$Mleg011), event=dados$Mleg01)
fit_KM_p <- survfit(censura.anos ~ dados$populacao,
  ctype=1, stype=2, conf.type="log-log")

# gráfico: curva kaplan-meier
ggsurvplot(fit_KM_p, data = dados,
  xlim = c(1992, 2021), # reduzindo o eixo x
  title = "Função sobrevivência, Kaplan-Meier",
  break.time.by = 5,
  font.title = c(20, "bold.italic"),
  xlab = "Ano", pval = TRUE,
  conf.int = TRUE,
  ylab = "Probabilidade de Sobrevivência",
  ggtheme = theme_bw(base_size = 10))
```

Com as curvas categorizadas por faixas de população, podemos observar que o Estatuto das Cidades teve grande impacto na probabilidade de adoção de Planos Diretores nos municípios acima de 20 mil habitantes, sobretudo naqueles

acima de 30 mil. Observamos também que, para além do efeito imediato da aplicação da lei, ela também consolidou uma taxa estável e constante de novas ações em todas as faixas ao longo da década e meia seguinte.

Gráfico 3. Função de Sobrevivência (não adoção) estratificada - Kaplan-Meier.



FONTE: O autor (2025)

É possível ainda comparar formalmente a diferença de sobrevivência entre os grupos da variável categórica. O teste estatístico mais comum para isso é o teste Log-Rank. (UCLA, s.d.)

```
logrank <- survdiff(censura.anos ~ dados$populacao, data = dados)
logrank
```

O output com p-valor abaixo de 0,05 reflete a evidência de que existe diferença significativa na sobrevivência entre pelo menos dois grupos¹⁰.

```
Chisq= 2473 on 3 degrees of freedom, p= <2e-16
```

Em síntese, esse tipo de análise tem um apelo gráfico bastante significativo uma vez que é possível literalmente visualizar quais grupos exibem maior ou menor chance de adotar a política. Para garantir maior confiabilidade, o ideal é que as análises gráficas sejam acompanhadas de testes estatísticos específicos para

¹⁰ Não foi possível reproduzir o teste que compara os grupos dois a dois. Provável depreciação ou erro no pacote.

avaliar em que medida a diferença entre os grupos na amostra podem ou não ser generalizadas para a população.

2.7. Resultados

Concluimos assim este primeiro capítulo tendo explorado noções básicas da análise de sobrevivência aplicada: conceitos, estruturas de dados, variável dependente e os conceitos de modelagens *paramétricas*, *semi* e *não-paramétricas*. Também aprendemos a aplicar alguns procedimentos para obter análises exploratórias dos dados, suas características, dispersão temporal e curvas Kaplan-Meier.

Com essas ferramentas, pudemos observar empiricamente o que sabia-se instituído em Lei. Até aqui, podemos avaliar que este é um caso de regulamentação bem sucedida: os maiores municípios comportaram-se como determinado pelos legisladores (linhas lilás e verde). Além disso, identificamos também um efeito consolidador de iniciativas de planejamento urbano em municípios que não estavam obrigados pela lei (linhas azul e vermelha). Por fim, podemos identificar que entre as faixas de 20 mil e 30 mil habitantes ainda há uma considerável quantidade de Municípios que não instituíram planos diretores apesar de vinculados por lei para fazê-lo (linha verde).

Observar o comportamento das variáveis dependentes e independentes é o primeiro passo na compreensão do fenômeno. Como proceder se estivermos interessados em identificar o impacto das variáveis na adoção e mesmo na velocidade em que ela pode acontecer ? Para isso, precisaremos de modelos estatísticos mais sofisticados, que abordaremos no capítulo seguinte.

3 EVENT HISTORY ANALYSIS COMO ANÁLISE DE SOBREVIVÊNCIA

3.1. Introdução

No capítulo anterior, abordamos noções básicas de como implementar uma análise de sobrevivência básica aplicada à adoção de políticas públicas, particularmente, a adoção dos Planos Diretores municipais. A análise foi basicamente exploratória dos dados disponíveis. Vimos na prática o que se espera de uma base de dados típica. Vimos também como observar nossos dados em sua dimensão temporal e como ela se expressa através da tabela de vida e do estimador Kaplan-Meier. Aprendemos ainda a comparar grupos.

Identificamos que municípios adotam planos diretores há décadas, mas que desde a redemocratização os planos diretores se multiplicaram. Constatamos ainda que houve um crescimento repentino no número de planos na segunda metade da década de 2000, o que era de se esperar, dado que a lei 10.257/01 instou os municípios a fazê-lo. A partir de então um novo padrão mais consolidado de adoções se estabeleceu.

Neste capítulo, primeiro levantamos as bases teóricas que fundamentam a formulação de hipóteses sobre a difusão de políticas públicas. Em seguida, apresentamos o modelo de Cox e utilizamos o modelo para testar as hipóteses levantadas inicialmente. As hipóteses podem ser encontradas no quadro 3.

3.2. Teoria e determinantes da difusão de políticas

Embora grande parte da ação governamental ocorra de forma incremental, a inovação política é um fato. O fenômeno da inovação não deve ser confundido ou resumido à *invenção*. Não se trata de criar algo novo, mas em reproduzir com adaptações soluções existentes em outro lugar (SILVA; KAUCHAKJE, 2018, p. 30). Por essa razão, os estudos sobre a difusão de políticas públicas estão atrelados ao estudo da inovação política.

Por muito tempo, os estudos empíricos sobre inovação e difusão de políticas concentraram-se nos estados norte-americanos, mas, gradualmente, passaram também a abordar a difusão em âmbito internacional (BERRY, 2007). Diversos modelos foram propostos. Berry (2007) lista alguns deles: políticas públicas podem ser copiadas entre unidades de análise semelhantes; ou copiadas de líderes em

inovação; difundidas regionalmente, causadas por razões internas, estimuladas pela competição política, entre outros.

No Brasil, o estudo de difusão de políticas foi introduzido como forma de complementar os estudos sobre descentralização política. Para Coelho (2009; 2012) a abordagem sobre difusão no Brasil surge para acrescentar perspectivas à consolidada pesquisa sobre descentralização de políticas públicas.

Desde então, um apanhado das pesquisas empíricas quantitativas em difusão de políticas públicas no Brasil se caracteriza pela identificação de determinantes de diversas naturezas. Algumas hipóteses levantadas tratam de:

- Incentivos eleitorais, como o efeito da competição eleitoral (BATISTA, 2018; SUGYIAMA, 2008; COELHO, 2012), mas também o efeito da reeleição (BASTOS, 2015) e do ano eleitoral (COELHO, CAVALCANTE, TURGEON, 2016).
- Investigam também a interação política entre executivo e legislativo em nível municipal (SPADA, 2010; WAMPLER, 2008).
- Ideologia e alinhamento político. A exemplo da ideologia/partido e o alinhamento partidário entre esferas de governo (SPADA, 2010; COELHO, CAVALCANTE, TURGEON, 2016).
- Redes sociais ou padrões de boa gestão. Redes partidárias e proximidade geográfica (COELHO, CAVALCANTE, TURGEON, 2016). O efeito regional de redes (BASTOS, 2015)
- Determinantes internos. Vários são levados em consideração como população, seus segmentos, níveis de pobreza, produção interna per capita, transferências governamentais, entre outros, relevantes para a política em questão.
- Padrões regionais, de interação, influência e proximidade são considerados sobre diferentes linhas argumentativas, como a interação entre gestores, interação de eleitores ou a semelhança e percepção dos problemas enfrentados.

O presente trabalho não dará conta de explorar todas as teorias existentes sobre o assunto e como se adequam ao federalismo brasileiro. Dados os propósitos da pesquisa, selecionamos algumas variáveis comumente utilizadas pela literatura

para dar continuidade ao trabalho, emulando os trabalhos sobre difusão de políticas públicas no federalismo brasileiro.

Assim sendo, as variáveis escolhidas devem refletir hipóteses comumente testadas pela ciência política nacional e se adequarem à investigação da política pública investigada, a adoção dos planos diretores municipais. Seguem os determinantes selecionados e as hipóteses formuladas a partir deles:

Primeiro. Certas políticas públicas são profundamente influenciadas por determinantes internos. De acordo com Berry e Berry (2007, p. 231) esses determinantes podem ser características políticas, econômicas e sociais da unidade analisada. Preferimos caracterizá-lo como a própria agenda setorial. Ou seja, quando a demanda setorial é severa, a temática política se sobrepõem às demais variáveis políticas. Assim, municípios mais pobres respondem rapidamente à demanda por políticas sociais. “Onde há demandas para maior oferta de assistência social, cobertura de saúde ou educação, provavelmente decisões de governo por reformas serão tomadas” (COELHO, CAVALCANTE E TURGEON, p.149, 2016). No caso dos planos diretores, temos que cidades maiores são mais complexas e, portanto, demandam maior planejamento urbano. A *população municipal*¹¹ poderia ser considerada uma proxy da necessidade de planejamento e , portanto, da urgência da agenda temática/setorial.

Segundo. Outro aspecto comumente discutido pela literatura diz respeito aos incentivos do sistema político-eleitoral. Para uma das diversas formulações possíveis, sistemas políticos mais competitivos levam seus atores a recorrerem a ações “inovadoras” e que os diferenciem dos demais competidores. Segundo esse argumento, a inovação política é um ativo que pode ser eleitoralmente utilizada. (SUGIYAMA, 2008; COELHO, 2012; SUGIYAMA, 2012; COELHO, CAVALCANTE e TURGEON, 2016; BATISTA, 2017). Assim, selecionamos a *competição eleitoral* para expressar os incentivos do sistema político-eleitoral na adoção de políticas públicas. O indicador é implementado pela fração de votos válidos do eleitorado de um município, que representa a diferença eleitoral entre os dois candidatos mais bem posicionados em cada eleição¹². A variável tem relação negativa com o risco de adoção da política pública. Quanto menor o índice, menor a diferença eleitoral entre

¹¹ Censo IBGE 2022.

¹² O indicador utilizado foi criado pelo autor a partir de dados do TSE para as 6 eleições, entre 2000 e 2020. Foi então extraída a média das frações.

os candidatos, portanto, maior a competitividade e maior a pressão pela adoção da “inovação política”.

Terceiro. Outro tema caro à análise da descentralização de políticas públicas consiste na constatação de que boa parte dos municípios sofre com importantes restrições administrativas, financeiras e burocráticas. A desigualdade entre as realidades e capacidades administrativas dos municípios é elemento central no dilema da coordenação federativa de políticas públicas. Municípios precisam lidar com diferentes níveis de limitação de recursos financeiros e humanos, expertise administrativo e capacidade fiscal. Assim, a literatura sobre governança estabelece a relação entre as capacidades materiais dos municípios e a efetiva adoção de políticas públicas (BATISTA, 2015, 2018). Uma leitura cuidadosa da literatura permite compreender que diversos aspectos da questão podem ser levantados e, naturalmente, diversas formas de implementar a variável investigada. Utilizaremos os dados do Índice de gestão fiscal (IFGF), produzido pela Firjan, como indicador da limitação da capacidade de gestão financeira dos municípios¹³.

Por fim, selecionamos o índice de desenvolvimento humano dos municípios (IDH-M) como variável de controle, por representar de forma abrangente a diversidade da realidade socioeconômica brasileira, o progresso e o bem-estar das suas populações. Como é característico das variáveis de controle, seu uso visa reduzir o efeito de variáveis omitidas do modelo sobre a variável dependente, reduzindo confusões e melhorando a comparabilidade.

No próximo capítulo, inserimos ainda uma variável que identifique possíveis efeitos da autocorrelação espacial sobre o modelo. Os modelos de difusão regional consideram que as unidades analisadas, os municípios, não são independentes. Elas se relacionam e suas interações podem ter pesos variados sob diferentes critérios (Berry e Berry, 2007, p. 224). Em uma dessas formulações, têm maior peso as relações que se dão com unidades geograficamente mais próximas. Por falta de um indicador que consiga captar a complexidade dessa relação, utilizamos o código das regiões intermediárias do IBGE¹⁴.

¹³ O Índice De Gestão Fiscal é um valor contínuo que varia entre zero e 1, sendo: Gestão crítica, entre 0 e 0,4, e Gestão de excelência, entre 0,8 e 1. O Índice leva em consideração: (a) a capacidade de arrecadação; (b) os gastos com folha de pagamento; (c) nível de investimentos; (d) o custo da dívida municipal; (e) uso dos restos a pagar. Mais uma vez, o autor utilizou a média para o período entre 2010 e 2020.

¹⁴ Em 2017 o IBGE lançou uma nova divisão geográfica do Brasil. Nela, as regiões intermediárias guardam relação com as antigas mesorregiões. As novas regiões foram propostas com base em evidências empíricas sobre a malha urbana brasileira. (IBGE, 2017)

Quadro 3.. Variáveis e Hipóteses levantadas

Variável	Variável Operacionalizada	Hipótese	Efeito esperado
Temática/ Severidade	População municipal	Quanto mais populoso, maior a chance (risco) de criação de Plano Diretor pelo município.	+
Competição eleitoral : Margem de vitória	Diferença de votos válidos entre o vencedor e o segundo mais votado, em fração do eleitorado local	Quanto maior a competição eleitoral local, maior a chance (risco) de criação de Plano Diretor pelo município.	-
Gestão fiscal	Qualidade da gestão fiscal: Firjan. Assume valores entre zero e 1	Menores restrições administrativas aumentam a chance (risco) de criação de Plano Diretor pelo município.	+
Controle	Índice de Desenvolvimento Humano (IDH-M)	Controle de variáveis omitidas	
Proximidade (fragilidade)	Regiões intermediárias: IBGE	Influências regionais aumentam a chance (risco) de criação de Plano Diretor pelo município.	

FONTE: O autor (2025)

3.3. Regressão de Cox

Uma vez definidas as variáveis que serão investigadas, a pergunta subsequente é: *Como é possível estruturá-las em um modelo estatístico?* Um dos avanços teóricos e empíricos de análises de difusão de políticas consiste na convergência criada entre as abordagens teóricas e as modelagens estatísticas possibilitadas pelos modelos de Cox. De acordo com (BERRY, 2007) o modelo estatístico de Cox é capaz de unificar os modelos de determinantes da difusão de políticas públicas.

Como discutido no capítulo anterior, os modelos estatísticos paramétricos e não paramétricos apresentam um *trade-off* entre dois aspectos principais: os primeiros exigem suposições mais rígidas sobre a distribuição dos dados, enquanto os segundos são mais flexíveis e não dependem dessas premissas. É nesse contexto que se destaca o modelo de Cox nas análises de difusão de políticas

públicas. Ele consiste em uma ponte entre os métodos não-paramétricos (como Kaplan-Meier) e os modelos paramétricos (exponencial, weibull, etc.). Oferece uma abordagem semi-paramétrica, que combina a flexibilidade de não precisarmos especificar uma distribuição para os dados do tempo até a adoção com a possibilidade de incluir variáveis explicativas de efeitos sobre o risco (BOX-STEFFENSMEIER e outros, p. 48, 2004)

Diferente de outras regressões, que assumem uma relação aditiva entre as variáveis explicativas e o intercepto, no modelo de Cox assume-se uma relação multiplicativa entre o risco de base ($h_0(t)$) e os coeficientes estimados. Isso ocorre, porque as covariáveis de interesse afetam o *risco de sobrevivência*, o risco instantâneo de ocorrência do evento (da falha), e não diretamente afetam o tempo de sobrevivência. Dessa forma, os efeitos são calculados como razão de risco exponencial, *hazard ratio* ($\exp(\beta)$).

Regressão Linear :
$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$$

Regressão de Cox :
$$h(t|X) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

3.4. Implementando o modelo de Cox

Até este ponto, apresentamos de forma sucinta a fundamentação teórico-metodológica que conecta a área de pesquisa ao modelo estatístico de Cox, bem como os critérios que orientam a escolha dos determinantes explicativos incluídos na análise. Nosso passo seguinte consiste na implementação prática.

3.4.1. Variáveis Independentes.

Primeiramente carregaremos os pacotes de análise necessários assim como a nova base de dados “cap2. csv”, especialmente preparada com as variáveis de interesse deste segundo capítulo. Além disso, garantimos que as variáveis estão sendo interpretadas pelo *software* nos tipos corretos (*data type*).

```
if(require(tidyverse) == F) install.packages("tidyverse"); require(tidyverse)
if(require(openxlsx) == F) install.packages("openxlsx"); require(openxlsx)
if(require(survival) == F) install.packages("survival"); require(survival)
if(require(survminer) == F) install.packages("survminer"); require(survminer)
if(require(reshape2) == F) install.packages("reshape2"); require(reshape2)
```



```
setwd("C:/direcione/para/a/localização/correta/")
cap2 <- read.csv2("cap2.csv", header=TRUE, sep = ",")

cap2$CodMun <- as.factor(cap2$CodMun)
cap2$media_comp <- as.numeric(cap2$media_comp)
cap2$IFGF.medio <- as.numeric(cap2$IFGF.medio)
cap2$IDHM <- as.numeric(cap2$IDHM)
cap2$Pop <- as.numeric(cap2$Pop)
```

Em segundo lugar, é fundamental conhecer a distribuição das variáveis independentes com que trabalharemos. No nosso caso, todas as variáveis são numéricas e, portanto, o comando *summary()* é útil.

```
> summary(cap2$Pop)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   771   5454   11732   38298   25765 12396372      1
> summary(cap2$IFGF.medio)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
0.07167 0.35164 0.48240 0.48821 0.62267 0.97290      5
> summary(cap2$media_comp)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0220  0.1240  0.1760  0.1946  0.2460  0.9570      2
> summary(cap2$IDHM)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.4180  0.5990  0.6650  0.6592  0.7180  0.8620      7
```

A variável *cap2\$Pop* é a única discreta. Varia entre 771 e mais de 12 milhões de habitantes. A variável *cap2\$IFGF.medio* é índice que varia entre 0,07 (extremo da gestão fiscal crítica) e 0,97 (gestão fiscal de excelência). A variável *cap2\$media_comp* é a fração do eleitorado que representa a diferença eleitoral entre os dois primeiros candidatos a prefeito e varia entre 0,02 (2%, nos municípios mais competitivos) e 0.95 (95%, nos municípios menos competitivos. Esse valor alto indica que o município costuma ter eleições com apenas 1 candidato). Por fim, a variável *cap2\$IDHM* variou entre 0,41 (municípios de baixo desenvolvimento) e 0,86 (municípios de alto desenvolvimento).

Para conhecer melhor o comportamento das variáveis é recomendado observá-las graficamente. Isso pode ser feito de modo simplificado para variáveis numéricas com a combinação de duas funções nativas do R *plot()* e *density()*.

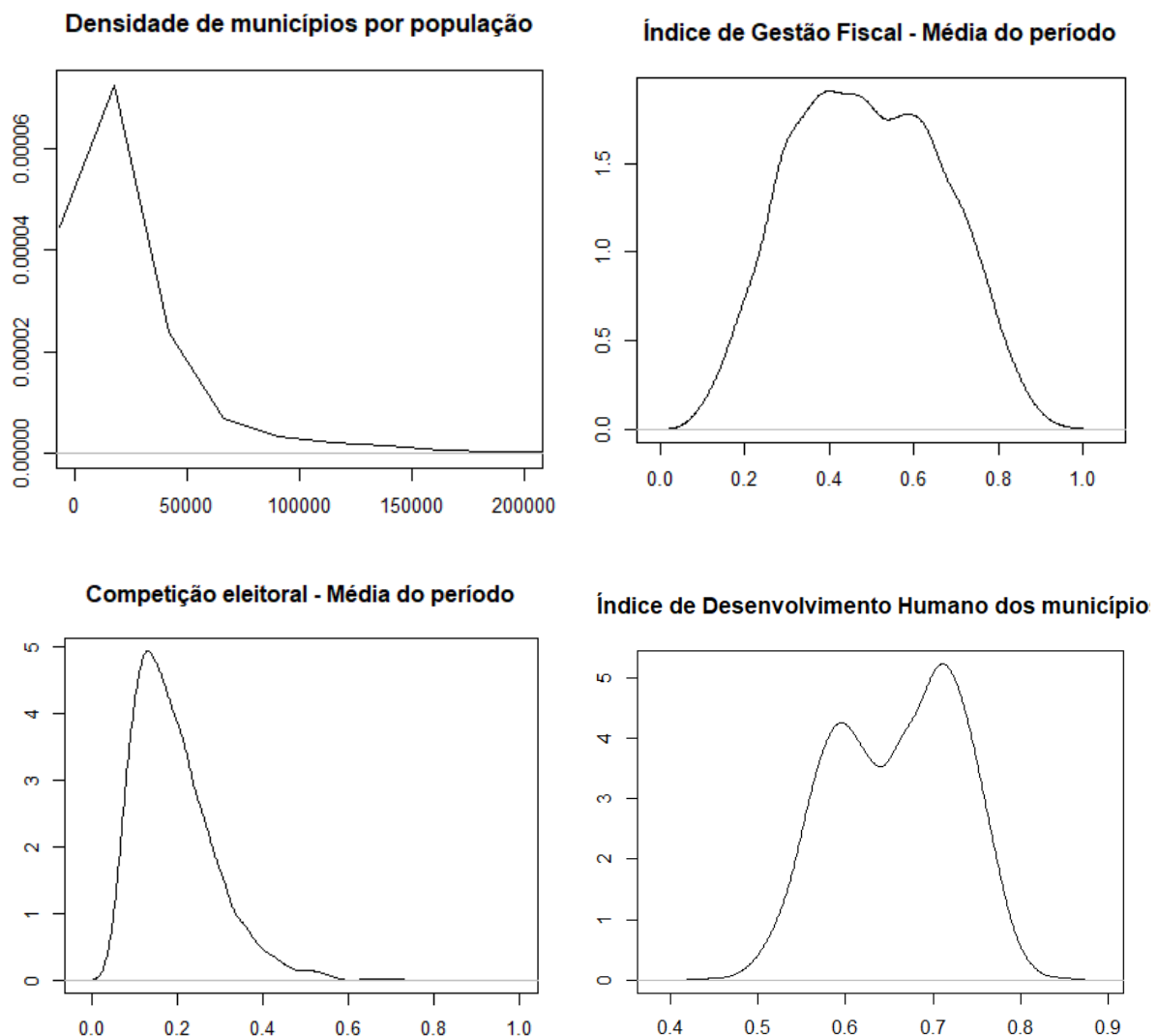
```
options(scipen = 999)
```

```

par(mfrow=c(2,2))
plot(density(cap2$Pop, na.rm = TRUE), xlim = c(0, 200000),
     main = "Densidade de municípios por Log da população")
plot(density(cap2$media_comp, na.rm = TRUE),
     main = "Competição eleitoral - Média do período")
plot(density(cap2$IFGF.medio, na.rm = TRUE),
     main = "Índice de Gestão Fiscal - Média do período")
plot(density(cap2$IDHM, na.rm = TRUE),
     main = "Índice de Desenvolvimento Humano dos municípios")
par(mfrow=c(1,1))

```

Gráfico 4. Distribuição das variáveis independentes: População, Índice de Gestão Fiscal, Competição Eleitoral e Índice de Desenvolvimento Humano.



FONTE: O autor (2025)

Como podemos observar, as variáveis têm grandezas diferentes. Enquanto 3 variáveis são índices que vão de zero a 1 (media_comp, IFGF.medio e IDHM), a

população (Pop) chega a casa dos milhões. Inclusive, para fins de visualização, precisamos limitar a longa cauda à direita (eixo x) do gráfico para essa variável.

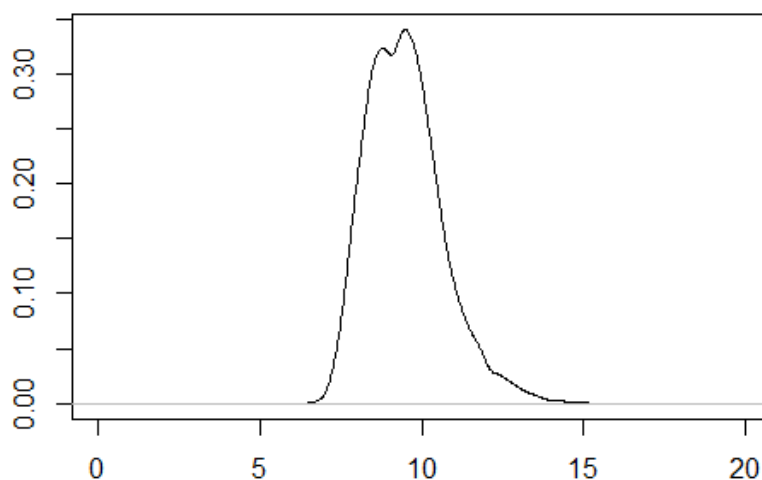
Tanto para fins didáticos quanto para fins substantivos da modelagem da pesquisa, é recomendado reduzir a variabilidade das variáveis, tornando-as mais semelhantes entre si. Isso pode ter efeitos positivos na convergência do modelo, no cálculo dos intervalos de confiança e dos testes de significância. Existem diferentes possibilidades de transformação, como a divisão por um valor significativo, a exemplo da média da variável, ou ainda o uso da função logaritmo. No nosso caso aplicaremos a função logaritmo $\log()$ (logaritmo natural, base $e = 2,718$) para suavizar a discrepância entre a variável 'Pop' e as demais.

```
cap2$Pop <- log(as.numeric(cap2$Pop))
summary(cap2$Pop)
plot(density(cap2$Pop, na.rm = TRUE), xlim = c(0, 20),
     main = "Densidade de municípios por log da população")
```

Dessa forma, temos:

```
> summary(cap2$Pop)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 6.648  8.604   9.370   9.480 10.157 16.333     1
```

Gráfico 5. Distribuição da População, transformada pelo logaritmo natural



FONTE: O autor (2025)

3.4.2. Variável dependente.

Uma vez analisadas as variáveis independentes, podemos partir para a construção do modelo. Para construir o modelo de regressão de cox precisamos criar um objeto que servirá como variável dependente. Como discutido no capítulo anterior, a variável dependente do modelo de sobrevivência é uma variável que inclui ao menos duas informações (podendo ter mais a depender de sua implementação). Por essa razão, existe uma função no pacote survival feita especificamente para a sua implementação. A função `Surv()` é flexível e permite diferentes modos de estruturação da variável. Nós já a utilizamos na construção do estimador kaplan-meier (no capítulo anterior), mas agora vamos entender mais de seus detalhes.

Na sua implementação mais simples os dados apresentam uma observação por unidade de análise e duas variáveis relevantes, as quais serão utilizadas como os dois elementos da variável dependente : (i) o tempo de acompanhamento, ou seja, o tempo que a unidade foi monitorada e (ii) o desfecho identificado até o final do monitoramento. No nosso caso, as duas variáveis correspondem a: (i) o tempo até que a política pública seja adotada para os casos de desfecho positivo (adoção) ou o tempo até o fim do escopo do estudo para os casos de desfecho negativo (não adoção) e (ii) a variável binária indicando desfecho positivo ou negativo.

Primeiro, vamos realizar alguns ajustes na base de dados. Na nossa base, o tempo de acompanhamento está registrado como os anos do calendário e devemos transformá-lo em tempo de acompanhamento. Para isso, utilizaremos o ano da redemocratização (1988) como ano de início do acompanhamento e removeremos as observações que já haviam adotado a política. Essa é uma escolha particular da pesquisa, que se baseia em elementos como: (i) relevância acadêmica do período de estudo selecionado; (ii) relevância do período identificada nos próprios dados; (iii) constância das condições e consequente comparabilidade dos eventos; (iv) dados disponíveis para a análise. Além disso, algumas remoções serão necessárias, como aquelas unidades que adotaram a política antes da data de referência e as linhas com dados faltantes.

```
# filtragens
cap2.f <- cap2
cap2.f <- cap2.f[!(cap2.f$CodMun %in% c("5300108", "5101837")), ]
```

```
cap2.f$Mleg011_ <- as.numeric(cap2.f$Mleg011)-1988
cap2.ff <- cap2.f %>% filter(Mleg011_ > 0)
cap2.ff <- na.omit(cap2.ff)
```

Agora criamos nossa variável dependente (*adoctempo*). Assim como apresentado anteriormente, o objeto *surv* é criado a partir das duas variáveis inseridas como parâmetros da função (*time=Mleg011* e *event=Mleg011*).

```
adoctempo <- Surv(time = as.numeric(cap2.ff$Mleg011_, type="counting"),
                  event = cap2.ff$Mleg01)
```

3.4.3. Modelo e Interpretação.

Para fins didáticos, optamos por criar um único modelo com a combinação das quatro variáveis. Vamos chamá-lo de 'm1'. Para a criação do modelo de Cox, utilizamos a função *coxph()* e dispomos a variável dependente '*adoctempo*' contra as variáveis independentes separadas pelo símbolo de adição '*~ Pop + media_comp ...*'.

```
options(scipen = 0)
m1 <- coxph(adoctempo ~ Pop + media_comp + IFGF.medio + IDHM,
            data = cap2.ff)
m1
survminer::ggforest(m1, fontsize = 1, data = cap2.ff)
```

Para interpretar os coeficientes, é preciso ter em mente, como já mencionado, que os efeitos das variáveis explicativas da regressão de Cox são calculados como razão de risco, *hazard ratio* ($\exp(\beta)$). Por essa razão, os *hazard ratios* menores que 1 ($\exp(\beta) < 1$) indicam um efeito negativo, enquanto os maiores que 1 ($\exp(\beta) > 1$) indicam efeitos positivos. Assim, quanto mais próximos de 1, os *Hazard Ratios* estarão indicando efeitos de magnitude inexpressiva. Por fim, os *Hazard Ratios* nunca serão negativos.

```
coxph(formula=adoctempo ~ Pop+media_comp+IFGF.medio+IDHM, data=cap2.ff)
```

	coef	exp(coef)	se(coef)	z	p
Pop	0.43137	1.53936	0.01298	33.226	< 2e-16
media_comp	-0.25241	0.77693	0.20434	-1.235	0.21673
IFGF.medio	0.40740	1.50290	0.14888	2.736	0.00621
IDHM	3.07405	21.62935	0.36745	8.366	< 2e-16

```
Likelihood ratio test=1491 on 4 df, p=< 2.2e-16
n= 5488, number of events= 2893
```

Desse modo, para um *Hazard Ratio* de 21.629 (como em IDHM, $\exp(\text{coef})$), para cada variação de 1 unidade no IDHM¹⁵, temos o aumento do risco de mais de 21 vezes (2.162%), mantendo-se tudo mais constante. Perceba que no modelo de Cox, pressupõe-se que os riscos sejam proporcionais e, por essa razão, o *Hazard Ratio* é uma constante e indica a razão do risco entre unidades com e sem a unidade adicional de comparação mencionada. Portanto, a taxa, *Hazard Ratio*, permite sobretudo a comparação entre as variáveis do modelo. Neste caso, a variação de uma unidade da variável independente é muito grande, compreendendo toda a amplitude dos dados, o que dificulta a interpretação.

Para *Hazard Ratios* inferiores a 1 (como em *media_comp*, $\exp(\text{coef})$), de 0.777, temos um impacto negativo da variável independente sobre a dependente. Para cada unidade adicional no indicador de competição eleitoral (*media_comp*), espera-se um risco de apenas 77,7% do risco original. Ou seja, o risco é reduzido em aproximadamente 22,3%. Perceba que o indicador *media_comp* tem sentido invertido. Menores valores indicam maior competição eleitoral. Portanto, o aumento no indicador representa a redução da competição eleitoral. A variável *media_comp* (assim como o IDHM) tem amplitude de zero a um. Desse modo, a variação de 77,7% representa uma mudança de um extremo ao outro da amplitude da variável.

Dando sequência à análise dos valores encontrados pelo modelo, temos ainda o efeito positivo das variáveis log natural da população $\log(\text{Pop})$, índice de gestão fiscal *IFGF.medio* e *IDHM*.

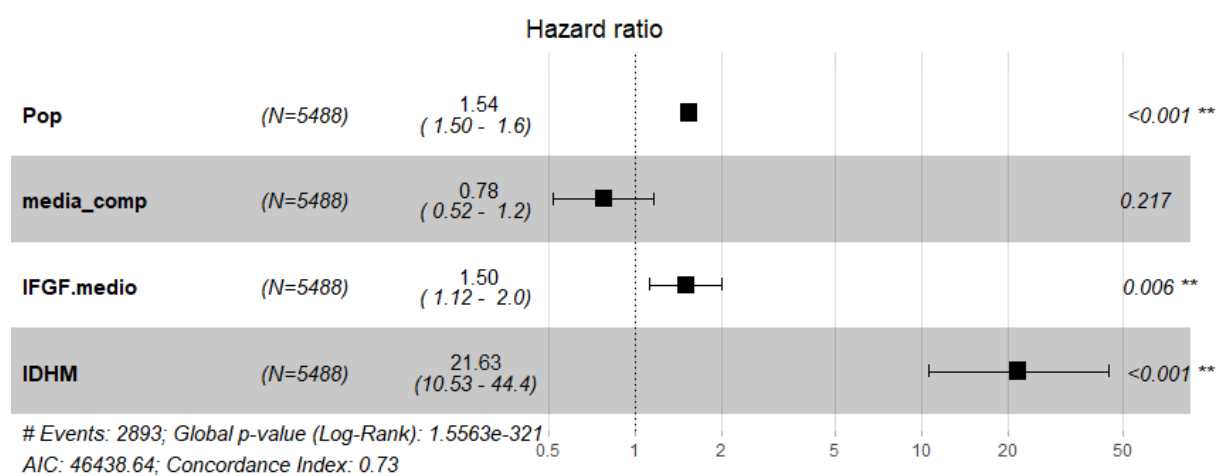
Além dos coeficientes e seus exponenciais, é preciso identificar se esses valores encontrados são considerados estatisticamente significativos. Aqui, observamos o *p-valor* apresentado na coluna final do output do modelo. Em nosso caso, identificamos a significância estatística (< 0.05) das variáveis *Pop*, *IFGF.medio* e *IDHM*, mas não da variável *media_comp*. Podemos assim concluir que apesar do impacto identificado pelo coeficiente desta última variável, os dados não fornecem

¹⁵ Uma clara interpretação dos coeficientes indica sua variação para cada unidade de variação da variável independente, que por sua vez deve corresponder a uma parte da variação total possível (a amplitude). No caso, uma unidade do IDHM corresponde a toda a amplitude da variável, o que torna a interpretação pouco informativa. Decidimos mantê-la desta forma para fins pedagógicos. O objetivo é familiarizar o leitor com a interpretação dos coeficientes assim como são impressos no *output*.

evidências suficientes para rejeitar a hipótese nula, ou seja, que não há impacto da variável. Portanto, não é seguro que o impacto realmente exista.

Para facilitar a compreensão e interpretação do modelo, é possível exibir graficamente os *hazard Ratios* e seus intervalos de confiança. No gráfico 4, estão dispostos os mesmos indicadores do output. Os intervalos de confiança (representados pelo “whisker” ou “bigode”) que cortam o valor 1 (no caso, apenas *media_comp*) não têm resultado estatisticamente significativo. Os intervalos das demais variáveis são confiáveis e suas magnitudes representam efeitos positivos.

Quadro 4.. Output gráfico dos resultados do modelo de Cox



FONTE: O autor (2025)

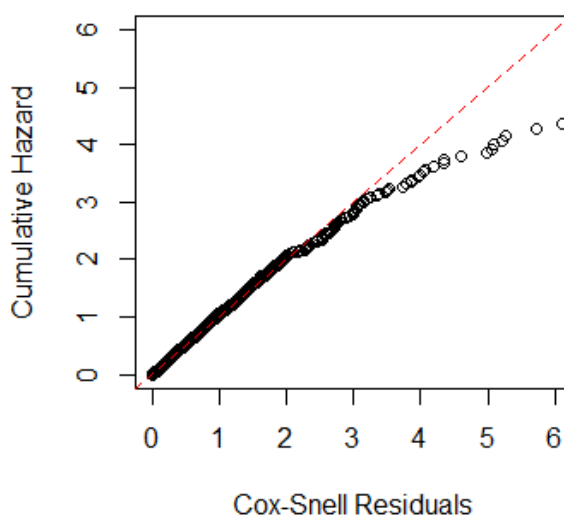
Por fim, outra forma comumente utilizada para avaliar o ajuste global do modelo de Cox consiste na avaliação dos Resíduos de Cox-Snell. Os Resíduos de Cox-Snell representam a diferença entre as estimativas dos tempos de sobrevivência calculadas a partir do modelo implementado e os tempos efetivamente registrados pelos dados (BOX-STEFFENSMEIER e outros, p. 120, 2004). Para que possamos observar a adequação do modelo, plotamos a função dos riscos residuais (Cox-Snell) acumulados, estimada através do método de Nelson-Aalen, contra a função de risco esperada sob o modelo especificado. Espera-se que o gráfico apresente uma linha reta de 45 graus, indicando que os resíduos seguem uma distribuição exponencial.

```
cox_snell <- predict(m1, type = "expected") # Obter os resíduos
surv_cs <- Surv(cox_snell, event = cap2.ff$Mleg01)
fit_cs <- survfit(surv_cs ~ 1) # Ajustar a Kaplan-Meier dos resíduos
H_cs <- -log(fit_cs$surv) # estimativa Nelson-Aalen
plot(fit_cs$time, H_cs,
```

```
# xlim = c(0, 6), ylim = c(0, 6), # iguala limites dos eixos
xlab = "Resíduos de Cox-Snell",
ylab = "Risco Acumulado",
main = "Gráfico dos Resíduos de Cox-Snell")
abline(0, 1, col = "red", lty = 2) # linha 45°
```

Como podemos observar, o modelo faz boas estimativas até certo ponto, quando ele começa a superestimar o risco de adoção (ou subestimar a sobrevivência) dos municípios com tempos maiores de sobrevivência. Essa falha na estimativa pode representar uma violação dos pressupostos do modelo, como discutiremos no próximo capítulo.

Gráfico 6. Distribuição da População, transformada pelo logaritmo natural



FONTE: O autor (2025)

3.5. Resultados

Encerramos o segundo capítulo após explorarmos os fundamentos teóricos que articulam a difusão de políticas públicas com os modelos de análise de sobrevivência, finalizando com aplicação prática do modelo de Cox. Os resultados obtidos até este ponto sugerem que fatores como o tamanho da população, a qualidade da gestão fiscal e níveis mais elevados de desenvolvimento humano estão positivamente associados ao risco de adoção de planos diretores pelos municípios. Por outro lado, a avaliação do ajuste do modelo apresentou indícios de inconsistências. No próximo capítulo, voltamo-nos à avaliação da robustez do modelo proposto, buscando aprimoramentos e ajustes necessários.

4 VERIFICAÇÃO DE PRESSUPOSTOS E AJUSTES DO MODELO

4.1. Introdução.

No capítulo anterior, após discutir as teorias que fundamentam as análises de difusão de políticas públicas, implementamos um modelo de cox. Realizamos uma análise básica dos resultados do modelo. Neste capítulo, aprofundaremos a compreensão do modelo com seus pressupostos metodológicos e os avaliaremos na prática. Por fim, propomos um novo modelo, mais enxuto para os dados selecionados.

Ajustar um modelo estatístico é um processo cíclico, que envolve sucessivos refinamentos. Um tutorial, no entanto, tende a apresentar esse percurso de forma linear, como se houvesse uma sequência direta até o resultado final. Na prática, a construção do modelo exige várias interações e ajustes graduais. Embora essas interações possam ser discutidas sob a perspectiva da necessidade de garantia da neutralidade da pesquisa científica – no sentido de que o processo científico não seja manipulado pelo pesquisador –, o fato é que a calibração de um modelo sempre envolve ajustes progressivos e contínuos. Nas próximas páginas, tentamos reproduzir esse processo, equilibrando a realidade das interações sucessivas com a finalidade pedagógica de um tutorial.

4.2. Verificando pressupostos.

No capítulo anterior nossa análise chegou a resultados parciais a respeito do efeito das variáveis selecionadas (população, competição eleitoral, gestão fiscal e desenvolvimento humano) sobre o risco de adoção de planos diretores nos municípios brasileiros. No entanto, antes de considerarmos os resultados como fortes evidências sobre o assunto, é preciso avaliar se o modelo cumpre os requisitos metodológicos.

Assim como outros modelos estatísticos, o modelo de cox também tem pressupostos que precisam ser atendidos para se garantir a validade dos resultados. Alguns dos pressupostos se assemelham aos de outras regressões, diferenciando-se, por vezes, pela forma de implementação. Outros, são específicos, visto a peculiaridade de nossa variável dependente.

Na sequência, discutimos cada um deles em ordem definida por critérios didáticos, não de importância. Mais uma vez, alinhados com o objetivo do nosso

trabalho, focaremos em explicações curtas, didáticas e que apontem para as aplicações práticas de cada um dos conceitos.

```
if(require(tidyverse) == F) install.packages("tidyverse"); require(tidyverse)
if(require(survival) == F) install.packages("survival"); require(survival)
if(require(survminer) == F) install.packages("survminer"); require(survminer)
if(require(reshape2) == F) install.packages("reshape2"); require(reshape2)
if(require(car) == F) install.packages("car"); require(car)

setwd("C:/direcione/para/a/localização/correta/")
cap2 <- read.csv2("cap2.csv", header=TRUE, sep = ",")

cap2$CodMun <- as.factor(cap2$CodMun)
cap2$media_comp <- as.numeric(cap2$media_comp)
cap2$IFGF.medio <- as.numeric(cap2$IFGF.medio)
cap2$IDHM <- as.numeric(cap2$IDHM)
cap2$Pop <- log(as.numeric(cap2$Pop))
```

4.2.1. Independência entre censura e sobrevivência.

Esse pressuposto reflete a necessidade de que fazer parte ou continuar fazendo parte da análise não pode estar relacionado à ocorrência do evento de interesse (adoção da política). A censura aqui se refere a todas as formas da “censura à direita” - quando a unidade de análise alcança o fim de observação sem ocorrência do evento de interesse, quando da descontinuidade do acompanhamento da unidade no estudo - assim como a outras formas de censura.

Em outras palavras, esse pressuposto pode ser lido como uma independência entre o tempo de sobrevivência e o recorte da pesquisa, momento em que haverá a censura de todas as unidades que permanecem no estudo. Mas ele também inclui a ausência de um viés de seleção. Na medida em que há maior (ou menor) probabilidade de censura (exclusão do estudo) de unidades que sobrevivem mais (ou menos) ao estudo, pode-se entender que está ocorrendo uma *seleção adversa de casos* ao longo da pesquisa, o que é indesejado.

No nosso contexto, a independência entre censura e sobrevivência depende sobretudo de dois fatores. O primeiro refere-se à independência do pesquisador, sua preocupação ética com a produção científica e com sua condução transparente. O segundo fator refere-se à qualidade dos dados utilizados.

Quanto ao primeiro quesito, é essencial que o pesquisador não manipule, ainda que inconscientemente, os critérios de inclusão ou exclusão de casos com

base nos resultados que espera obter. Esse cuidado se realiza com a definição prévia dos critérios de análise e o registro transparente das decisões tomadas ao longo da pesquisa. A pressão por resultados estatisticamente significativos não pode comprometer a neutralidade do estudo.

Quanto à qualidade dos dados, em nosso caso, dependemos em grande medida da qualidade dos dados governamentais. Apesar de o Brasil ser reconhecido como produtor de dados relativamente de boa qualidade, ainda há espaço para a melhora. Cabe ao pesquisador examinar os dados objetivos e verificar possíveis inconsistências que sugiram a falta de autenticidade ou integridade dos dados. Isso inclui checar a completude das informações, avaliar a consistência temporal das variáveis com fatos externos conhecidos e estar atento a lacunas que possam existir. Daí a importância da análise exploratória cuidadosa, realizada em nosso primeiro capítulo.

4.2.2. Distribuição homogênea da variável dependente

Embora o modelo de Cox seja semi-paramétrico e não assuma uma distribuição específica para os tempos de sobrevivência, a distribuição da censura ainda deve ser razoável. A proporção de eventos observados deve ser suficiente para estimar os coeficientes de forma confiável. Uma alta proporção de censuras em determinados momentos do intervalo pode reduzir o poder estatístico do modelo ou indicar que outros fenômenos ou variáveis não estão sendo incluídas no modelo.

Também já tivemos a oportunidade de discutir que o *Estatuto das Cidades* impulsionou um grande boom de adoções dos planos diretores, impulso esse que se iniciou após o prazo de institucionalização da lei. Por essa razão, é razoável acreditar que novas condições (incentivos, modelos de referência, capacidades institucionais, burocráticas e financeiras, etc.) foram criadas. Também é razoável afirmar que essa alteração de condições não está sendo captada pelo modelo. Como observado, o padrão de adoção é muito distinto antes e depois da lei.

Em suma, uma escolha equilibrada nesse ponto é limitar o escopo da pesquisa, reduzindo-a para o período após a promulgação da lei. Essa decisão, reduzirá a variabilidade no padrão de adoções, garantirá maior homogeneidade nas condições reais dos municípios e possivelmente impactará positivamente outros pressupostos, como o da proporcionalidade dos riscos.

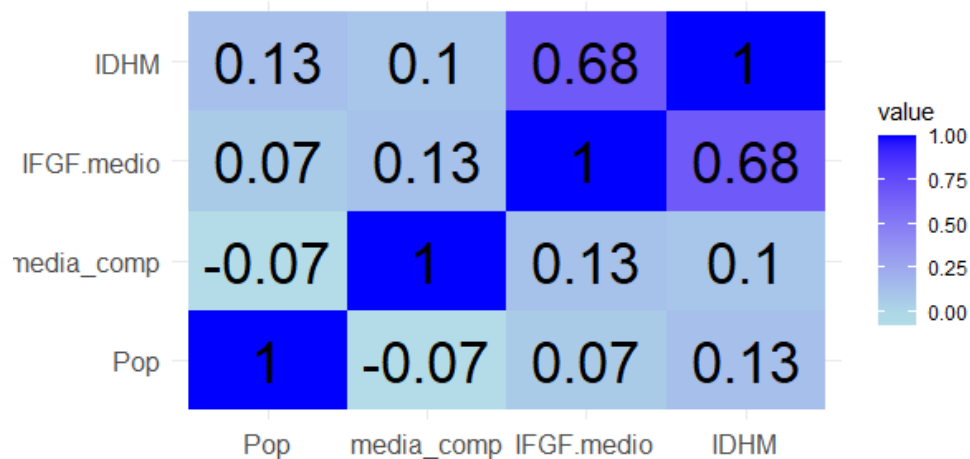
```
cap2.f <- cap2
cap2.f <- cap2.f[!(cap2.f$CodMun %in% c("5300108", "5101837")), ]
cap2.f$Mleg011_ <- as.numeric(cap2.f$Mleg011)-2003 # alterando o ano
cap2.ff <- cap2.f %>% filter(Mleg011_ > 0)
cap2.ff <- na.omit(cap2.ff)
```

4.2.3. Ausência de multicolinearidade

Consiste em pressuposto comum a outras regressões. As covariáveis incluídas no modelo não devem ser altamente correlacionadas entre si. A multicolinearidade pode levar a estimativas instáveis e dificultar a interpretação dos coeficientes. Embora não seja uma exigência estrita do modelo, é recomendável que as variáveis independentes não apresentem alta correlação entre si, pois isso pode comprometer a interpretação dos resultados, já que passam a explicar, em grande parte, a mesma variação da variável dependente. Esse fenômeno, conhecido como multicolinearidade, pode reduzir os efeitos estimados das covariáveis e diminuir sua significância estatística. (Kleinbaum; Klein, 2012).

Para testar a multicolinearidade, realizou-se uma correlação entre as variáveis independentes do modelo. No caso de modelos com várias variáveis independentes, é prático criar uma matrix de correlação. Aqui apresentamos a matriz em forma de heatmap

```
# Matriz de correlação
m_corr <- cor(cap2.ff[c(3,9,10,8)], use = "complete.obs")
m_longa <- reshape2::melt(m_corr) # transforma para matriz longa
# Heatmap de correlações
ggplot(m_longa, aes(x = Var1, y = Var2, fill = value, label =
round(value, 2))) +
  geom_tile() + # Cria os blocos do heatmap
  geom_text(color = "black", size = 10) + # Adiciona os números
nas células
  scale_fill_gradient2(low = "white", mid = "lightblue", high =
"blue", midpoint = 0) + # Escala de cores
  theme_minimal() +
  labs(title = "Heatmap de Correlações", x = "", y = "")
```

Gráfico 7. Heatmap de correlações entre variáveis independentes.**FONTE:** O autor (2025)

Outro teste referenciado pela literatura consiste na utilização da função *vif()* (pacote ‘car’) para calcular a inflação das variâncias do modelo gerada pelas variáveis independentes. Nesse teste estatístico, tem-se que valores menores que 5 identificam ausência de multicolinearidade preocupante. Valores entre 5 e 10 representam alguma multicolinearidade, que pode representar um problema. Valores maiores que 10 representam multicolinearidade forte e a variável pode precisar ser removida. Para aplicar a função é preciso reconstruir o modelo

```
adoctempo <- Surv(time = as.numeric(cap2.ff$Mleg011_, type="counting"),
  event = cap2.ff$Mleg01)
m2 <- coxph(adoctempo ~ Pop+ media_comp+ IFGF.medio+ IDHM, data=cap2.ff)
car::vif(m2)
```

Pop	media_comp	IFGF.medio	IDHM
1.206473	1.022675	1.990107	2.160597

Nos dois testes, o que encontramos é uma baixa correlação entre as variáveis explicativas e um baixo índice de inflação das variâncias. A única correlação moderada-alta encontrada (0.68) acontece entre *IFGF.medio* e *IDHM*.

Assim, concluímos que não parece haver multicolinearidade importante no modelo. Por outro lado, como já foi anteriormente mencionado, o ajuste do modelo estatístico é um processo iterativo e de melhorias graduais e progressivas. Dessa forma, podemos manter essa informação em mente e cogitar em um momento futuro a supressão de uma das variáveis moderadamente colineares.

4.2.4. Linearidade das variáveis contínuas

O modelo de Cox assume que as covariáveis contínuas têm um efeito linear sobre o *logarítmo do hazard* (*risco instantâneo*). Assim como em outras regressões lineares, assume-se que o efeito das variáveis independentes é linear, ou seja, aumentam ou diminuem proporcionalmente ao crescimento ou redução da variável. No caso da regressão de cox, a linearidade diz respeito não ao efeito sobre o tempo, mas à função de risco, mas especificamente ao seu logarítmo. Para garantir essa suposição, algumas abordagens são comuns. Uma delas já foi aplicada nesse trabalho, quando, no capítulo anterior, e reproduzido neste, aplicamos a função logaritmo natural à variável população, dando-lhe distribuição mais suave. Outras transformações possíveis são a raiz quadrada, a divisão pela média, entre outras.

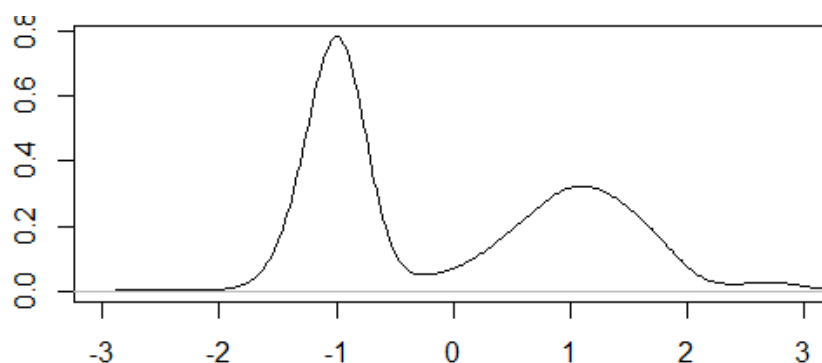
Além da transformação das variáveis independentes, outra abordagem comum é a análise do *resíduo de martingale* (BOX-STEFFENSMEIER e outros, p. 121, 2004). Ele mede a diferença entre o número observado de eventos de interesse e o número esperado de eventos, com base no modelo ajustado. Se o resíduo for próximo de zero, significa que o modelo está bem calibrado. Valores positivos indicam que o modelo subestimou o risco, enquanto valores negativos indicam que o modelo superestimou o risco. Para melhor interpretação é possível plotar a densidade dos resíduos transformados¹⁶ para uma distribuição simétrica em torno de zero, usando *type = "deviance"*.¹⁷

```
martingale_resid <- residuals(m2, type = "deviance") # ou "martingale"
cap2.ff$r.marting <- martingale_resid # acrescenta os resíduos à base
plot(density(martingale_resid), xlim = c(-3,3),
     main = "Densidade dos Resíduos de Martingale transformados")
```

Como é possível observar, os resíduos apresentam um padrão subestimação e superestimação de boa parte dos riscos. O modelo tende a subestimar significativamente uma parte dos riscos (pico em 1), enquanto superestima outra parte (pico em - 1). Além disso, a baixa densidade em zero aponta para a não linearidade das variáveis independentes, modelo com pouco poder explicativo ou a violação do pressuposto da proporcionalidade de riscos, como discutimos a seguir.

¹⁶ BOX-STEFFENSMEIER chama a transformação de “deviance residuals” (BOX-STEFFENSMEIER e outros, p. 123, 2004)

¹⁷ É ainda possível plotar os resíduos contra as covariáveis para verificar se há relações não-lineares ou outros problemas. Não o faremos aqui, mas adicionamos os resíduos à base de dados para facilitar o processo.

Gráfico 8. Densidade dos resíduos de Martingale transformados (m2)

FONTE: O autor (2025)

4.2.5. Proporcionalidade dos Riscos.

O pressuposto da proporcionalidade dos riscos é o pressuposto mais importante do modelo de Cox¹⁸, ainda que frequentemente não seja testado (BOX-STEFFENSMEIER e outros, p. 67, 2024). Por esse pressuposto assume-se que os efeitos das covariáveis são constantes ao longo do tempo e que, portanto, por estarem sob influência das mesmas variáveis, a razão do risco entre os indivíduos permanece também constante. Assim como o pressuposto da linearidade das variáveis contínuas, o pressuposto da proporcionalidade dos riscos requer que haja proporcionalidade. Lá, as variáveis independentes variam proporcionalmente a variação dos riscos. Aqui, o risco entre dois indivíduos, ou seja, o risco relativo entre as unidades de análise permanece constante ao longo do tempo. Dessa forma, a restrição criada pela proporcionalidade dos riscos é mais rigorosa, pois se baseia em características matemáticas específicas do modelo de Cox, que requerem certos comportamentos que nem sempre refletem a realidade dos dados.

A suposição de proporcionalidade dos riscos é o principal custo que se paga pela simplicidade e flexibilidade do modelo de Cox. É uma das condições mais frequentemente violadas e, ao mesmo tempo, uma das mais importantes de se preservar, para que a análise permaneça clara e direta — como é desejável nesse tipo de abordagem.

Para se avaliar a proporcionalidade dos riscos, comumente usa-se os *resíduos de Schoenfeld*. Eles são definidos como a diferença entre os valores observados das covariáveis no momento de um evento e os valores esperados

¹⁸ O modelo de cox também é conhecido como modelo dos riscos proporcionais de Cox.

dessas covariáveis com base no risco relativo estimado pelo modelo de Cox. (BOX-STEFFENSMEIER e outros, p. 121, 2024)

A função `cox.zph()` realiza o chamado *Teste de Grambsch-Therneau*, que testa formalmente a suposição de proporcionalidade dos riscos. Ela avalia se há uma relação significativa entre os resíduos de Schoenfeld e o tempo. A existência dessa correlação indica que o efeito da covariável varia ao longo do tempo. A função `cox.zph()`, *no entanto*, não retorna a magnitude da correlação. Assim, precisamos, além do teste, calcular manualmente a magnitude da correlação de Spearman entre os resíduos de Schoenfeld e o tempo, o rho (ρ).

```
schoen_resid <- residuals(m2, type = "schoenfeld") # resíduos
res.m2 <- cox.zph(m2, transform="identity") # teste de proporcionalidade
rho_manual <- cor(schoen_resid, res.m2$time, # magnitude da correlação
                  method = "spearman")

rho_manual
res.m2
survminer::ggcoxzph(res.m2) # gráfico dos resíduos por variável
```

Como é possível observar, apesar da baixa correlação entre os resíduos de Schoenfeld e o tempo, o teste da hipótese da proporcionalidade encontra significância estatística em todas as variáveis e para o modelo como um todo (GLOBAL). P-valores menores que 0.05 nos fazem rejeitar a hipótese nula e concluir que há uma correlação significativa entre os resíduos e o tempo.

> rho_manual		> res.m5			
Pop	-0.09632786		chisq	df	p
media_comp	0.04233103	Pop	3.23	1	0.0725
IFGF.medio	0.23642204	media_comp	7.08	1	0.0078
IDHM	0.20052153	IFGF.medio	174.83	1	<2e-16
		IDHM	128.91	1	<2e-16
		GLOBAL	188.08	4	<2e-16

4.2.6. Independência dos tempos de sobrevivência.

O último pressuposto a ser discutido diz respeito à independência dos tempos de sobrevivência — ou seja, dos tempos até a adoção. Esse pressuposto estabelece que a decisão de um município adotar determinada política pública não deve interferir na decisão de outro. Em outras palavras, todo o efeito sobre o risco

de adoção deve decorrer das variáveis explicativas incluídas no modelo, e não da influência entre as unidades de análise.

Na prática, esse pressuposto é frequentemente violado, especialmente em estudos de difusão de políticas públicas. A interdependência entre municípios pode ocorrer de diversas formas: quando eles pertencem a grupos com condições comuns (como um risco basal compartilhado), ou quando há influência direta entre eles — fenômeno conhecido como *efeito de contágio*. É comum que municípios adotem políticas ao se espelharem em experiências vizinhas, ao se beneficiarem do conhecimento técnico de outros ou até mesmo motivados por disputas que aceleram a adoção de inovações frente a municípios concorrentes (BERRY, 2007)

Esse tipo de interação entre unidades, embora viole o pressuposto de independência, representa um aspecto essencial da realidade nos processos de difusão. Autores especializados na área tratam dessa violação não como um erro a ser eliminado, mas como um componente que precisa ser modelado adequadamente (BERRY, p. 229, 2007)

Existem diferentes estratégias para lidar com a interdependência. Em casos de difusão regional, é possível incorporar a *autocorrelação espacial* ao modelo, identificando e dimensionando a influência entre unidades vizinhas. Quando os efeitos de contágio se alteram ao longo do tempo, uma abordagem mais robusta envolve o uso do *modelo estendido de Cox*, conhecido como *modelo tempo-dependente de Cox*. Esse modelo permite não apenas representar grupos de contágio, mas também incorporar variações temporais nesses efeitos. No entanto, com aplicação mais complexa, exige dados em séries temporais transversais (cross-section time series), o que ultrapassa o escopo deste trabalho.

Alternativamente, é possível lidar com a heterogeneidade não observada entre grupos de forma mais modesta, por meio de termos de *fragilidade* (*frailty()*) ou ajustar o modelo via *estratificação* (*strata()*). O uso do termo de *fragilidade* é indicado quando se presume que há fatores não observados influenciando o risco entre diferentes grupos, permitindo que cada grupo tenha uma função de risco própria. Já a estratificação é útil quando há heterogeneidade observada entre os grupos — ela não estima a magnitude dessa heterogeneidade, mas ajusta o modelo permitindo riscos basais distintos entre os estratos, mantendo constantes os coeficientes das variáveis explicativas. Ambas as abordagens necessitam identificar uma variável categórica que represente os grupos — o que nem sempre reflete a

complexidade dos padrões reais de contágio ou autocorrelação espacial entre os municípios.

Por simplicidade, adotamos a função *frailty()* com base em uma variável categórica que representa as regiões intermediárias do IBGE¹⁹. O modelo foi então especificado novamente com o acréscimo do termo, permitindo estimar o impacto da heterogeneidade não observada sobre os tempos de adoção. O termo fragilidade introduz um efeito aleatório que permite identificar heterogeneidade entre os grupos. Variâncias, quando próximas a zero, indicam nenhuma variabilidade introduzida, sugerindo que o modelo de Cox sem fragilidade já descreveria bem os dados.

```
# comportamento da variável
cap2.ff$cod.interm <- as.factor(cap2.ff$cod.interm)
length(table(cap2.ff$cod.interm))
hist(table(cap2.ff$cod.interm), breaks=20,
      main = "Histograma de Regiões por quantidade de Municípios",
      xlab = "Número de municípios", ylab = "Número de regiões")
# acrescentando o termo de fragilidade ao modelo
cox_frailty <- coxph(adoctempo ~ Pop + media_comp + IFGF.medio + IDHM +
                    frailty(cod.interm), data = cap2.ff)
cox_frailty
summary(cox_frailty)
```

No nosso *output*, a estatística **Variance of random effect = 0.366** indica uma variabilidade moderada atribuída ao termo de fragilidade. Além disso, o p-valor significativo (2.5e-98) reforça a relevância estatística desse efeito no modelo. Assim temos boas razões para acreditar na existência de efeito da interação entre os municípios, justificando a inclusão do termo de fragilidade.

4.3. Propondo um novo modelo e Resultados

Com base nos resultados do modelo anterior, propomos um novo modelo mais parcimonioso, incluindo apenas a variável que apresentou indícios de atender ao pressuposto da proporcionalidade dos riscos: a população, transformada pelo logaritmo natural ($\log(\text{Pop})$). Além disso, acrescentamos um termo de fragilidade, utilizando a variável de código das regiões intermediárias do IBGE (*frailty(cod.interm)*), a fim de capturar possíveis efeitos não observados relacionados à proximidade geográfica.

¹⁹ Ver nota explicativa 14.

```
# Novo modelo
m3 <- coxph(adoctempo ~ Pop + frailty(cod.interm), data = cap2.ff)
summary(m3)
```

Para avaliar o ajuste do modelo, calculamos os resíduos de Cox-Snell, ajustando uma curva de Kaplan-Meier contra esses resíduos e comparando-a com a linha de 45°, que representa o ajuste ideal:

```
cox_snell <- predict(m3, type = "expected") # Obter resíduos Cox-Snell
surv_cs <- Surv(cox_snell, event = cap2.ff$Mleg01)
fit_cs <- survfit(surv_cs ~ 1) # Ajustar curva Kaplan-Meier dos resíduos
H_cs <- -log(fit_cs$surv) # estimativa Nelson-Aalen
plot(fit_cs$time, H_cs,
     # xlim = c(0, 6), ylim = c(0, 6), # iguala limites dos eixos
     xlab = "Resíduos de Cox-Snell",
     ylab = "Risco Acumulado",
     main = "Gráfico dos Resíduos de Cox-Snell")
abline(0, 1, col = "red", lty = 2) # linha 45°
```

Os resultados do novo modelo são mais robustos do que os obtidos anteriormente. A variável $\log(\text{Pop})$ apresentou efeito positivo e estatisticamente significativo sobre o risco de adoção dos planos diretores (p-value = 1.0e-224). Interpretando o coeficiente estimado ($\beta = 0.4989$), podemos afirmar que um aumento de aproximadamente 2,7 vezes na população (isto é, um incremento de 1 unidade no logaritmo natural da população) está associado a um aumento de cerca de 64,7% no risco de adoção do plano diretor ($\exp(\text{coef}) = 1.647$).

O termo de fragilidade também foi estatisticamente significativo (p-valor = 3.8e-129), com uma variância estimada de 0.446, indicando moderada heterogeneidade entre as regiões intermediárias. Isso sugere que os municípios inseridos em diferentes contextos regionais possuem riscos distintos de adoção, justificando a inclusão do efeito aleatório no modelo.

```
> summary(m3)
coxph(formula = adoctempo ~ Pop + frailty(cod.interm), data = cap2.ff)

n= 5274, number of events= 2679

             coef      se(coef)      se2  Chisq      DF      p
Pop           0.4989 0.01559   0.01534 1024.1     1.0 1.0e-224
frailty(cod.interm)                        937.5 114.8 3.8e-129

exp(coef) exp(-coef) lower .95 upper .95
```

```

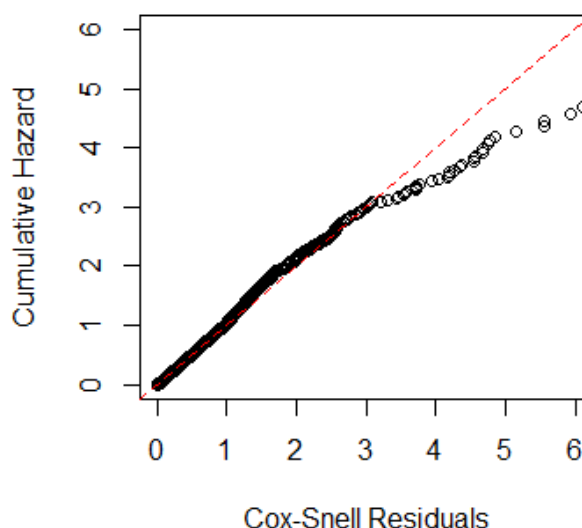
Pop      1.647      0.6072      1.597      1.698

Iterations: 10 outer, 34 Newton-Raphson
      Variance of random effect= 0.4463689      I-likelihood = -21224
Degrees of freedom for terms=      1.0 114.8
Concordance= 0.773 (se = 0.004 )
Likelihood ratio test= 2171 on 115.8 df,      p=<2e-16

```

Quanto à avaliação gráfica com os resíduos de Cox-Snell, observa-se que o novo modelo continua apresentando um bom ajuste nos tempos iniciais, com a curva seguindo a linha de 45°. Entretanto, à medida que o tempo avança, o modelo tende a superestimar o risco de adoção. Ainda assim, a inclinação da curva é menos acentuada do que a observada no modelo anterior, indicando uma leve melhora no ajuste geral.

Gráfico 10. Resíduos de Cox-Snell



FONTE: O autor (2025)

```

# resíduos de schoenfeld
schoen_resid3 <- residuals(m3, type = "schoenfeld")
res.m3 <- cox.zph(m3, transform = "identity")
rho_manual3 <- cor(schoen_resid3, res.m3$time, method="spearman")
rho_manual3
res.m3
survminer::ggcoxzph(res.m3)
# resíduos de martingale
martingale_resid3 <- residuals(m3, type = "deviance") # ou "martingale"
cap2.ff$r.marting3 <- martingale_resid3 # add resíduos à base
plot(density(martingale_resid3), xlim = c(-3,3),

```

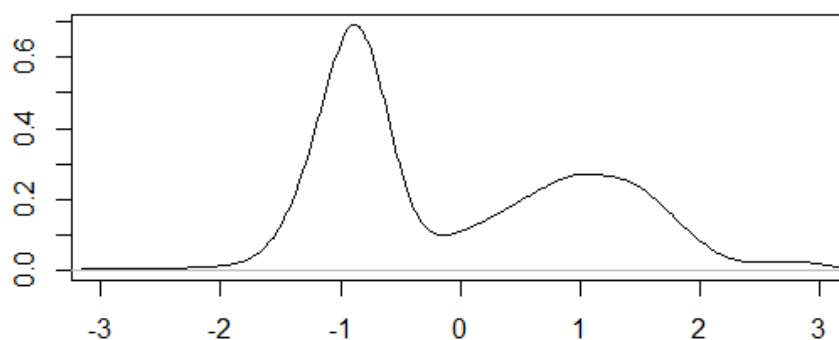
```
main = "Densidade dos Resíduos de Martingale")
```

A avaliação dos *resíduos de Schoenfeld* e do *teste de Grambsch-Therneau* indicam que não há evidências de violação do pressuposto de proporcionalidade dos riscos para a variável $\log(Pop)$. O teste não rejeita a hipótese nula, de ausência de relação entre os resíduos e o tempo de adoção, como demonstram o p-valor não significativo ($p = 0.43$) e a baixa correlação observada ($\rho = -0.114$). Esses resultados reforçam a adequação do modelo ao pressuposto central da regressão de Cox:

```
> rho_manual3      [1] -0.1141705
> res.m3
      chisq      df      p
Pop    0.592    0.97 0.43
GLOBAL 0.592  115.79 1.00
```

Por outro lado, a avaliação da linearidade da variável contínua por meio dos *resíduos de Martingale* mostra pouca evolução em relação aos modelos anteriores. A distribuição dos resíduos ainda revela padrões de subestimação e superestimação do risco, com baixa densidade em torno de zero, indicando que o efeito da variável $\log(Pop)$ sobre o risco de adoção pode não ser estritamente linear. Esse comportamento sugere uma possível violação da suposição de linearidade entre a covariável e o logaritmo da função de risco (hazard).

Gráfico 10. Densidade dos Resíduos de Martingale transformados (m3)



FONTE: O autor (2025)

Dado que o modelo possui apenas uma variável independente, não há necessidade de verificar os demais pressupostos.

5 CONCLUSÃO

Neste trabalho apresentamos um guia didático para aplicação da análise de sobrevivência, com foco na difusão de políticas públicas. A partir da perspectiva metodológica da Event History Analysis (EHA), conduzimos um percurso que conecta a literatura sobre difusão de políticas públicas no Brasil com os fundamentos e aplicações práticas da técnica estatística.

No primeiro capítulo, apresentamos os fundamentos da análise de sobrevivência, a estrutura dos dados e a operacionalização da variável dependente. Discutimos os principais conceitos teóricos e apresentamos a implementação de análises descritivas, destacando a relevância da curva de Kaplan-Meier para a visualização da sobrevivência. A política escolhida para ilustração empírica, os Planos Diretores Municipais, adequou-se ao propósito didático do trabalho, permitindo identificar padrões de adoção coerentes com marcos legais e outros referenciais conhecidos.

Na sequência, o segundo capítulo introduziu os modelos estatísticos propriamente ditos, com ênfase na regressão semi-paramétrica de Cox. Foram descritas as hipóteses teóricas, operacionalizadas as variáveis independentes e dependentes e o modelo foi implementado. Destacamos, nesse ponto, como fatores como população municipal, competição eleitoral e capacidade fiscal poderiam influenciar a chance de adoção dos Planos Diretores. O modelo de Cox captou o efeito dessas variáveis sobre o tempo até a adoção da política.

O terceiro capítulo abordou a verificação dos pressupostos do modelo de Cox, como a proporcionalidade dos riscos, independência dos tempos de sobrevivência e ausência de multicolinearidade. Verificou-se a fragilidade na verificação de alguns pressupostos. Na última seção, foram propostos ajustes ao modelo para melhor adequá-lo aos dados. Esse exercício reforçou o caráter instrucional do trabalho, permitindo ao leitor acompanhar etapas pouco exploradas na literatura. O exercício reforçou tanto a importância das etapas, para a correta aplicação da metodologia, quanto o processo real de ajuste de um modelo.

Como principais resultados, verificou-se que a adoção dos Planos Diretores ocorre com maior intensidade em municípios mais populosos, além de ser influenciada por aspectos regionais. A análise permitiu não apenas visualizar

padrões gerais de difusão, mas também interpretar a relevância de cada variável com base na razão de risco.

A principal contribuição deste trabalho é sua proposta didática. Ao sistematizar, passo a passo, os procedimentos para realizar uma análise de sobrevivência em R, espera-se que o material sirva como porta de entrada para pesquisadores iniciantes e intermediários que desejem aplicar esse tipo de técnica na investigação da difusão de políticas públicas. A disponibilização dos códigos e dos dados reforça o compromisso com a transparência e a reprodutibilidade científica, alinhando-se às boas práticas de ciência aberta.

Entre as limitações que podem ser apontadas, mencionamos, em primeiro lugar o não aprofundamento no tratamento de dados. O tratamento e a análise exploratória de dados brutos ou mais complexos pode demandar do investigador conhecimentos que estão muito além das possibilidades desse trabalho. Sugere-se nesta etapa o uso de materiais específicos. Em segundo lugar, a modelagem de efeitos da autocorrelação espacial e de variáveis tempo-dependentes é um passo fundamental para o campo de pesquisa e carece de literatura que a aborde didaticamente. Embora relevantes, esses elementos não foram contemplados neste trabalho, mas se apresentam como candidatos promissores para produções futuras que desejem aprofundar ou dar continuidade à proposta deste trabalho.

Espera-se, portanto, que este guia facilite o uso da análise de sobrevivência no campo da ciência política, mas também incentive aplicações empíricas mais avançadas, contribuindo para o fortalecimento da pesquisa quantitativa na área de políticas públicas.

REFERÊNCIAS

ABRUCIO, Fernando; SANO, Hironobu; SEGATTO, Catarina Ianni. Cooperação interestadual: o papel do Consad e do Consed na disseminação de reformas e políticas públicas. In: FARIA, C. A.; COELHO, D.; SILVA, S. (Orgs.). Difusão de políticas públicas. São Bernardo do Campo - Editora UFABC, p. 103-132, 2016.

ARRETCHE, Marta. Estado Federativo e Políticas Sociais: Determinantes da Descentralização. Rio de Janeiro: Editora Revan. 3ª ed., 2011.

BASTOS, Mario Rubem do Couto. A Implementação de Políticas Públicas com Agenciamento e Múltiplos Principais: A Análise da Difusão do Programa Bolsa Família. Tese de Doutorado em Economia, Universidade de Brasília, 2015.

BATISTA, Mariana. A difusão da lei de acesso à Informação nos municípios brasileiros: fatores internos e externos. Cadernos Enap, Caderno 53, Brasília: Enap, 2017.

BERRY, Frances Stokes; BERRY, William D. Innovation and Diffusion Models in Policy Research. In: Theories of the Policy Process, edited by Paul. A. Sabatier. Boulder: Westview. 2007.

CARVALHO, Maria do Carmo Brant de. Política de Assistência Social e seu reordenamento Institucional. Material didático do Curso para Gestores de Fundo Municipal de Assistência Social. FUNDAP, mimeo, 1997.

CARVALHO, Gregory dos Passos. Difusão de planos municipais de saneamento básico: condicionantes da formação da agenda governamental em perspectiva comparada. Dissertação de mestrado, Brasília: UNB. 2015.

CARVALHO, Grégory dos Passos; COELHO, Denilson Bandeira. Difusão da Estratégia Saúde da Família: fatores explicativos da adesão dos municípios de São Paulo, Brasil, no contexto coercitivo de indução pela esfera federal. *Ciência & Saúde Coletiva*, v. 28, n. 8, ago. 2023. Disponível em: <https://www.scielo.br/j/csc/a/jrTTqxZLfDFgjHSfz4m5WQ/?lang=pt>. Acesso em: 5/4/2025.

CAVALCANTI, Pedro. Descentralização de políticas públicas sob a ótica neoinstitucional: uma revisão de literatura. Revista de Administração Pública. Vol. 45, n. 6. Pp. 1781-1804. 2011.

COELHO, Denilson Bandeira. Political Competition and the diffusion of conditional cash transfers in Brazil. Brazilian Political Science Review. N. 6, v. 2, pp. 56-87, 2012.

FAGUET, Jean-Paul. Decentralization and local government performance. Technical Consultation on Decentralization. Rome: FAO, 1997

PALOTTI, Pedro Lucas; PEREIRA, Celina; CAMÕES, Marizaura, Reis. A adoção da carreira de gestor governamental no Brasil: Há um processo de difusão em andamento?. In: FARIA,

C. A.; COELHO, D.; SILVA, S. (Orgs.). "Difusão de políticas públicas". São Paulo: Ed. UFABC, 2016. p. 133-155.

SOUZA, Celina. Para entender a nossa barafunda federativa. Insight Inteligência (Rio de Janeiro), v. 61, pp. 60-70, 2013.

SPADA, Paolo. Political Competition and the Diffusion of Policy Innovations in local government: The case of Participatory Budgeting in Brazil. Paper presented at LASA Congress. 2010.

_____. The Diffusion of Participatory Governance Innovations: A Panel Data Analysis of the Adoption and Survival of Participatory Budgeting in Brazil. 2014.

STRANG, David. Adding Social Structure to Diffusion Models: An Event History Framework. Sociological Methods & Research. Vol. 19. Ed. 3. Pp. 324-353. 1991

SUGIYAMA, Natasha B. Bottom-up Policy Diffusion: National Emulation of a Conditional Cash Transfer Program in Brazil. Publius: The Journal of Federalism. V. 4, pp. 25-51. 2012

SUGIYAMA, Natasha B. Theories of Policy Diffusion: Social Sector Reform in Brazil. Comparative Political Studies, 41 (2), 193-216. 2008.

TATEMOTO, Letícia Caroline Barche. Lei De Acesso À Informação e Sua Implementação Nos Executivos Estaduais. Tese de Doutorado em Ciência Política. Biblioteca Depositária: FAFICH, UFMG, Belo Horizonte. 2016

WAMPLER, B. 2007. Participatory Budgeting in Brazil: Contestation, Cooperation, and Accountability. University Park; The Pennsylvania State University Press.

_____. A difusão do Orçamento Participativo brasileiro: "boas práticas" devem ser promovidas?. Opinião Pública, vol. 14, nº 1, p.65-95. Campinas, 2008.

_____. The spread and impact of participatory budgeting: Brazil, Peru, China, South Korea and the Philippines. In: FARIA, C. A.; COELHO, D.; SILVA, S. (Orgs.). "Difusão de políticas públicas". São Paulo: Ed. UFABC, 2016. p. 157-185.

REFERÊNCIAS EM METODOLOGIA

AQUINO, Jackson. R para cientistas sociais. UESC, 2014. Disponível em: https://www.uesc.br/editora/livrosdigitais_20140513/r_cientistas.pdf Acesso em: 5/4/2025

BOX-STEFFENSMEIER, Janet M. and JONES, Bradford S. Event History Modeling: a Guide for Social Scientists. New York: Cambridge University Press, 2004.

CARVALHO, Marília Sá; ANDREOZZI, Valeska Lima; CODEÇO, Claudia torres; CAMPOS, Dayse Pereira; BARBOSA, Maria Tereza Serrano; SHIMAKURA, Silvia Emiko. Análise de sobrevivência: teoria e aplicações em saúde. Editora FIOCRUZ. 2011.

EPIDEMIOLOGIST'S R HANDBOOK. Survival analysis. Disponível em: https://www.epirhandbook.com/pt/new_pages/survival_analysis.pt.html. Acesso em: 5/4/2025

FLORES, Alejandro Quiroz. *Survival analysis: a new guide for social scientists*. Cambridge: Cambridge University Press, 2022. (*Elements in Quantitative and Computational Methods for the Social Sciences*). ISBN 9781009054508.

GOLUB, Jonathan. Survival analysis. In: BOX-STEFFENSMEIER, Janet M.; BRADY, Henry E.; COLLIER, David (org.). *The Oxford handbook of political methodology*. Oxford: Oxford University Press, 2008. cap. 23, p. 530–546.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. *Cidades e Estados: Informações básicas dos municípios brasileiros*. Rio de Janeiro: IBGE, 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/educacao/10586-pesquisa-de-informacoes-basicas-municipais.html> . Acesso em: 5 abr. 2025.

KLEINBAUM, D. G., & KLEIN, M. (2012). *Survival Analysis: A Self-Learning Text* (3rd ed.). New York: Springer. <https://doi.org/10.1007/978-1-4419-6646-9>

THERNEAU, T. M. A Package for Survival Analysis in R. 2023. Disponível em: <https://CRAN.R-project.org/package=survival>. Acesso em: 5 abr. 2025.

UCLA – INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION. *Survival analysis with R*. Disponível em: https://stats.oarc.ucla.edu/wp-content/uploads/2025/02/survival_r_full.html. Acesso em: 5 abr. 2025.

WICKHAM, Hadley. Tidy data. *Journal of Statistical Software*, [S.l.], v. 59, n. 10, p. 1–23, 2014. DOI: <https://doi.org/10.18637/jss.v059.i10>.