

RESUMO

Não temos tempo a perder: uma introdução à análise de sobrevivência

O que fazer quando a variável dependente é o tempo até a ocorrência de um determinado evento? Este artigo apresenta uma introdução intuitiva à análise de sobrevivência com foco no estimador de Kaplan-Meier. Nosso público alvo são estudantes de graduação e pós-graduação em fases iniciais de treinamento. Metodologicamente, o desenho de pesquisa utiliza simulação básica para apresentar a implementação computacional e a interpretação substantiva dos coeficientes. Todas as rotinas computacionais foram reportadas com o objetivo de aumentar a transparência e garantir a replicabilidade dos resultados. Ainda adotamos o protocolo TIER. Com este artigo esperamos difundir a utilização da análise de sobrevivência em Ciência Política e contribuir com o avanço da literatura sobre metodologia política.

Palavras-chave: métodos quantitativos; análise de sobrevivência; estimador de Kaplan-Meier; ciência política; SPSS.

ABSTRACT

What to do when the variable of interest is the time until the occurrence of a particular event? This paper presents an intuitive introduction to survival analysis focusing on the Kaplan-Meier estimator. Our target audience are undergraduate and graduate students in the early stages of training. Methodologically, the research design uses data with basic simulation to present in detail the computational implementation and the substantive interpretation of the results. All computational routines were duly reported in order to increase transparency and ensure replicability of the results. Moreover, we adopted the protocol TIER. With this article we hope to disseminate the use of survival analysis in Political Science and contribute to the literature on political methodology.

Keywords: quantitative methods; survival analysis; Kaplan-Meier estimator; political science; SPSS.

Não temos tempo a perder: uma introdução à análise de sobrevivência¹

Rodrigo Lins²

Dalson Figueiredo Filho³

Enivaldo Rocha⁴

INTRODUÇÃO

Quanto tempo dura uma constituição? Quanto tempo demora para um indivíduo reincidir criminalmente? Quanto tempo demora a formação de um gabinete parlamentar? Em muitas situações de pesquisa, a variável de interesse é o tempo transcorrido até a ocorrência de um determinado evento, o que a literatura denomina de falha (failure) (ALLISON, 1984). Diferente de variáveis quantitativas e discretas, o tempo tem características específicas que requerem a utilização de técnicas especiais. Este artigo apresenta uma introdução à análise de sobrevivência com foco no estimador Kaplan-Meier (KM). Esse modelo foi escolhido por representar o início dos estudos de análise de sobrevivência. Metodologicamente, o desenho de pesquisa utiliza simulação básica para apresentar a implementação computacional e a interpretação substantiva dos resultados.

Mas qual é a importância de um artigo pedagógico? A principal motivação é a escassez de aplicações nas Ciências Sociais em geral e na Ciência Política em particular. Por exemplo, Krueger e Lewis-Beck (2008) examinaram artigos publicados entre 1990 e 2005 na *American Political Science Review* (APSR), *American Journal of Political Science* (AJPS), e *Journal of Politics* (JOP) e não identificaram nenhuma ocorrência de trabalhos. No Brasil, o cenário é semelhante. Revisamos os arquivos das revistas DADOS, *Brazilian Political*

1 Materiais de replicação estão disponíveis em: osf.io/dpebu. Agradecemos ao Berkeley Initiative for Transparency in the Social Sciences (BITSS) e ao grupo de Métodos de Pesquisa em Ciência Política da Universidade Federal de Pernambuco (UFPE) pelo treinamento e pelos comentários, respectivamente. Erros eventuais são todos nossos.

2 Doutorando em Ciência Política (DCP/UFPE) e membro do grupo Métodos de Pesquisa em Ciência Política.

3 Professor Adjunto do Departamento de Ciência Política da Universidade Federal de Pernambuco (DCP/UFPE).

4 Professor Titular do Departamento de Ciência Política da Universidade Federal de Pernambuco (DCP/UFPE).

Science Review (BPSR), Revista de Sociologia e Política (RSP) e Revista Brasileira de Ciências Sociais (RBCS) e encontramos apenas dois trabalhos que utilizam alguma forma de análise de sobrevivência: Coêlho (2012) e Coêlho, Cavalcante e Turgeon (2016)⁵. Ambos os trabalhos, no entanto, tinham como objetivo o estudo de difusão de políticas públicas.

Outro elemento que justifica a relevância de um artigo didático é a ausência da análise de sobrevivência de cursos intensivos de metodologia. Revisamos as programações do curso de verão da IPSA-USP entre 2010 e 2016 e não encontramos nenhuma oferta de cursos sobre análise de sobrevivência. Examinamos ainda as ementas dos cursos de análise de dados da Universidade de São Paulo, Universidade Federal de Minas Gerais e Universidade Federal de Pernambuco e, salvo melhor catalogação, também não encontramos iniciativas pedagógicas ao ensino da análise de sobrevivência⁶.

Dessa forma, enquanto a análise de sobrevivência é amplamente utilizada em outras disciplinas como Medicina (MEIER-KRIESCHE, SCHOLD E KAPLAN, 2004), Economia (BOOCKMANN et al., 2007) e Engenharia (WU E LIU, 2002), ela ainda não foi incorporada ao arsenal metodológico das Ciências Sociais. E, na ausência de iniciativas de difusão da técnica, é bastante improvável que esse cenário seja modificado.

Com o objetivo de superar essa lacuna, o restante do artigo está organizado da seguinte maneira. A primeira parte apresenta as características da análise de sobrevivência. Depois disso, explicamos as principais propriedades do estimador de Kaplan-Meier. A terceira seção ilustra o passo a passo da aplicação prática e a interpretação dos resultados a partir de dados simulados. Por fim, apresentamos as conclusões.

ANÁLISE DE SOBREVIVÊNCIA: O QUE É E PARA QUE SERVE? ⁷

O termo análise de sobrevivência surgiu a partir de pesquisas biomédicas, uma vez que o método teve origem no interesse de pesquisadores no estudo de mortalidade, levando em conta o tempo transcorrido entre o diagnóstico de uma dada doença e a morte do paciente (GUO, 2010). O primeiro estudo a utilizar tal formato foi publicado por John Graunt em 1662. Intitulado “Natural and Political Observations Made Upon the Bills of Mortality”, o autor calculou a estatística de mortalidade a partir de relatórios semanais que detalhavam os números e as causas das mortes no subúrbio de Londres (GUO, 2010).

Tecnicamente, a análise de sobrevivência pode ser definida como um conjunto de procedimentos estatísticos usados para se analisar dados cuja variável de interesse é o tempo até a ocorrência de um dado evento (BOX-STEFFENSMEIER E JONES, 1997; COLOSIMO E

5 Utilizamos as ferramentas de busca dos sites das próprias revistas. Os termos de busca foram: survival analysis; event history analysis; duration models; survival models; failure-time models. Também foram feitas buscas com a tradução para o português.

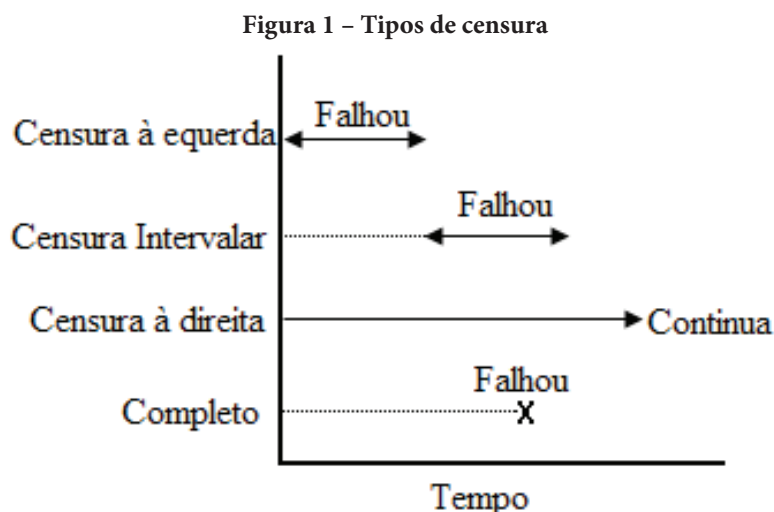
6 Em 2014, o professor Bernardo Lanza ofertou uma disciplina de análise de sobrevivência durante o Programa de Treinamento Intensivo em Metodologia Quantitativa (MQ) da Universidade Federal de Minas Gerais (UFMG). Similarmente, o professor Ernesto Amaral disponibilizou eletronicamente os slides de um curso introdutório ofertado em 2010, ver: < <http://www.ernestoamaral.com/docs/dcp859b4-102/Aula132.pdf>>. A professora Ani Katchova da Ohio State University oferta um curso on-line, ver < <https://www.youtube.com/watch?v=fTX8GghbBPc>>.

7 Para ter informações sobre as outras técnicas e abordagens de análise de sobrevivência, ver: Allison (1985); Davis e Lawrance (1989); Morita, Lee e Mowday (1993); Box-Steffensmeier e Zorn (2002); Akritas (2004); Box-Steffensmeier e Jones (2004); Colosimo e Giolo (2006); e Hosmer, Lemeshow e May (2008).

GIOLO, 2006; HOSMER, LEMESHOW E MAY, 2008). A literatura utiliza diferentes termos para denominar a ocorrência do evento que pode significar a morte de um paciente ou a cura de uma doença. Nas Ciências Sociais, pode representar o tempo até a reincidência criminal (SCHMIDT E WITTE, 1989) ou até que uma constituição seja substituída ou sofra uma emenda (NEGRETTO, 2012)⁸.

De acordo com Kleinbaum e Klein (2005), a análise de sobrevivência tem três objetivos: (1) estimar e interpretar funções de sobrevivência e/ou de risco dos dados de sobrevivência; (2) comparar funções de sobrevivência e/ou risco; e (3) avaliar a relação das variáveis explicativas e o tempo de sobrevivência. Em termos não técnicos, a meta é comparar grupos diferentes e avaliar o impacto de variáveis para a sobrevivência dos casos. Usualmente, a variável dependente em análises de sobrevivência contém duas informações. A primeira é o tempo que leva até que o caso experimente a falha (evento) e a outra é uma variável dicotômica indicando o seu status (0 = censurado; 1 = experimentou o evento).

Outra importante característica de desenhos de pesquisa dos modelos de análise de sobrevivência são as observações censuradas (AGRESTI E FINLAY, 2012). A censura equivale a observação parcial da resposta que foi interrompida por alguma razão, não permitindo a observação completa do tempo de falha⁹. Em outras palavras, a censura ocorre quando temos algumas informações sobre o tempo de sobrevivência de uma observação, mas não temos o tempo exato (KLEINBAUM E KLEIN, 2005). A figura 1 ilustra diferentes tipos de censura.



Fonte: elaboração dos autores, a partir de: http://help.synthesis8.com/weibull_alta8/data_types.htm

A censura à esquerda acontece quando o tempo de sobrevivência é menor do que a duração da pesquisa. Por exemplo, quando um presidente sofre um impeachment no terceiro ano de um estudo de 10 anos. Por sua vez, tem-se censura intervalar quando se sabe em que intervalo a falha aconteceu, mas não se sabe o tempo exato (COLLET, 1994). Isso acontece quando o evento acontece entre dois momentos de mensuração, como quando um trabalhador de minas não continha problemas de respiração no momento da primeira mensura-

8 Para outras aplicações em Ciência Política, ver: King et al. (1990); Gasiorowski (1995); Bernhard, Reenock e Nordstrom (2003); Slovik (2008); Mainwaring e Pérez-Liñán (2013); e Martínez (2015).

9 Ver: < <http://www.de.ufpb.br/~tarciana/MDS/Aula1.pdf> >

ção, mas apresenta os sintomas na segunda aferição. Por sua vez, a censura à direita ocorre quando o estudo chegou ao fim sem que um caso tenha sofrido o evento. Por exemplo, o programa de acompanhamento de ex-detentos terminou e o indivíduo não reincidiu em atividades criminosas. Por fim, tem-se a censura aleatória quando se sabe o período de início e fim da observação, mas ela não faz mais parte do estudo por outros motivos, como um paciente que se muda durante um experimento, ou, no caso de um estudo semelhante ao de Schmidt e Witte (1989), um ex-presidiário que se mudou da região examinada. Os modelos de análise de sobrevivência podem trabalhar com observações censuradas à direita e aleatórias. As censuras à esquerda, por outro lado, não podem ser usadas.

Para Colosimo e Giolo (2006), é importante definir três elementos que compõem o tempo de falha¹⁰. O primeiro é o tempo inicial, que pode ser o momento em que se inicia a observação do estudo ou o início de algum tratamento. O segundo é a escala de medida, que é o tempo, medido em horas, dias, anos, mandatos ou sessões legislativas, por exemplo. O terceiro elemento é o evento de interesse, conhecido como falha. O evento definido como uma falha costuma ser algo indesejado, como a morte (UNGER et al., 2014) ou a queda de um regime político (MAINWARING E PÉREZ-LIÑÁN, 2013). No entanto, o evento também pode ser um resultado (BOX-STEFFENSMEIER E JONES, 1997), como no caso da sanção do senado para indicações presidenciais (OSTRANDER, 2015) ou transição do desemprego para o estado de ocupado (ANTIGO E MACHADO, 2006).

De forma semelhante às regressões, é preciso observar a significância das estimativas, que no modelo de Kaplan-Meier para dois ou mais grupos é obtido através do teste de log-rank¹¹. Ele é usado para testar a hipótese nula de que a probabilidade de ocorrência do evento é a mesma entre os grupos (BLAND E ALTMAN, 2004). Em situações em que o p-valor não é significativo ($p\text{-valor} > 0,05$) devemos inferir que as curvas de sobrevivência dos grupos são iguais. Contrariamente, quando o resultado é significativo ($p\text{-valor} < 0,05$) infere-se que as curvas são diferentes, ou seja, rejeita-se a hipótese nula em função da hipótese alternativa.

Finalmente, existe uma série de ferramentas em análise de sobrevivência. Além da curva de Kaplan-Meier – foco deste trabalho – existem outras opções, como o estimador de Nelson-Aalen¹², o estimador da tabela de vida¹³, e o modelo de regressão de Cox¹⁴. O quadro 1 resume as principais características de uma análise de sobrevivência.

10 Para uma discussão detalhada sobre os tipos de falha ver Prinja, Gupta e Verma (2010).

11 Existem outros testes de hipótese, mas como esclarecem Colosimo e Giolo (2006), o teste log-rank costuma ser o mais utilizado em análises de sobrevivência. Por isso, optamos por ele. Este teste é baseado no número de vezes da ocorrência do evento e calculado da seguinte maneira: o número de falhas por grupo e o número esperado se não houvesse diferença entre os grupos (BLAND E ALTMAN, 2004).

12 Para um passo a passo do estimador de Nelson-Aalen no R e no SAS, ver [http://sas-and-r.blogspot.com.br/2010/05/example-739-nelson-aalen-estimate-of.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+SASandR+\(SAS+and+R\)](http://sas-and-r.blogspot.com.br/2010/05/example-739-nelson-aalen-estimate-of.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+SASandR+(SAS+and+R))

13 Para uma melhor compreensão da estimativa de Tabela de Vida, ver Sambath, Ramanujam e Chinaiyan (2013).

14 O modelo de Regressão de Cox é o mais utilizado em análises de sobrevivência. Ele permite analisar o impacto de variáveis explicativas no tempo de sobrevivência das observações. Para uma aplicação no Stata, ver http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/. Para uma aplicação do modelo de Cox com time change covariates ver Zhou (2001).

Quadro 1 – Característica de Análises de Sobrevivência

Item	Descrição
Variável Dependente	Tempo até a ocorrência de um evento
Evento de Interesse	Falha
Tempo Inicial	Momento de início do estudo
Escala de Medida	Tempo real (dias, semanas, meses, anos)
Dados	Censurados
Teste de Hipótese	Log-Rank

Fonte: elaboração dos autores.

Em síntese, a análise de sobrevivência deve ser utilizada quando a variável dependente é o tempo até a ocorrência de um determinado evento. A ocorrência do evento é denominada falha e pode ser mensurada a partir diferentes unidades (dias, semanas, meses, anos). Uma característica central dos dados é a censura que exige técnicas específicas de análise. Como nosso foco é o estimado Kaplan-Meier, iremos explorar o teste de hipótese através do teste de log-rank.

CONCEITOS-CHAVE PARA O ESTIMADOR DE KAPLAN-MEIER

Nesta seção apresentamos o estimador Kaplan-Meier (KM) e suas principais características. O estimador é uma estatística não-paramétrica que tem como objetivo aferir a função de sobrevivência (survival function), vejamos:

$$S(t) = \frac{n^{\circ} \text{ de casos vivos no início} - n^{\circ} \text{ de casos que falharam}}{n^{\circ} \text{ de casos vivos no início}} \quad (1)$$

Onde t representa o tempo, maior do que zero. A função de sobrevivência possui uma interpretação bastante direta. Ela representa a probabilidade de que um dado caso não falhe – ou seja, sobreviva – em um determinado momento do estudo. Colosimo e Giolo (2006) explicam que a função é “a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver ao tempo t ” (COLOSIMO e GIOLO, 2006: 20). Guo (2010) segue a mesma linha ao interpretar a função de sobrevivência como a mensuração da “probability of not having the event (surviving to, or remaining in the subject set of having no event) by time t ” (GUO, 2010: 36). Por exemplo, imagine uma base de dados com 20 observações e nenhuma falha. Tem-se então $20-0/20=1$. Ou seja, todos os casos sobreviveram. A probabilidade de sobrevivência é 1. Contrariamente, em um cenário em que todas as observações experimentam a falha tem-se: $20-20/20=0$. Ou seja, ninguém sobreviveu.

O estimador de Kaplan-Meier não utiliza a função de sobrevivência. No entanto, a maior parte dos pacotes estatísticos permite analisar graficamente a função de risco. Por isso, também a explicamos brevemente: a função de risco (hazard function) ou de taxa de falha pode ser compreendida como uma mensuração de risco. Para Box-Steffensmeier e Jones (1997), essa função indica o risco que um caso incorre em um momento específico, dado que o evento ainda não aconteceu. Respostas para perguntas como “qual o risco de um regime cair?” se tornam mais interessantes. Por isso, ela é apontada como a “mais importante

função de análise de sobrevivência” (GUO, 2010: 35) e como “mais informativa do que a função de sobrevivência” (COLOSIMO E GIOLO, 2006: 23). O valor da função de risco é sempre positivo, mas como não é uma probabilidade, pode exceder o valor 1.

Em síntese, a função de sobrevivência indica a chance de uma observação sobreviver em um momento específico do estudo. Já a função de risco representa a taxa de risco de falha da observação, levando em conta que ela ainda não havia experimentado o evento até um momento específico. Nossa sugestão é a de que os pesquisadores que trabalham com o estimador do produto-limite¹⁵ utilizem ambas as funções com o objetivo de aumentar a robustez das análises. O Cambridge Dictionary of Statistics (2010) define substantivamente as duas funções da seguinte forma:

Survival Function: The probability that the survival time of an individual is longer than some particular value. A plot of this probability against time is called a survival curve and is a useful component in the analysis of such data (EVERITT E SKRONDAL, 2010, p. 423).

Hazard Function: The risk that an individual experiences an event (death, improvement etc.) in a small time interval, given that the individual has survived up to the beginning of the interval. It is a measure of how likely an individual is to experience an event as a function of time (EVERITT E SKRONDAL, 2010, p. 201).

Graficamente, a curva de Kaplan-Meier serve para descrever o tempo de sobrevivência para todos os casos e identificar o período da falha para as observações que sofreram falha e o momento de censura para os casos censurados. Ainda é possível obter a estatística descritiva dos dados (média, mediana, desvio padrão, etc.) para melhor compreender a distribuição do tempo de sobrevivência. Por fim, também é possível realizar testes de hipótese quando se tem dois ou mais grupos. Nestes casos, é elaborada uma curva de função de sobrevivência para cada um deles (FÁVERO, 2014). Os limites de intervalo de tempo em uma curva de Kaplan-Meier são o tempo de falha, que devem ser organizados de forma ascendente. A mensuração da média e da mediana também permite observar as diferenças substantivas entre os grupos.

SIMULAÇÃO

Depois de apresentar as principais características do estimador de Kaplan-Meier, o próximo passo é descrever a base de dados¹⁶. A simulação consiste em uma matriz de dados com 40 observações, que serão tratados como países democráticos. A variável de interesse é a sobrevivência do regime, ou seja, a quantidade de anos que se leva até que essas democracias sejam derrubadas por regimes autoritários. O evento de interesse é a queda do regime democrático e foi codificado como “1” em caso de retrocesso institucional e “0” caso con-

15 A ideia do estimador do produto-limite é, na realidade, o limite de probabilidades condicionais. Isso é: a probabilidade de uma observação sobreviver a t_2 , tendo em conta que ele sobreviveu a t_1 .

16 É importante ressaltar que os dados foram criados aleatoriamente, sem nenhuma ligação com países reais. Dessa forma, buscamos apenas ilustrar o que foi debatido até agora, e não realizar qualquer análise inferencial ou teste de hipótese substantiva. O banco construído está no apêndice e também disponível no Dataverse (doi:10.7910/DVN/BYQSUX) e no Academia.

Não temos tempo a perder: uma introdução à análise de sobrevivência

trário. A escala de medida será representada em anos e o estudo terá um alcance de cinco décadas. Ou seja: aqueles países que sobreviverem ao menos meio século serão censurados à direita.

Os países estão distribuídos em dois grupos de mesmo tamanho ($n = 20$). Trataremos um grupo como formado por democracias parlamentaristas e o outro por democracias presidencialistas. A variável dummy, portanto, está configurada da seguinte forma: 0 = parlamentarista e 1 = presidencialista. Os dados foram simulados para testar a hipótese de que a sobrevivência dos regimes democráticos é mais fácil no parlamentarismo (LINZ, 1990; LINZ E VALENZUELA, 1994; CHEIBUB E LIMONGI, 2002; LIJPHART, 2008). Dessa forma, construímos as variáveis com uma distribuição de Bernoulli, com probabilidade de falha de 0,5 para o parlamentarismo e 0,8 para o presidencialismo¹⁷.

O próximo passo é analisar as informações¹⁸. A tabela 1 apresenta um resumo do processamento dos dados.

Tabela 1 – Resumo do processamento

No total	No de Eventos	Censurado	
		No	Porcentagem
40	25	15	37,5%

Fonte: elaboração dos autores.

O banco de dados contém 40 casos, dos quais 25 experimentaram o evento, ou seja, ocorreu retrocesso institucional (62,5%). Os censurados totalizam 15 países, representando 37,5% da amostra. Em seguida deve-se examinar a tabela de sobrevivência.

Tabela 2 – Tabela de Sobrevivência

			Proporção cumulativa sobrevivendo ao tempo			
	Ano	Status	Estimativa	Erro padrão	No de eventos cumulativos	No de casos restantes
1	3	1	0,975	0,025	1	39
2	7	1	0,950	0,034	2	38
3	10	1	0,925	0,042	3	37
4	12	1	0,900	0,047	4	36
5	16	1	0,875	0,052	5	35
6	18	1	0,850	0,056	6	34
7	19	1	0,825	0,060	7	33
8	21	1	0,800	0,063	8	32

17 O passo para criá-las é: Transformar à Calcular variável. Na caixa variável de destino é preciso digitar o nome da nova variável (no nosso banco, intitulamos Censura) e optar por Rv.Bernoulli em Funções e variáveis especiais.

18 O comando a ser seguido é: Analisar à Sobrevivência à Kaplan-Meier... Na caixa “Tempo” deve-se colocar a variável que mede o tempo de sobrevivência dos casos. Em “Status” se coloca a variável que indica as falhas ou censuras. Em seguida, ao clicar em Opções, deve-se marcar Tabela de Sobrevivência, Sobrevivência de Média e Mediana e Quartis. Em gráficos, preencher Sobrevivência e Risco.

9	24	1	0,775	0,066	9	31
10	26	1	0,750	0,068	10	30
11	28	1	0,725	0,071	11	29
12	29	1	0,700	0,072	12	28
13	30	1	0,675	0,074	13	27
14	33	1	0,650	0,075	14	26
15	34	1	0,625	0,077	15	25
16	35	1	0,600	0,077	16	24
17	37	1	0,575	0,078	17	23
18	38	1	0,550	0,079	18	22
19	39	1	0,525	0,079	19	21
20	41	1	-	-	20	20
21	41	1	0,475	0,079	21	19
22	44	1	-	-	22	18
23	44	1	0,425	0,078	23	17
24	48	1	0,400	0,077	24	16
25	49	1	0,375	0,077	25	15
26	50	0	-	-	25	14
27	50	0	-	-	25	13
28	50	0	-	-	25	12
29	50	0	-	-	25	11
30	50	0	-	-	25	10
31	50	0	-	-	25	9
32	50	0	-	-	25	8
33	50	0	-	-	25	7
34	50	0	-	-	25	6
35	50	0	-	-	25	5
36	50	0	-	-	25	4
37	50	0	-	-	25	3
38	50	0	-	-	25	2
39	50	0	-	-	25	1
40	50	0	-	-	25	0

Fonte: elaboração dos autores.

A coluna “Ano” indica exatamente o ano de ocorrência da falha. Por exemplo, o primeiro retrocesso institucional ocorreu no terceiro ano de observação. A segunda no sétimo ano. A terceira falhou após uma década. E assim sucessivamente. A coluna “Status” indica se houve (1) ou não a falha (0). Dos 25 países que sofreram retrocesso institucional, 13 foram nos 30 primeiros anos observados. Nos anos 41 e 44, dois países sofreram as falhas. Em casos como esses, onde há repetição, os dados só aparecem uma única vez.

A tabela 3 sumariza a média e a mediana para os casos agrupados, com intervalo de con-

fiança de 95%.

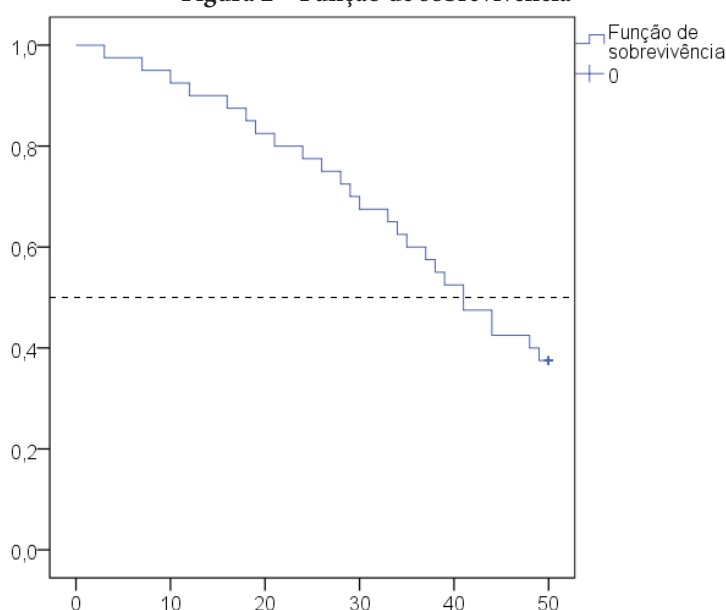
Tabela 3 – Médias e Medianas para tempo de sobrevivência

		Média				Mediana	
Intervalo de Con- fiança de 95%	Inter- valo de Con- fiança de 95%						
Estimativa	Erro Padrão	Limite Inferior	Limite Superior	Estimativa	Erro Padrão	Limite Inferior	Limite Superior
36,900	2,269	32,452	41,348	41,000	3,790	33,572	48,428

Fonte: elaboração dos autores.

Em média, os países sobrevivem 36,9 anos, com um erro padrão de 2,269. A mediana é um pouco mais alta, atingindo 41 anos, com um erro padrão de 3,790. Como o banco só possui censura à direita, o sinal (+) só aparece no último ano de observação (50). Esses são os casos que não haviam sofrido o evento de interesse no fim do estudo e totalizam 15 países. A figura 2 apresenta a curva de sobrevivência de todos os casos agrupados (parlamentaristas + presidencialistas).

Figura 2 – Função de sobrevivência¹⁹

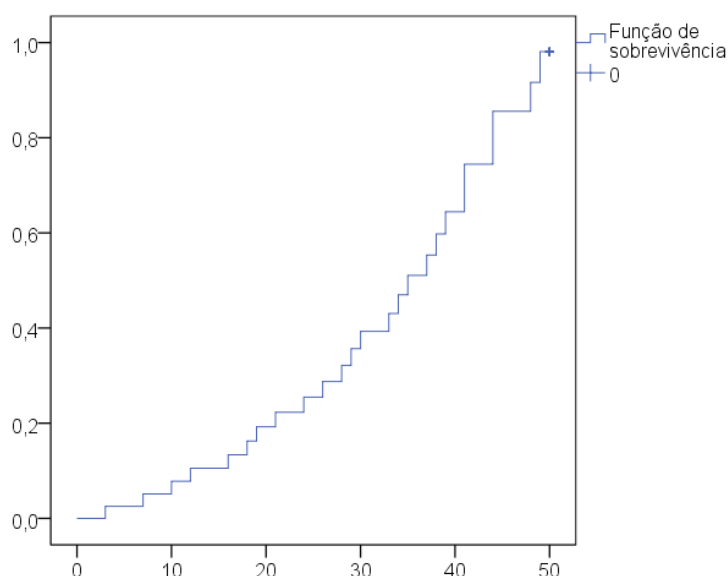


Fonte: elaboração dos autores.

A linha pontilhada representa a mediana do tempo de sobrevivência (0,5). Isso implica dizer que 50% dos casos de falha, levando em conta todas as democracias, já haviam acontecido aproximadamente nos primeiros 40 anos. A figura 3 apresenta a função de risco para os dados agrupados.

¹⁹ Os gráficos apresentados aqui foram editados no output do SPSS.

Figura 3 – Função de risco



Fonte: elaboração dos autores.

Levando em conta que a taxa de falha é ascendente, espera-se que quanto mais o tempo passar, maior será a chance de que as democracias sofram um retrocesso institucional.

Uma vez que a análise foi realizada com os dados agrupados, podemos examinar o desempenho dos dois grupos separadamente: países parlamentaristas versus presidencialistas²⁰. A tabela 4 apresenta a análise segmentada por grupo.

Tabela 4 – Resumo de processamento de dados

Grupo	No total	No de Eventos	Censurado	
			N	Porcentagem
Parlamentarismo	20	10	10	50%
Presidencialismo	20	15	5	25%
Geral	40	25	15	37,5%

Fonte: elaboração dos autores.

A tabela 4 mostra que dos 40 casos, 20 são democracias parlamentaristas e outros 20 são presidencialistas. No primeiro grupo houve 10 falhas (50%) e 10 censuras (50%). Já no segundo, ocorreram 15 quedas de democracia (75%) e cinco censuras (25%). Em termos substantivos, esses resultados sinalizam que o sistema presidencialista é mais passível de retrocesso institucional quando comparado ao parlamentarista. A tabela 5 apresenta tabela de sobrevivência por grupo.

²⁰ Para realizar a análise segmentada por grupo é preciso repetir a mesma sequência do exemplo anterior, incluindo na caixa “Fator” a variável que indica a separação dos grupos (os países parlamentaristas e presidencialistas). Em Comparar Fator, marcar Posição de Log, para poder observar o teste log-rank.

Não temos tempo a perder: uma introdução à análise de sobrevivência

Tabela 5 – Tabela de Sobrevivência por grupo

		Ano	Status	Proporção cumulativa sobrevivendo ao tempo		No de eventos cumulativos	No de casos restantes
				Estimativa	Erro padrão		
Parlamentarismo	1	19	1	0,950	0,049	1	19
	2	26	1	0,900	0,067	2	18
	3	29	1	0,850	0,080	3	17
	4	34	1	0,800	0,089	4	16
	5	37	1	0,750	0,097	5	15
	6	39	1	0,700	0,102	6	14
	7	41	1	0,650	0,107	7	13
	8	44	1	0,600	0,110	8	12
	9	48	1	0,550	0,111	9	11
	10	49	1	0,500	0,112	10	10
	11	50	0	-	-	10	9
	12	50	0	-	-	10	8
	13	50	0	-	-	10	7
	14	50	0	-	-	10	6
	15	50	0	-	-	10	5
	16	50	0	-	-	10	4
	17	50	0	-	-	10	3
	18	50	0	-	-	10	2
	19	50	0	-	-	10	1
	20	50	0	-	-	10	0
Presidencialismo	1	3	1	0,950	0,049	1	19
	2	7	1	0,900	0,067	2	18
	3	10	1	0,850	0,080	3	17
	4	12	1	0,800	0,089	4	16
	5	16	1	0,750	0,097	6	15
	6	18	1	0,700	0,102	7	14
	7	21	1	0,650	0,107	8	13
	8	24	1	0,600	0,110	9	12
	9	28	1	0,550	0,111	10	11
	10	30	1	0,500	0,112	11	10
	11	33	1	0,450	0,111	12	9
	12	35	1	0,400	0,110	13	8
	13	38	1	0,350	0,107	14	7
	14	41	1	0,300	0,102	15	6
	15	44	1	0,250	0,097	15	5
	16	50	0	-	-	15	4
	17	50	0	-	-	15	3
	18	50	0	-	-	15	2
	19	50	0	-	-	15	1

	20	50	0	-	-	15	0
--	----	----	---	---	---	----	---

Fonte: elaboração dos autores.

No parlamentarismo, metade dos casos falharam, ou seja, experimentaram retrocesso institucional e metade foram censurados. A primeira falha aconteceu no ano 19 e a última no ano 49. Já no presidencialismo, as falhas acontecem mais cedo e em maior número. Nos dez primeiros anos, três falhas acontecem, sendo a primeira já no terceiro ano de estudo. A tabela 6 apresenta as médias e as medianas para o tempo de sobrevivência por grupo.

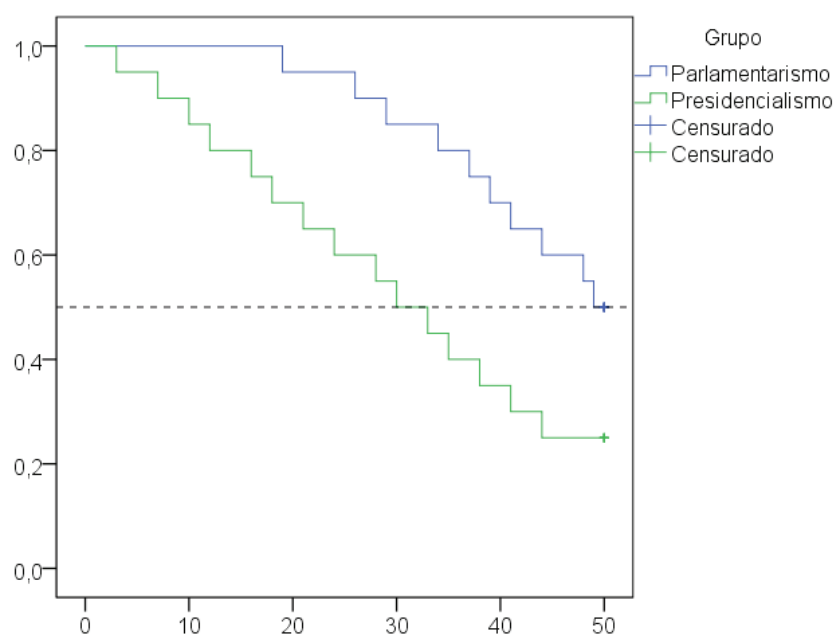
Tabela 6 – Médias e Medianas para tempo de sobrevivência por grupo

Grupo	Média				Mediana			
	Intervalo de Confiança de 95%				Intervalo de Confiança de 95%			
	Estimativa	Erro Padrão	Limite Inferior	Limite Superior	Estimativa	Erro Padrão	Limite Inferior	Limite Superior
Parlamentarismo	43,300	2,091	39,202	47,398	49,000	-	-	-
Presidencialismo	30,500	3,483	23,673	37,327	30,000	5,590	19,043	40,957
Geral	36,900	2,269	32,452	41,348	41,000	3,790	33,572	48,428

Fonte: elaboração dos autores.

A média do tempo de sobrevivência no parlamentarismo (43,3 anos) é maior do que no presidencialismo (30,5 anos), assim como a estimativa de mediana. Enquanto no parlamentarismo o valor é 39, no presidencialismo, a estimativa da mediana é de 30 anos. A figura 4 ilustra a função de sobrevivência por grupo.

Figura 4 – Funções de Sobrevivência



Fonte: elaboração dos autores.

Comparativamente, esses resultados indicam que o desenho institucional parlamentarista garante maior sobrevivência às democracias quando comparado com os sistemas presiden-

cialistas. Isso porque a curva de sobrevivência do grupo de países parlamentaristas é mais alta do que a do grupo de países presidencialistas. Além disso, é possível perceber que a mediana das falhas dos regimes presidencialistas é de aproximadamente 30 anos, enquanto a dos regimes parlamentaristas é de aproximadamente 50 anos.

Por fim, deve-se avaliar a magnitude e a significância da diferença entre as funções de sobrevivência de cada sistema. A tabela 7 sumariza essas informações a partir do teste de log-rank.

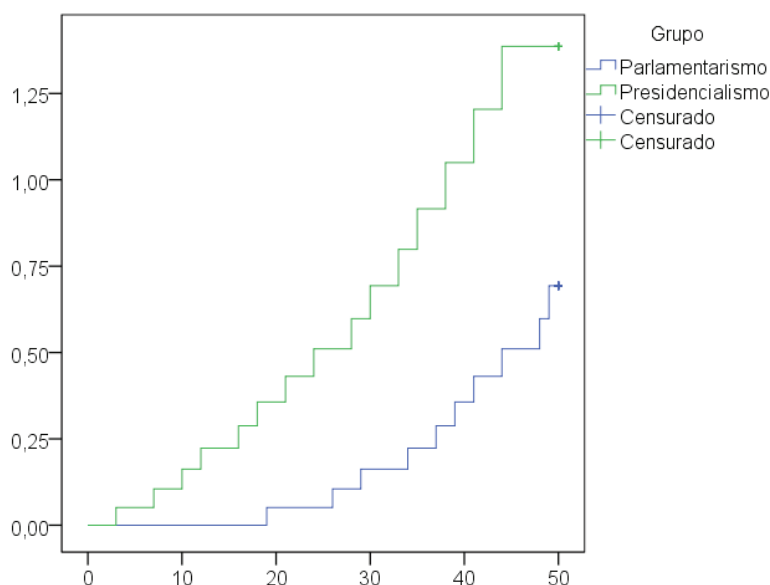
Tabela 7 – Comparações globais

	Qui-quadrado	gl	Sig.
Log Rank	5,093	1	0,024

Fonte: elaboração dos autores.

Observamos um qui-quadrado de 5,093 e um p-valor de 0,024. Dessa forma, deve-se rejeitar a hipótese nula em função da hipótese alternativa. Em termos menos técnicos, isso significa que a diferença entre os grupos é significativa. A figura 5 apresenta a função de risco, separada por grupo.

Figura 5 – Função de risco por grupo



Fonte: elaboração dos autores.

Levando em conta a função de risco dos grupos, a diferença entre eles é ainda maior. A partir do quinto ano de observação, já há um desprendimento dos dois grupos até que se chega no ano 50, quando a função de risco dos casos censurados no presidencialismo é o dobro daqueles censurados no parlamentarismo. Ou seja, tanto a função de sobrevivência quanto a função de risco indicam que as democracias parlamentaristas são mais longevas do que as presidencialistas.

CONCLUSÃO

A análise de sobrevivência ainda é muito pouco utilizada na pesquisa empírica em Ciências Sociais, principalmente quando comparada à análise de regressão. Este artigo buscou ajudar na disseminação da técnica a partir de uma abordagem intuitiva

Inicialmente, apresentamos os principais conceitos e as especificidades inerentes à análise de sobrevivência. Diferentemente das ferramentas estatísticas mais comumente utilizadas pelos pesquisadores da área, o estimador de Kaplan-Meier não está preocupado com o evento em si, mas com o tempo até a ocorrência de um determinado fenômeno.

Em seguida, apresentamos o passo a passo da implementação computacional a partir do SPSS. Reportamos as tabelas de sobrevivência, além da visualização gráfica de tais informações. Ainda reproduzimos o teste de significância mais utilizado, o log-rank. Todas as rotinas computacionais foram reportadas com o objetivo de aumentar a transparência e garantir a replicabilidade dos resultados. Com este artigo esperamos difundir a utilização da análise de sobrevivência em Ciência Política e contribuir com o avanço da literatura sobre metodologia política no Brasil.

REFERÊNCIAS

- AGRESTI, Alan e FINLAY, Barbara. (2012). Métodos Estatísticos para as Ciências Sociais. Porto Alegre: Penso.
- AKRITAS, Michael G. (2004). Nonparametric Survival Analysis. In: Statistical Science, vol. 19, nº 4, p. 615-623.
- ALLISON, Paul D. (1985). Survival Analysis of Backward Recurrence Times. In: Journal of the American Statistical Association, vol. 80, nº 390, p. 315-322.
- ANTIGO, Mariângela Furlan e MACHADO, Aná Flávia. (2006). Transições e Duração do Desemprego: uma revisão da literatura com novas evidências para Belo Horizonte. In: Nova Economia, vol. 16, nº 3, p. 375-406.
- BERNHARD, Michael; REENOCK, Christopher; NORDSTROM, Timothy. (2003). Economic Performance and Survival in New Democracies: is there a honeymoon effect? In: Comparative Political Studies, vol. 36, nº 4, p. 404-431.
- BLAND, J Martin e ALTMAN, Douglas G. (2004). The Logrank Test. In: BMJ, vol. 328, p. 1073.
- BOTELHO, Francisco; SILVA, Carlos; e CRUZ, Francisco. (2009). Epidemiologia Explicada: análise de sobrevivência. In: Acta Urológica, vol. 26, nº4, p. 33-38.
- BOX-STEFFENSMEIER, Janet e JONES, Bradford S. (1997). Time Is of the Essence: Event History Models in Political Science. In: American Journal of Political Science, vol. 41, nº 4, p. 1414-1461.
- _____. (2004). Event History Modeling: a guide for social scientists. Cambridge University Press.
- BOX-STEFFENSMEIER, Janet e ZORN, Christopher. (2002). Duration Model of Repeated Events. In: The Journal of Politics, vol. 64, nº 4, 1069-1094.
- BUENO DE MESQUITA, Bruce e SIVERSON, Randolph M. (1995). War and the Survival of Political Leaders: a comparative study of regime types and political accountability. In: The American Political Science Review, vol. 89, nº 4, p. 841-855.
- CHEIBUB, José Antonio e LIMONGI, Fernando. (2002). Democratic Institutions and Regime Survival: parliamentary and presidential democracy reconsidered. In: Annual Review of Political Science.
- COÊLHO, Denilson Bandeira. (2012). Political Competition and the Diffusion of Conditional Cash Transfers in Brazil. In: Brazilian Political Science Review, vol. 6, nº 2, p.56-87.
- COÊLHO, Denilson Bandeira; CAVALCANTE, Pedro; TURGEON, Mathieu. (2016). In: Revista de Sociologia e Política, vol. 24, nº 58, p. 145-165.
- COLLET, D. (1994). Modelling Survival Data in Medical Research. Chapman & Hall.
- COLOSIMO, Enrico A. e GIOLO, Suely R. (2006). Análise de Sobrevivência Aplicada. Editora Edgard Blucher.

- DAVIS, T. P. e LAWRENCE, A. J. (1989). The Likelihood for Competing Risk Survival Analysis. In: *Scandinavian Journal of Statistics*, vol. 16, nº 1, p. 23-28.
- EVERITT, B. S.; e SKRONDAL, A. (2010). *The Cambridge Dictionary of Statistics*, 4th ed. Cambridge University Press.
- FÁVERO, Luiz Paulo. (2014). *Métodos Quantitativos com Stata*. Rio de Janeiro: Elsevier.
- GASIOROWSKI, Mark J. (1995). Economic Crisis and Political Regime Change: an event history analysis. In: *The American Political Science Review*, vol. 89, nº 4, p. 882-897.
- GUO, Shenyang. (2010). *Survival Analysis*. Oxford University Press (Pocket guide to social work research methods).
- HIROI, Taeko e OMORI, Sawa. (2013). Causes and Triggers of Coups d'État: an event history analysis. In: *Politics and Policy*, vol. 41, nº 1, p. 39-64.
- HOSMER, David W.; LEMESHOW, Stanley; MAY, Susanne. (2008). *Applied Survival Analysis: regression modeling of time-to-event data*. New Jersey: Wiley-Interscience.
- KING, Gary; ALT, James E.; BURNS, Nancy Elizabeth; LAVER, Michael. (1990). A Unified Model of Cabinet Dissolution in Parliamentary Democracies. In: *American Journal of Political Science*, vol. 34, nº 3, p. 846-871.
- KLEINBAUM, David G. e KLEIN, Mitchel. (2005). *Survival Analysis: a self-learning text*. Springer.
- KNUST, Carl Henrik e NYGARD, Havard Møkleiv. (2015). Institutional Characteristics and Regime Survival: why are semi-democracies less durable than autocracies and democracies? In: *American Journal of Political Science*, vol. 59, nº 3, p. 656-670.
- KRUEGER, James S. e LEWIS-BECK, Michael S. (2008). Is OLS Dead? In: *The Political Methodologist*, vol. 15, nº 2, p. 2-4.
- LIJPHART, Arend. (2008). *Modelos de Democracia: desempenho e padrões de governo em 36 países*. Rio de Janeiro: Civilização Brasileira.
- LINZ, Juan. (1990). The Perils of Presidentialism. In: *Journal of Democracy*, vol. 1, nº 1, p. 51-69.
- LINZ, Juan e VALENZUELA, Arturo (ed.). (1994). *The Failure of Presidential Democracy*, vol. 1. John Hopkins University Press.
- MORITA, June G.; LEE, Thomas W.; MOWDAY, Richard T. (1993). The Regression-Analog to Survival Analysis: a selected application to turnover research. In: *The Academy of Management Journal*, vol. 36, nº 6, p. 1430-1464.
- MAINWARING, Scott e PÉREZ-LIÑÁN, Aníbal. (2013). *Democracies and Dictatorships in Latin America: emergence, survival, and fall*. Cambridge University Press.
- MARTÍNEZ, Christopher A. (2015). Presidential Survival in South America: rethinking the role of democracy. In: *International Political Science Review*, p. 1-16.
- NEGRETTO, Gabriel L. (2012). Replacing and Amending Constitutions: the logic of constitutional change in Latin America. In: *Law & Society Review*, vol. 46, nº 4, p. 749-779.

OSTRANDER, Ian. (2015). The Logic of Collective Inaction: senatorial delay in executive nomination. In: American Journal of Political Science, vol. 00, nº 0, p.1-10.

SAMBATH, Kripasankar Anandan; RAMANUJAM, Ramanan; CHINNAIYAN, Ponnujara. (2013). Survival Analysis: Kaplan-Meier and life table estimates for time to event clinical trial tuberculosis data. In: Concepts in Pure and Applied Science, vol. 1, nº 1, p. 17-21.

SCHMIDT, Peter e WITTE, Ann Dryden. (1989). Predicting Criminal Recidivism using 'split population' survival models. In: Journal of Econometrics, nº 40, p. 141-159.

SLOVIK, Milan. (2008). Authoritarian Reversals and Democratic Consolidation. In: American Political Science Review, vol. 102, nº 2, p. 153-168.

_____. (2015). Which Democracies Will Last? Coups, Incumbent Takeovers, and the Dynamic of Democratic Consolidation. In: British Journal of Political Science, vol. 45, nº 4, p. 715-738.

ZHOU, Mai. (2001). Understanding the Cox Regression Model with Time-Change Covariates. In: The American Statistician, vol. 55, nº 2, p. 153-155.

ANEXOS

Quadro 2 – Banco de dados

Casos	TempoSob	Censura	Grupo
1	50	0	0
2	50	0	0
3	50	0	0
4	48	1	0
5	26	1	0
6	34	1	0
7	19	1	0
8	49	1	0
9	50	0	0
10	50	0	0
11	39	1	0
12	41	1	0
13	29	1	0
14	44	1	0
15	50	0	0
16	50	0	0
17	37	1	0
18	50	0	0
19	50	0	0
20	50	0	0
21	3	1	1
22	7	1	1
23	50	0	1

24	10	1	1
25	12	1	1
26	41	1	1
27	50	0	1
28	16	1	1
29	18	1	1
30	21	1	1
31	44	1	1
32	50	0	1
33	24	1	1
34	28	1	1
35	50	0	1
36	30	1	1
37	33	1	1
38	50	0	1
39	35	1	1
40	38	1	1

Fonte: elaboração dos autores.



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



Departamento de
Ciência Política

Programa de Pós-Graduação
em Ciência Política

