

A IMPORTÂNCIA DA ANÁLISE DE SOBREVIDA PARA OS ESTUDOS EPIDEMIOLÓGICOS E DE SAÚDE

Marcos Manoel Honorato¹
Samuel Oliveira de Amorim²
Lorena Guimarães Ferreira Honorato³
Sheyla Mara Silva de Oliveira⁴
Renata Maria de Carvalho Cremaschi⁵
Fernando Morgadinho Santos Coelho⁶

RESUMO: Os estudos epidemiológicos e de saúde são muito versáteis e exigem metodologias que se adequem aos propósitos de cada pesquisa. A análise de sobrevida é um conjunto de procedimentos metodológicos que podem ser de grande utilidade para se avaliar desfechos e fatores a ele relacionados ao longo do tempo. Embora seja mais utilizada em estudos oncológicos, destacamos neste estudo a utilização em outras áreas, como economia e educação. Através de técnicas paramétricas e não paramétricas, torna-se possível mensurar a ocorrência de evento e o impacto das variáveis explicativas para o desfecho. As curvas de Kaplan-Meier são de fácil compreensão e os testes de Log-rank, Tarone-Ware e Gehan-Breslow são importantes para confirmar se existe diferença estatisticamente significativa entre grupos analisados. Abordamos também as medidas de análise de risco, como *hazard rate* e *hazard rate ratio*, que são importantes para quantificar a probabilidade de ocorrência de eventos em diferentes grupos. Discutimos os modelos de Cox e de Weibull, ferramentas úteis para avaliar o papel de um conjunto de variáveis explicativas no desfecho no tempo. O artigo é um ensaio teórico que explica de forma clara e objetiva as bases teóricas para a utilização dessa metodologia e sugere uma maior difusão sobre o assunto no ensino de graduação e pós-graduação.

Palavras-chave: Epidemiologia e bioestatística. Saúde. Análise de sobrevida.

1880

ABSTRACT: Epidemiological and health studies are highly versatile and require methodologies that suit the purposes of each research. Survival analysis is a set of methodological procedures that can be of great utility to assess outcomes and related factors over time. Although more commonly used in oncological studies, we highlight in this study its utilization in other areas such as economics and education. Through parametric and non-parametric techniques, it becomes possible to measure event occurrence and the impact of explanatory variables on the outcome. Kaplan-Meier curves are easily understandable, and Log-rank, Tarone-Ware, and Gehan-Breslow tests are important to confirm if there is a statistically significant difference between analyzed groups. We also address risk analysis measures such as hazard rate and hazard rate ratio, which are important for quantifying the probability of event occurrence in different groups. We discuss Cox and Weibull models, useful tools for evaluating the role of a set of explanatory variables in the outcome over time. The article is a theoretical essay that clearly and objectively explains the theoretical foundations for the use of this methodology and suggests greater dissemination of the subject in undergraduate and graduate education.

Keywords: Epidemiology and biostatistics. Health. Survival analysis.

¹ Médico neurologista, Doutor em Ciências (Programa de Pós-graduação em Neurologia-Neurociências da UNIFESP), docente do Curso de Medicina da Universidade do Estado do Pará (UEPA).

² Acadêmico de medicina pela Universidade do Estado do Pará (UEPA).

³ Assistente social, Mestre em Promoção à Saúde, Universidade Luterana do Brasil (ULBRA).

⁴ Enfermeira, Doutora em Ciências, Universidade do Estado do Pará (UEPA).

⁵ Médica neurologista, Doutora em Ciências (Programa de Pós-graduação em Psicobiologia da UNIFESP).

⁶ Médico neurologista, Livre-docente do Departamento de Neurologia e Neurocirurgia da UNIFESP.

INTRODUÇÃO

A Estatística é uma ciência complexa que se fundamenta na necessidade do homem de analisar dados de uma determinada realidade e a partir deles inferir significados que os tornem úteis para a Sociedade. Os dados transformam-se, através da estatística, em informações capazes de ajudar a compreender ou mudar essa realidade, seja no campo da Economia, na Educação ou na Saúde (Ignácio, 2010).

A Bioestatística é uma vertente que busca, de uma maneira geral, interpretar dados relacionados à saúde, auxiliando cientificamente a Epidemiologia, a qual estuda a forma como o processo saúde-doença está caracterizado em determinado cenário geográfico e populacional. Através da estatística descritiva geram-se dados que expressam numericamente a ocorrência de um fenômeno epidemiológico, como uma doença ou um agravamento à saúde. Já a estatística inferencial tem como objetivo identificar fatores de risco ou consequências relacionados a esse fenômeno ou agravamento, quantificando sua relevância. Todos esses processos devem servir, afinal, para que seja tomada a melhor decisão individual ou coletiva diante da realidade estudada (Ayres, 2011).

Nesse contexto, a análise de sobrevivência é definida como um conjunto de procedimentos e técnicas estatísticas usados para avaliar o tempo necessário para que aconteça um determinado evento em uma amostra, e também estabelecer a influência de variáveis categóricas (independentes) para que esse evento (desfecho) ocorra ao longo de um período de tempo (Miot, 2017). Ela pode ser realizada na maioria dos softwares estatísticos disponíveis no mercado, porém sua utilização exige alguns conhecimentos específicos acerca dos significados dos parâmetros e resultados obtidos (Ruhe, 2016).

1881

COMPREENDENDO A ANÁLISE DE SOBREVIVÊNCIA

Embora essa metodologia seja tradicionalmente utilizada em estudos oncológicos para analisar a mortalidade e os efeitos dos tratamentos ou procedimentos, pode ser interessante em pesquisas longitudinais onde haja um desfecho binário bem definido (variável dependente), como surto versus não surto, cura versus não cura, complicação versus não complicação etc. Por isso, alguns autores sugerem que o termo “análise do tempo para um evento” é mais apropriado (George et al., 2014; Miot, 2017). Também é bastante praticada na indústria e economia, onde pode ajudar a descrever o tempo para uma máquina ou peça parar de funcionar, correspondendo à chamada “taxa de falha” (David & Kleinbaum, 2016).

O processo estatístico gera uma função densidade de probabilidade $f(t)$, uma função de sobrevivência $S(t)$ e uma função de risco $h(t)$, expressas na figura 1 (David & Kleinbaum, 2016).

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

$$S(t) = P(T \geq t) = 1 - \int_0^t f(u) du, \text{ para todo } t > 0,$$

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

Figura 1 – Funções de densidade de probabilidade $f(t)$, sobrevivência $S(t)$ e risco $h(t)$ para análise de sobrevivência pelas estimativas de Kaplan-Meier (David & Kleinbaum, 2016).

Esse tipo de análise permite que os sujeitos da amostra sejam incluídos no estudo em tempos diferentes (Prinça et al., 2010). Também admite sem que haja prejuízo do poder estatístico a perda de informações decorrente da interrupção de seguimento (desistência, mudança de domicílio) ou não ocorrência do evento, às vezes pelo fim do período do estudo, o que é chamado de “censura” (David & Kleinbaum, 2016; Sainani, 2016). Também possibilita a avaliação de variados endpoints em um mesmo estudo, como recidiva, cura, morte, efeito colateral (Kartsonaki, 2016).

1882

A curva de sobrevivência pode ser obtida por meio de estimador de Kaplan-Meier. É um método não paramétrico que permite análise de amostras pequenas e sem distribuição normal (Kartsonaki, 2016). Essa curva é uma função da sobrevivência e se caracteriza por uma série de mudanças em degraus, onde cada degrau sinaliza o evento em um ou mais indivíduos, partindo sempre de 100% dos indivíduos (eixo y) (David & Kleinbaum, 2016).

Permite-se calcular o tempo mediano de sobrevivência (ou mediana de sobrevida), que é o tempo em que ocorre o desfecho em 50% da amostra como um todo e depois na presença ou ausência de cada variável explicativa (Sainani, 2016).

A figura 2 mostra um exemplo de gráfico elaborado a partir do estimador de Kaplan-Meier para um evento em um tempo de observação de 24 meses. Observe que o tempo mediano nessa amostra hipotética foi de 6 meses, ou seja naquele momento 50% dos indivíduos tinham atingido o evento. É possível perceber que não houve mudança na curva após 12 meses, indicando que a probabilidade de ocorrência do evento passou a ser improvável a partir de então.

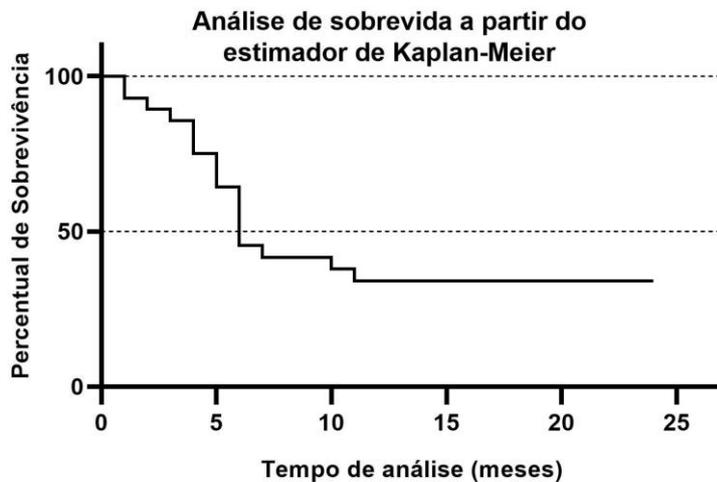


Figura 2 – Exemplo hipotético de um gráfico de análise de sobrevivência a partir do estimador de Kaplan-Meier.

Na presença de variáveis regressoras categóricas pode-se comparar o efeito das categorias ao longo do tempo. Quando as curvas relativas aos grupos a serem comparados não se cruzam, é provável que exista diferença estatística entre os grupos (Prinjal et al., 2010).

A figura 3 mostra um exemplo de curvas obtidas pelo estimador de Kaplan-Meier para uma amostra hipotética, onde a ocorrência de um evento (morte, por exemplo) é analisada em relação à variável sexo. Observe que as curvas dos dois subgrupos estão sempre próximas e se cruzam ou se superpõem várias vezes. Isso sugere que não há diferença significativa entre ser do sexo masculino ou feminino quanto ao risco de ocorrência do desfecho ao longo do tempo.

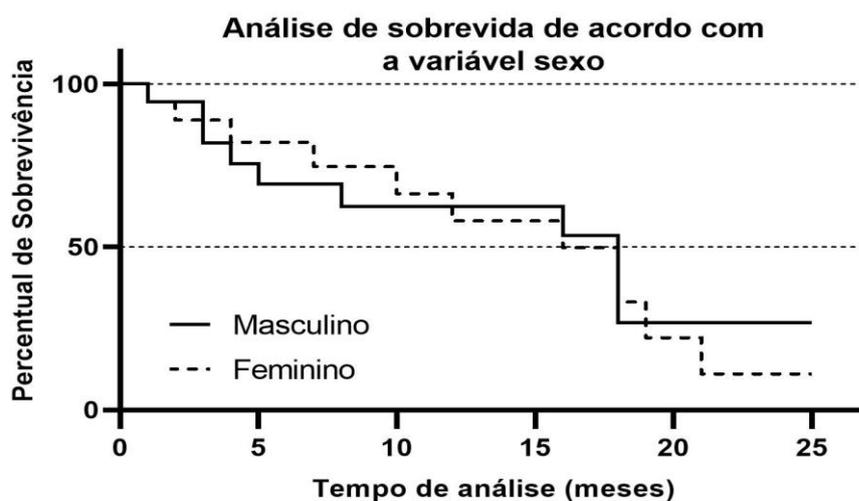


Figura 3 - Curva estimada por Kaplan-Meier para o evento em relação à variável sexo ao longo de 24 meses.

A figura 4 mostra um exemplo de curvas obtidas pelo estimador de Kaplan-Meier para uma amostra hipotética, onde a ocorrência de um evento (morte, por exemplo) é analisada em relação à variável tratamento. Observe que as curvas dos dois subgrupos são bem divergentes e não se cruzam. Isso sugere haver diferença significativa entre receber medicamento A ou B quanto ao risco de ocorrência do desfecho (morte, por exemplo) ao longo do tempo, sendo que o tratamento com a droga B seria mais favorável, o que poderia ajudar na tomada de decisão em outra amostra com as mesmas características.

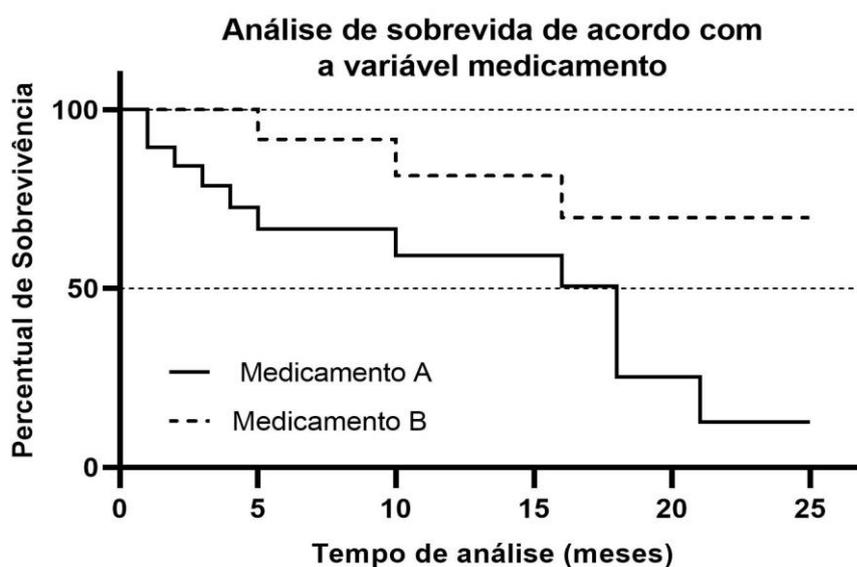


Figura 4 - Curva estimada por Kaplan-Meier para o evento em relação à variável tratamento ao longo de 24 meses.

TESTES DE HIPÓTESES EM ANÁLISE DE SOBREVIDA

Além da curva de Kaplan-Meier, testes estatísticos de hipóteses, como Log-rank, Tarone-Ware e Gehan-Breslow, são úteis e mais eficientes do que um teste de correlação simples (chi quadrado de Pearson ou exato de Fisher, por exemplo) para confirmar se a diferença visual observada através do gráfico tem significância estatística. Isso se dá porque tanto o fator tempo quanto os dados censurados são valorizados (Miot, 2017).

O teste de Gehan-Breslow avalia a significância em um momento inicial no tempo de observação, quando a amostra é maior. O teste de Tarone-Ware é mais adequado para calcular se há diferença na parte intermediária do tempo de observação. O teste de Log-rank é considerado mais poderoso porque serve para avaliar o nível de significância estatística ao longo de todo o tempo de observação e sua sensibilidade é superior aos outros testes nos segmentos final (Miot, 2017).

MEDIDAS DE RISCO EM ANÁLISE DE SOBREVIDA

Uma vez que exista diferença estatística entre dois grupos em uma análise de sobrevivida, ela pode ser calculada sob forma de “*hazard rate*”, que sinaliza qual a probabilidade dos indivíduos de cada grupo sofrerem o evento em determinado instante (Flynn, 2012). A medida de risco entre os dois grupos é chamada de “*hazard rate ratio*”, sendo na realidade o quociente dos *hazard rate* de cada grupo (Sainani, 2016).

Por exemplo, se analisarmos novamente o exemplo da figura 4, a diferença entre os grupos seria confirmada através dos testes de hipóteses ($p = 0,01$) e o cálculo do Hazard rate ratio seria 4,9, o que indica que o grupo de indivíduos submetidos ao tratamento B têm quase 5 vezes menor chance de atingirem o evento (morte) ao longo do tempo de observação.

MODELOS ESTATÍSTICOS

A modelagem estatística serve para avaliar o poder de explicação de variáveis regressoras em conjunto (Flynn, 2012). No contexto da análise de sobrevivência, o método semi-paramétrico de Cox é o mais conhecido, mas existem modelos paramétricos, sendo Weibull um deles (Benitez-Parejo et al., 2011),

O modelo de Cox é bastante flexível. Isto facilita seu uso em pesquisas na área de saúde. A proporcionalidade das taxas de falhas configura-se como uma exigência para sua utilização, o que pode ser sugerida através das curvas de sobrevivida de Kaplan-Meier das variáveis categóricas (curvas que não se aproximam nem se cruzam) e do teste de Log-rank (Lin & Zelterman, 2002). Para verificar se há violação da suposição das taxas de falhas proporcionais, podem ser elaborados gráficos de resíduos padronizados de Schoenfeld no tempo para cada covariável ajustada. Espera-se que não haja uma grande dispersão dos pontos (Fitrianto & Jiin, 2013; Zhang, 2017).

A razão da taxa de falha é a medida de comparação entre os grupos no modelo de Cox e expressa quantas vezes o evento ocorre a mais em um grupo do que no outro ao longo do tempo de observação (Collett, 2015; George et al., 2014).

O modelo de Weibull, embora menos divulgado e de compreensão mais complexa, pode ser mais fidedigno em evidenciar o papel de múltiplas variáveis em um desfecho no tempo, sendo definida pelas funções descritas na figura 5 (Collett, 2015).

$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}, t \geq 0,$	(densidade)
$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\},$	(sobrevivência)
$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \alpha, \gamma > 0$	(risco).

Figura 5 – Funções de densidade de probabilidade $f(t)$, sobrevivência $S(t)$ e risco $h(t)$ no modelo de Weibull (Collett, 2015)

No modelo de Weibull, a diferença entre os grupos é expressa na forma de razão de tempo mediano, o qual estabelece quantas vezes o tempo mediano para o evento ocorrer é maior em um grupo comparado ao outro (Prinja et al., 2010).

Quando se faz um estudo de análise de sobrevida, vários modelos estatísticos podem ser elaborados combinando diferentes covariáveis. O critério de informação de Akaike (AIC) é um teste estatístico amplamente usado para comparar diferentes modelos possíveis e determinar qual deles se ajusta melhor aos dados (George et al., 2014).

A modelagem estatística pelo método semiparamétrico de Cox gera uma curva logarítmica para cada variável, enquanto o método paramétrico de Weibull resulta em uma função exponencial capaz de estimar a ocorrência do desfecho no tempo com as covariáveis selecionadas (Mittal et al., 2014). Assim é possível identificar fatores preditivos e, simultaneamente, excluir confundidores (ElHafeez et al., 2012; Saha- Chaudhuri & Juwara, 2021).

CONSIDERAÇÕES FINAIS

As técnicas de análise de sobrevida se constituem em uma valiosa ferramenta no campo da bioestatística e podem ser úteis tanto em estudos retrospectivos quanto em coortes prospectivas, onde permitem observar a ocorrência de um evento no tempo e também analisar o impacto de variáveis individuais e em conjunto no desfecho estudado, auxiliando na tomada de decisão mais assertiva.

Sugere-se que essa metodologia seja mais divulgada e se torne mais utilizada em pesquisas de cursos de graduação e pós-graduação.

REFERÊNCIAS BIBLIOGRÁFICAS

AYRES, J. R. d. C. M. Desenvolvimento histórico-epistemológico da Epidemiologia e do conceito de risco. *Cadernos de saúde pública*, 27, 1301-1311, 2011.

BENITEZ-PAREJO, N.; DEL ÁGUILA, M. R.; PEREZ-VICENTE, S.. Survival analysis and Cox regression. *Allergologia et immunopathologia*, 39(6), 362-373, 2011.

COLLETT, D. Modelling survival data in medical research. CRC press, 2015.

DAVID, G.; KLEINBAUM, K. Survival analysis: a self-learning text. Springer-Verlag New York, 2016.

ELHAFEEZ, S. A. et al.. An overview on standard statistical methods for assessing exposure-outcome link in survival analysis (Part II): the Kaplan-Meier analysis and the Cox regression method. *Aging clinical and experimental research*, 24(3), 203-206, 2012.

FITRIANTO, A.; JIIN, R. L. T. Several types of residuals in cox regression model: an empirical study. *Int J Math Anal*, 7, 2645-2654, 2013.

FLYNN, R. Survival analysis. *Journal of Clinical Nursing*, 21(19pt20), 2789-2797, 2012.

GEORGE, B.; SEALS, S.; ABAN, I. Survival analysis and regression models. *Journal of nuclear cardiology*, 21(4), 686-694, 2014.

IGNÁCIO, S. A. Importância da estatística para o processo de conhecimento e tomada de decisão. *Revista Paranaense de Desenvolvimento-RPD*(118), 175-192, 2010. 1887

KARTSONAKI, C. Survival analysis. *Diagnostic Histopathology*, 22(7), 263-270, 2016.

Lin, H.; ZELTERMAN, D. Modeling survival data: Extending the cox model, 2002.

MIOT, H. A. (2017). Análise de sobrevivência em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, 16(4), 267-269, 2017.

MITTAL, S. et al. A. High-dimensional, massive sample- size Cox proportional hazards regression for survival analysis. *Biostatistics*, 15(2), 207- 221, 2014.

PRINJA, S.; GUPTA, N.,; VERMA, R. Censoring in clinical trials: review of survival analysis techniques. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 35(2), 217, 2010.

RUHE, C. Estimating survival functions after stcox with time-varying coefficients. *The Stata Journal*, 16(4), 867-879, 2016.

SAHA-CHAUDHURI, P.; JUWARA, L. Survival analysis under the Cox proportional hazards model with pooled covariates. *Statistics in Medicine*, 40(4), 998-1020, 2021.

SAINANI, K. Introduction to Survival Analysis. *PM & R: the Journal of Injury, Function, and Rehabilitation*, 8(6), 580-585, 2016.

ZHANG, Z. Survival analysis in the presence of competing risks. *Annals of translational medicine*, 5(3), 2017.