

**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA**

**PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA (PIBIC)**

**RELATÓRIO TÉCNICO**

**MEDIAÇÃO EM ANÁLISE DE SOBREVIVÊNCIA**

**Michelle Pereira Vale dos Passos**

*Bolsista do CNPq*

**Leila Denise Alves Ferreira Amorim**

*Orientadora e Docente do Departamento de Estatística/UFBA*

Salvador – BA

2020

## **Agradecimentos**

Michelle Pereira Valle dos Passos teve bolsa de iniciação científica do CNPq, vinculada ao PIBIC-UFBA 2019-2020. O projeto ao qual se vinculou o projeto de iniciação científica foi aprovado e desenvolvido no Departamento de Estatística, sem financiamento, em parceria com Prof. Marcelo Magalhães Taddeo.

## LISTA DE ILUSTRAÇÕES

- Figura 1: Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para o tempo de aleitamento materno segundo a covariável dificuldade para amamentar nos primeiros dias pós-parto ..... pág.14
- Figura 2: Resíduos padronizados estimados pelo método de Kaplan-Meier para avaliação do modelo log-normal ..... pág.16
- Figura 03: Resíduos padronizados de Schoenfeld do modelo de Cox para análise dos dados de aleitamento materno ..... pág.18
- Figura 4: Diagrama das relações estimadas na análise de mediação sobre a influência da mídia .... pág.24
- Figura 5: Diagrama das relações na análise de mediação sobre a higiene do sono ..... pág.27
- Figura 6: Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para avaliação do modelo Weibull ..... pág.28
- Figura 7: Resíduos padronizados de Schoenfeld do modelo de Cox para análise de mediação sobre a influência da higiene do sono ..... pág.29
- Figura 8: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na ausência de censura e  $\hat{a}\hat{b} = 0,5$ , utilizando os modelos AFT e de Cox ..... pág.34
- Figura 9: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 50% de censura e  $\hat{a}\hat{b} = 0,5$ , utilizando os modelos Weibull e de Cox ..... pág.35
- Figura 10: Probabilidade de cobertura percentual (CP%) do intervalo de confiança construído baseado no erro-padrão estimado, com  $\alpha = 0,2$  ..... pág.36
- Figura 11: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 25% de censura e  $\hat{a}\hat{b} = 0,5$ , utilizando os modelos Weibull e de Cox ..... pág.40
- Figura 12: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 70% de censura e  $\hat{a}\hat{b} = 0,5$ , utilizando os modelos Weibull e de Cox ..... pág.41
- Figura 13: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na ausência de censura e  $\hat{a}\hat{b} = 0,1$ , utilizando os modelos Weibull e de Cox ..... pág.41
- Figura 14: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 25% de censura e  $\hat{a}\hat{b} = 0,1$ , utilizando os modelos Weibull e de Cox ..... pág.42
- Figura 15: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 50% de censura e  $\hat{a}\hat{b} = 0,1$ , utilizando os modelos Weibull e de Cox ..... pág.42

Figura 16: Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 70% de censura e  $\hat{a}\hat{b} = 0,1$ , utilizando os modelos Weibull e de Cox ..... pág.43

Figura 17: Probabilidade de cobertura percentual (CP%) do intervalo de confiança construído baseado no erro-padrão estimado, com  $a = 10$  ..... pág.43

## LISTA DE TABELAS

Tabela 1: Testes logrank e de Wilcoxon para testar a igualdade das curvas de sobrevivência segundo as covariáveis do estudo sobre aleitamento materno .....	pág.15
Tabela 2: Teste da razão de verossimilhanças para avaliar a adequação dos modelos de regressão Weibull e log-normal .....	pág.16
Tabela 3: Estimativas dos parâmetros do modelo de regressão log-normal ajustado aos dados de aleitamento materno .....	pág.16
Tabela 4: Estimativas dos parâmetros dos modelo de regressão Weibull e Cox ajustados aos dados de aleitamento materno .....	pág.17
Tabela 5: Testes da proporcionalidade dos riscos usando resíduos de Schoenfeld obtido no modelo de Cox.....	pág.18
Tabela 6: Estimativa média dos parâmetros dos modelos Weibull e de Cox e estimativas segundo a relação teórica entre os dois modelos para o cenário sem censura .....	pág.20
Tabela 7: Estimativa dos parâmetros da análise de mediação sobre a influência da mídia.....	pág.23
Tabela 8: Intervalos de 95% de confiança para o efeito mediado usando diferentes propostas para estimação dos erros-padrão na análise de mediação sobre a influência da mídia .....	pág.24
Tabela 9: Testes da proporcionalidade dos riscos usando resíduos de Schoenfeld obtido no modelo de Cox.....	pág.29
Tabela 10: Estimativa dos parâmetros dos modelos AFT e Cox da análise de mediação sobre a influência da higiene do sono em sua duração .....	pág.30
Tabela 11: Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados sem censura e $\alpha=10$ .....	pág.32
Tabela 12: Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados com censura e $\alpha=10$ .....	pág.33
Tabela 13: Estimativa média dos parâmetros dos modelos Weibull e de Cox e estimativas segundo a relação teórica entre os dois modelos para o cenário com censura .....	pág.38
Tabela 14: Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados sem censura e $\alpha=0,2$ .....	pág.39
Tabela 15: Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados com censura e $\alpha=0,2$ .....	pág.39

## SUMÁRIO

1	Análise de Sobrevida	pág.08
1.1	Aplicação 1: Aleitamento materno	pág.14
2	Mediação	pág.21
2.1	Aplicação 2: Mediação presumida da mídia	pág.23
3	Mediação em Análise de Sobrevida	pág.25
3.1	Aplicação 3: Mediação da higiene do sono	pág.26
3.2	Comparação entre o modelo de tempo de vida acelerado e o modelo de Cox	pág.30
	Referências Bibliográficas	pág.37
	Apêndice 1	pág.38
	Apêndice 2	pág.44
	Apêndice 3	pág.51

## RESUMO

Este relatório técnico está vinculado ao Plano de Trabalho Mediação em Análise de Sobrevivência do PIBIC/UFBA 2019-2020 e inclui o estudo e sumarização de tópicos relacionados à análise de sobrevivência, mediação clássica e mediação em análise de sobrevivência, incluindo a implementação desses métodos em softwares estatísticos, especificamente o software R versão 3.6.0. No presente relatório estão sumarizados os conceitos e metodologias fundamentais para entendimento de mediação em análise de sobrevivência. Foram implementados dois estudos de simulação, apresentados nas seções 1 e 3. O primeiro estudo de simulação objetivou verificar a relação entre o modelo de tempo de vida acelerado e o modelo de Cox, na ausência e presença de censuras, para diferentes tamanhos amostrais. Assim, verifica-se que as estimativas se afastam da relação preconizada entre os parâmetros dos dois modelos à medida que aumentam o tamanho amostral e o percentual de censura. O segundo estudo de simulação, por sua vez, comparou as estimativas de efeito indireto em dois modelos de mediação em análise de sobrevivência, variando percentual de censura, tamanho amostral e magnitude do efeito. A estimação dos efeitos indiretos foi feita pelos métodos do produto e da diferença utilizando os modelos de tempo de vida acelerado e o de Cox, na presença e ausência de censuras. As estimativas são similares utilizando os métodos do produto e da diferença dos coeficientes no modelo de tempo de vida acelerado, o que não é observado no modelo de Cox, sobretudo à medida que o tamanho amostral decresce. Neste relatório apresentamos também nas seções 1, 2 e 3 análises de dados conduzidas para examinar, respectivamente: o tempo máximo de duração do aleitamento materno desde o nascimento da criança até o desmame; como a importância do tema causa uma determinada reação mediada pela influência presumida da mídia; e como o conhecimento da importância e função do sono afeta sua duração, que é mediada pela influência da higiene do sono.

## 1. Análise de Sobrevivência

Em análise de sobrevivência a variável resposta é o tempo até que um evento de interesse ocorra. Este tempo é denominado tempo de falha e é composto por três elementos: tempo inicial, escala de medida e evento de interesse (falha). A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta (Colosimo e Giolo, 2006). Os dados de sobrevivência podem conter ou não a presença de truncamentos ou censuras. Estes truncamentos geralmente são confundidos com censuras, porém os truncamentos consistem em uma condição que exclui determinado indivíduo do estudo, já as censuras são observações parciais do tempo de falha.

As censuras são classificadas como:

**Censura tipo I:** Quando o estudo é terminado após um período pré-estabelecido de tempo.

**Censura tipo II:** Quando o estudo é terminado após a ocorrência do evento de interesse em um número pré-estabelecido de indivíduos.

**Censura do tipo aleatória:** Quando o paciente é retirado do estudo sem que a falha tenha ocorrido.

**Censura à direita:** Quando o tempo de ocorrência do evento de interesse está à direita do tempo registrado, ou seja, quando o evento de interesse ocorreu após o final do estudo.

**Censura à esquerda:** Quando o evento de interesse ocorreu antes do início do estudo ou do período de acompanhamento do indivíduo.

**Censura intervalar:** Quando o evento de interesse ocorreu em determinado intervalo de tempo, mas não se conhece o tempo exato.

Segundo Colosimo e Giolo (2006), uma representação simples do mecanismo de censura à direita é feita utilizando os pares  $(T_i, C_i)$ , para  $i = 1, \dots, n$ , que formam uma amostra aleatória de  $n$  pacientes, onde  $T$  é uma variável aleatória representando o tempo de falha de um paciente e  $C$  outra variável aleatória independente de  $T$ , representando o tempo de censura associado a este paciente. O que se observa para este paciente é, portanto,  $t = \min(T, C)$  e  $\delta = \begin{cases} 1, & \text{se } T \leq C \\ 0, & \text{se } T > C \end{cases}$ .

O tempo de sobrevivência é especificado pela variável aleatória  $T$  (maior ou igual à zero), em que  $T$  representa o tempo de falha. Podemos especificar a variável aleatória  $T$  por meio de algumas funções, tais como a **função de sobrevivência** que representa a probabilidade de uma observação não falhar até certo tempo  $t$  e pode ser escrita em termos probabilísticos através da fórmula:



$$S(t) = P(T \geq t)$$

Consequentemente a probabilidade de uma observação falhar até t é dada por:

$$F(t) = 1 - S(t).$$

A **função de taxa de falha** também chamada de **função de taxa de risco** representa a probabilidade instantânea da falha ocorrer em um intervalo de tempo muito pequeno e pode ser obtida através da fórmula:

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

Por fim, a **função de taxa de falha acumulada** não tem uma interpretação direta, mas pode ser útil na avaliação da bondade do ajuste de alguns modelos e é calculada através da fórmula:

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

Relações existentes entre essas funções facilitam seu cálculo, de modo que:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t))$$

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log S(t)$$

$$S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u) du\right\}.$$

Por conta da presença de censuras, para obter estimativas para a função de sobrevivência e a partir dela estimar outras estatísticas de interesse, pode-se considerar a utilização de técnicas não-paramétricas, tais como: Kaplan-Meier, Nelson-Aalen e Tabela de Vida. Neste projeto o estimador de Kaplan-Meier será utilizado por ser o estimador de máxima verossimilhança da função de sobrevivência (Colosimo e Giolo, 2006). O estimador de Kaplan-Meier, proposto por Kaplan e Meier (1958), é também chamado de estimador limite-produto, sendo definido como:

$$\hat{S}(t) = \prod_{j: t_j < t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j: t_j < t} \left( 1 - \frac{d_j}{n_j} \right)$$

em que  $t_1 < t_2 \dots < t_k$ , representam os k tempos distintos de falha,  $d_j$  o número de falhas em  $t_j$ ,  $j = 1, \dots, k$ , e  $n_j$  o número de indivíduos sob risco em  $t_j$ . O estimador de Kaplan-Meier possui ótimas propriedades, tais como: é não viciado para grandes amostras, fracamente consistente, converge assintoticamente para o processo

gaussiano e é o estimador de máxima verossimilhança da função de sobrevivência. É possível realizar a construção de intervalos de confiança e testar hipóteses para a função de sobrevivência. A expressão da variância assintótica do estimador de Kaplan-Meier é dada por:

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j: t_j < t} \frac{d_j}{n_j(n_j - d_j)}$$

Como  $\hat{S}(t)$ , para  $t$  fixo, tem distribuição assintótica Normal, pode-se obter um intervalo aproximado de  $100(1 - \alpha)\%$  de confiança para  $S(t)$  dado por:

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}(t))}$$

em que  $\alpha/2$  denota o  $\alpha/2$  - percentil da distribuição Normal padrão. Para valores extremos de  $t$ , este intervalo de confiança pode apresentar limite inferior negativo ou limite superior maior do que 1. Esse problema é resolvido utilizando uma transformação para a função de sobrevivência.

Muitas vezes as análises envolvendo os métodos não paramétricos em sobrevivência são realizadas para vários estratos populacionais, definidos por sexo, grupo etário, raça/cor, gravidade da doença, etc. Para realizar a comparação da função de sobrevivência entre grupos, pode-se considerar a utilização de testes não paramétricos, dentre os quais os mais conhecidos são o teste de Wilcoxon e log-rank. O teste mais utilizado em análise de sobrevivência é o teste de logrank e é apropriado quando a razão entre as taxas de falha dos grupos a serem comparados é aproximadamente constante. Esta propriedade é conhecida como riscos proporcionais. O teste de Wilcoxon, que também é muito usado na prática, utiliza uma ponderação na estatística do teste, em que o peso é igual ao número de indivíduos sob risco, sendo mais apropriado quando se deseja colocar mais relevância na parte inicial da escala do tempo (Colosimo e Giolo, 2006).

Além dos procedimentos não paramétricos, é possível a utilização de abordagem paramétrica para estimação em análise de sobrevivência. Os principais modelos probabilísticos utilizados incluem exponencial e Weibull, que vem se destacando por conta de sua adequação a várias situações práticas.

A distribuição Weibull foi proposta por Weibull (1939), e sua popularidade em aplicações práticas se deve ao fato dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função taxa de falha é monótona, podendo ser crescente, decresce ou constante (Colosimo e Giolo, 2006). Para uma variável aleatória  $T$  com distribuição Weibull tem-se a função de densidade da probabilidade dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} (t)^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}, \quad t \geq 0$$

em que  $\gamma$ , o parâmetro de forma, e  $\alpha$ , o de escala, são ambos positivos. O parâmetro  $\alpha$  tem a mesma unidade de medida de  $t$  e  $\gamma$  não tem unidade. Para esta distribuição, as funções de sobrevivência e de taxa de falha são, respectivamente,

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

e

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

para  $t, \alpha$  e  $\gamma \geq 0$ . Quando  $\gamma = 1$  tem-se a distribuição exponencial, sendo assim um caso particular da distribuição Weibull.

Para efetuar a escolha do “melhor” modelo que se ajusta os dados pode-se utilizar testes de hipóteses com modelos encaixados ou técnicas gráficas, tais como a comparação da função de sobrevivência do modelo proposto com o estimador de Kaplan-Meier, ou a linearização da função de sobrevivência (Colosimo e Giolo, 2006).

Após a realização da escolha do modelo probabilístico, os parâmetros do modelo são estimados pelo Método de Máxima Verossimilhança, que permite a incorporação das censuras e possui ótimas propriedades para grandes amostras. A estimação usando a função de verossimilhança a partir dos tipos de censura é apresentada a seguir:

**i) Censura do tipo I ou do tipo aleatório:**

$$L(\theta) = \prod_{i=1}^r f(t_i; \theta) \prod_{i=r+1}^n S(t_i; \theta)$$

ou equivalentemente por:

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

em que  $\delta_i$  é a variável indicadora de falha.

**ii) Censura do tipo II:**

$$L(\theta) \propto \prod_{i=1}^r f(t_i; \theta) \prod_{i=r+1}^n S(t_i; \theta)$$

No caso da distribuição Weibull, a função de verossimilhança é dada por:

$$L(\gamma, \alpha) = \prod_{i=1}^n \left[ \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \right]^{\delta_i} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}.$$

Não existem expressões fechadas para os estimadores de  $\gamma$  e  $\alpha$ . Por esta razão as estimativas de máxima verossimilhança para os parâmetros da distribuição Weibull são obtidas através de métodos numéricos como, por exemplo, o de Newton-Raphson (Colosimo e Giolo, 2006). Para a construção de intervalos de  $(1 - \alpha)\%$  de confiança é utilizada a distribuição assintótica do estimador de máxima verossimilhança  $\hat{\theta}$ . Para grandes amostras, essa propriedade estabelece, sob certas condições de regularidade, que  $\hat{\theta}$  tem distribuição Normal multivariada de média  $\theta$  e matriz de variância-covariância  $V(\hat{\theta})$ , ou seja,  $\hat{\theta} \sim N_k(\theta, \text{Var}(\hat{\theta}))$ , sendo  $k$  a dimensão de  $\theta$ . Sob certas condições de regularidade (Casella, 2002), a matriz de variância-covariância é aproximadamente o negativo da inversa da matriz de informação de Fisher:

$$\text{Var}(\hat{\theta}) \approx -[I_F(\theta)]^{-1},$$

em que  $I_F(\theta) = E \left[ \left( \frac{\partial \log L(\theta)}{\partial \theta} \right)^2 \right]$ .

De acordo com Colosimo e Giolo (2006), testes de hipóteses, como o de Wald, da Razão de Verossimilhanças e do escore, podem ser utilizados quando o objetivo consiste em testar hipóteses relacionadas ao vetor  $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$  de parâmetros ou a um subconjunto dele. O teste de Wald é geralmente o mais utilizado para testar hipóteses relativas a um único parâmetro  $\theta_j$ .

Usualmente existe interesse na inclusão de covariáveis na análise de sobrevivência para avaliar seu efeito sobre o tempo até a ocorrência do evento ou sobre sua taxa. A forma mais eficiente de acomodar o efeito dessas covariáveis é através da definição de um modelo de regressão apropriado para dados censurados, que pode ser paramétrico ou semiparamétrico. Neste relatório são apresentados dois modelos de regressão em sobrevivência: o modelo Weibull (paramétrico) e o modelo de Cox (semiparamétrico).

O modelo paramétrico com distribuição Weibull também é denominado modelo de tempo de vida acelerado (*AFT = accelerated failure time*, em inglês). Quando adequados ao ajuste dos dados, estes modelos são mais eficientes que os semiparamétricos, apesar de serem, em geral, menos flexíveis. Um parâmetro extra de escala ( $\sigma$ ) é incluído, como uma forma de generalizar o modelo, assim o modelo AFT é expresso por:

$$Y = \log(T) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \sigma v$$

com  $v = \log(\epsilon)$ , que segue uma distribuição do valor extremo padrão,  $\beta$  é o vetor de parâmetros associados às covariáveis, e  $\sigma = \frac{1}{\gamma}$  representa o parâmetro de escala. Neste caso, a função de sobrevivência para  $Y$  condicional a  $\mathbf{x}$  é expressa por

$$S(y|\mathbf{x}) = \exp\left\{-\frac{y - \mathbf{x}\beta}{\sigma}\right\}$$

e, para  $T$  condicional a  $\mathbf{x}$ , por

$$S(t|x) = \exp \left\{ - \left( \frac{t}{\exp\{x\beta\}} \right)^{\frac{1}{\sigma}} \right\}.$$

Para facilitar a interpretação dos resultados do modelo AFT, considera-se a razão de tempos médios entre dois grupos (covariável categórica) ou devido ao aumento de uma unidade da covariável (covariável quantitativa), que são definidos tomando-se o exponencial do correspondente parâmetro. Como exemplo, considere que  $x$  é uma variável binária ( $x=0$  ou  $1$ ), tal que:

$$\log T = \beta_0 + \beta_1 x + v$$

$$T = e^{\beta_0} e^{\beta_1 x} e^v$$

$$E[T|x] = e^{\beta_0} e^{\beta_1 x} E[e^v]$$

$$\frac{E[T|x = 1]}{E[T|x = 0]} = \frac{e^{\beta_0} e^{\beta_1} E[e^v]}{e^{\beta_0} E[e^v]} = e^{\beta_1}$$

A adequação do modelo ajustado é avaliada através de análise de resíduos, em que os resíduos de Cox-Snell (1968) e os resíduos padronizados são utilizados para examinar o ajuste global do modelo. Os resíduos de *martingale*, por sua vez, são usados para determinar a forma funcional (linear, quadrática, etc.) de uma covariável (em geral, contínua) e os *resíduos deviance* auxiliam no exame da acurácia do modelo para cada indivíduo sob estudo.

Os modelos semiparamétricos, que englobam o modelo de riscos proporcionais de Cox (Cox, 1972), são mais flexíveis e permitem facilmente a incorporação de covariáveis dependentes do tempo. O modelo pode ser representado pela forma:

$$\lambda(t) = \lambda_0(t) \exp\{x'\hat{\beta}\}$$

em que  $\hat{\beta}$  é o vetor de parâmetros associado às covariáveis. Para realizar a estimação dos coeficientes  $\hat{\beta}$ 's é necessária a utilização do método da máxima verossimilhança parcial. Para facilitar a interpretação dos coeficientes estimados considera-se a razão de riscos instantâneos no tempo  $t$ , que é calculada através da razão das taxas de falha de dois indivíduos. Para ilustrar este conceito, considere que tem-se apenas uma covariável binária  $x$  (0/1), tal que a função de taxa de falha do primeiro grupo será representada por  $\lambda_1(t)$  e a do segundo grupo por  $\lambda_2(t)$ . A razão destas duas funções de falha, também conhecida como risco relativo, é dada por  $\frac{\lambda_1(t)}{\lambda_2(t)} = \exp(\beta_1)$ , que é constante para todo tempo  $t$  de acompanhamento do estudo. A estimação no modelo de Cox é feita através do método da verossimilhança parcial (Cox, 1975).

Para verificar a qualidade geral do ajuste do modelo de Cox utilizam-se os resíduos de Cox-Snell. A avaliação da proporcionalidade dos riscos, que é uma suposição do

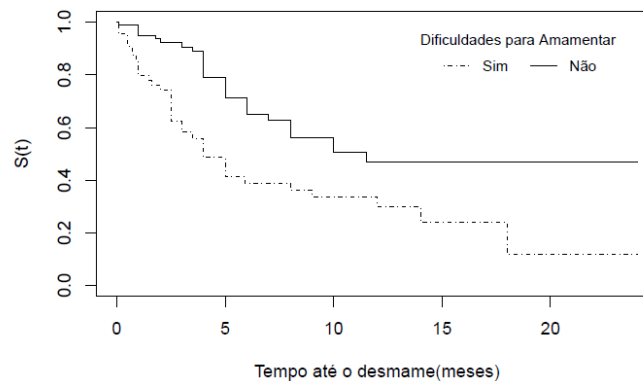
modelo de Cox, é feita através do resíduo de Schoenfeld. Avaliações podem também ser feitas através de métodos gráficos descritivos, que consistem nas seguintes etapas: (i) inicialmente dividir os dados em  $m$  estratos, usualmente de acordo com alguma variável; (ii) em seguida, estima-se  $\hat{\Lambda}_{0_j}(t)$  para cada estrato usando o estimador de Nelson-Aalen-Breslow. Se a suposição de proporcionalidade dos riscos for válida, as curvas do logaritmo de  $\hat{\Lambda}_{0_j}(t)$  versus  $t$  ou  $\log(t)$  devem apresentar diferenças aproximadamente constantes no tempo.

Nas duas subsecções seguintes apresentamos em (1.1) uma análise de dados com as metodologias para dados de sobrevivência descritas nesse relatório e em (1.2) um estudo de simulação para comparação das estimativas obtidas pelo modelo de tempo de vida acelerado e pelo modelo de Cox.

### 1.1 Aplicação 1: Aleitamento materno

Nesta secção ilustraremos os métodos descritos na secção 1 através de análise de dados, em que o evento de interesse é o tempo máximo de duração do aleitamento materno desde o nascimento da criança até o desmame (Colosimo e Giolo, 2006). O banco de dados possui informações de 153 crianças ( $n=153$ ), possui 56,7% dos dados censurados e é composto por 11 covariáveis, além da variável resposta e do indicador de censuras. As análises foram conduzidas no software R versão 3.6.0. A sintaxe das análises realizadas no R encontra-se no Apêndice 1 desse relatório.

Métodos não paramétricos foram utilizados para a realização da análise descritiva e exploratória. O método de Kaplan-Meier foi utilizado para estimação da função de sobrevivência segundo todas as covariáveis do estudo. A Figura 1 mostra que as mães que tiveram dificuldade para amamentar apresentam um tempo de desmame menor do que aquelas que não tiveram dificuldades. Mais de 50% das mulheres que não tiveram dificuldades para amamentar chegaram aos 10 meses, enquanto esse percentual foi menor que 40% para as mulheres que tiveram dificuldade para amamentar.



**Figura 1:** Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para o tempo de aleitamento materno segundo a covariável dificuldade para amamentar nos primeiros dias pós-parto.

Os testes logrank e Wilcoxon foram usados para a comparação entre as curvas de sobrevivência de acordo com as categorias de cada covariável do estudo. Os resultados para ambos os testes são apresentados na Tabela 1.

**Tabela 1:** Testes *logrank* e de Wilcoxon para testar a igualdade das curvas de sobrevivência segundo as covariáveis do estudo sobre aleitamento materno.

Covariável	Estatísticas de teste (valor $p$ )	
	<i>logrank</i>	Wilcoxon*
Experiência anterior de amamentação	03,95 (0,047)	05,70 (0,020)
Número de filhos vivos	02,60 (0,107)	02,40 (0,100)
Conceito materno sobre o tempo ideal de amamentação	06,15 (0,013)	07,90 (0,005)
Dificuldades para amamentar nos primeiros dias pós-parto	12,26 (0,001)	14,90 (0,670)
Tipo de serviço em que realizou o pré-natal	01,38 (0,241)	01,30 (0,300)
Recebeu exclusivamente leite materno na maternidade	07,47 (0,006)	07,00 (0,008)
A criança tem contato com o pai	01,84 (0,175)	01,50 (0,200)
Renda per capita	02,11 (0,146)	02,20 (0,100)
Peso ao nascimento	01,87 (0,171)	02,40 (0,100)
Tempo de separação mãe-filho pós-parto	02,60 (0,107)	01,60 (0,200)
Permanência no berçário	02,93 (0,087)	01,70 (0,200)

\*equivalente à modificação de Peto & Peto do teste de Gehan-Wilcoxon.

As covariáveis incluídas no modelo de regressão inicial para o tempo até o desmame foram escolhidas com base nos resultados da Tabela 1. O critério utilizado foi a apresentação de valores de  $p$  inferiores a 0,25 em pelo menos um dos testes de comparação das curvas de sobrevivência. Todas as covariáveis passaram por este critério. Após o ajuste do modelo Weibull foram mantidas no modelo apenas aquelas com significância estatística ( $p$ -valor < 10%). Deste modo, as covariáveis selecionadas foram: experiência anterior de amamentação, o conceito materno sobre o tempo ideal de amamentação, dificuldades para amamentar nos primeiros dias pós-parto e se recebeu exclusivamente leite materno na maternidade. Após a seleção das covariáveis para o modelo de regressão Weibull, comparou-se este modelo com aquele usando distribuição gama generalizada através do teste da razão de verossimilhanças. Procedimento similar foi adotado para avaliar a adequação do modelo de regressão log-normal (Tabela 2). As hipóteses utilizadas para testar a adequabilidade dos modelos de regressão Weibull e log-normal foram, respectivamente:

$H_0$ : O modelo Weibull é adequado

$H_0$ : O modelo log – normal é adequado

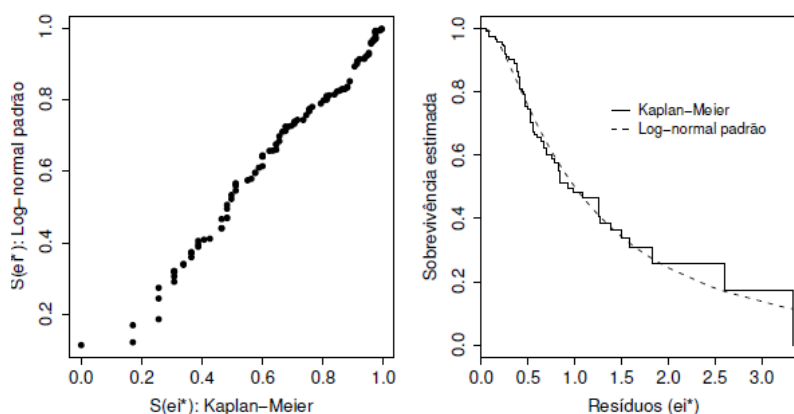
versus uma hipótese alternativa de que o modelo não é adequado. Este teste é usualmente realizado utilizando-se a estatística da razão de verossimilhanças em modelos encaixados.

**Tabela 02:** Teste da razão de verossimilhanças para avaliar a adequação dos modelos de regressão Weibull e log-normal.

Adequação ao Modelo de Regressão	TRV	P-valor
Weibull	5,347	0,0207
Log-normal	0,218	0,6406

Fonte: Análise de Sobrevivência Aplicada (Colosimo e Giolo, 2006)

Baseando-se nos resultados da Tabela 2, conclui-se que existem evidências de que o modelo de regressão log-normal é mais adequado para ajustar os tempos até o desmame. Realizando-se a análise dos resíduos de Cox-Snell foi confirmada a adequação do modelo log-normal (Figura 2).



**Figura 2:** Resíduos padronizados estimados pelo método de Kaplan-Meier para avaliação do modelo log-normal.

O gráfico à direita da Figura 2 apresenta a função sobrevivência dos resíduos estimados tanto pelo método de Kaplan-Meier quanto pelo modelo log-normal padrão. De acordo com estes resultados, nota-se que ambas as curvas estão bem próximas. O gráfico à esquerda, por sua vez, compara as estimativas obtidas pelos dois métodos, que parece se comportar como uma reta, apresentando um ajuste satisfatório. Deste modo, é possível concluir que o modelo de regressão log-normal é adequado para ajustar os tempos até o desmame. As estimativas dos coeficientes do modelo log-normal são apresentadas na Tabela 3.

**Tabela 3:** Estimativas dos parâmetros do modelo de regressão log-normal ajustado aos dados de aleitamento materno.

Covariável	Estimativa	Erro Padrão	valor <i>p</i>
Intercepto	3,293	0,304	< 0,0001
Experiência anterior de amamentação	-0,572	0,301	0,057
Conceito materno sobre o tempo ideal de amamentação	-0,631	0,290	0,029
Dificuldades para	-0,824	0,302	0,006



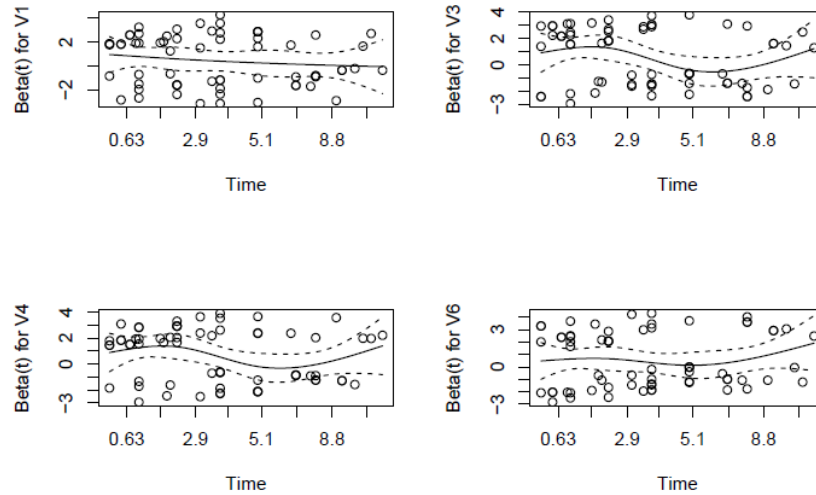
amamentar nos primeiros dias pós-parto			
Recebeu exclusivamente leite materno na maternidade	-0,680	0,293	0,020
Parâmetro de Forma	1,439	0,129	0,001

Como o objetivo deste relatório consiste em comparar o modelo de regressão Weibull com o modelo semiparamétrico de Cox, ambos os modelos também foram ajustados considerando o mesmo conjunto de covariáveis. As estimativas para os parâmetros destes modelos estão apresentadas na Tabela 4.

**Tabela 4:** Estimativas dos parâmetros dos modelos de regressão Weibull e Cox ajustados aos dados de aleitamento materno.

Modelo Covariável	Estimativa	Erro Padrão
<b>Weibull</b>		
Experiência anterior de amamentação	-0.5337	0.2802
Conceito materno sobre o tempo ideal de amamentação	-0.4927	0.2768
Dificuldades para amamentar nos primeiros dias pós-parto	-0.6821	0.2792
Recebeu exclusivamente leite materno na maternidade	-0.6522	0.2749
<b>Cox</b>		
Experiência anterior de amamentação	0,4707	0.2682
Conceito materno sobre o tempo ideal de amamentação	0,5793	0.2624
Dificuldades para amamentar nos primeiros dias pós-parto	0,7158	0.2639
Recebeu exclusivamente leite materno na maternidade	0,5781	0.2639

Para que o modelo de Cox possa ser ajustado aos dados é necessário que a suposição de riscos proporcionais seja atendida. A Figura 3 mostra o gráfico dos resíduos padronizados de Schoenfeld para a verificação da suposição de riscos proporcionais. A Tabela 5 apresenta o resultado do teste de hipóteses de proporcionalidade dos riscos para todas as covariáveis do modelo.



**Figura 3:** Resíduos padronizados de Schoenfeld do modelo de Cox para análise dos dados de aleitamento materno.

**Tabela 5:** Testes da proporcionalidade dos riscos usando resíduos de Schoenfeld obtido no modelo de Cox.

Covariável	Rho	valor $p$
Experiência anterior de amamentação	-0,1098	0,385
Conceito materno sobre o tempo ideal de amamentação	-0,1289	0,298
Dificuldades para amamentar nos primeiros dias pós-parto	-0,1047	0,419
Recebeu exclusivamente leite materno na maternidade	0.0918	0,435
GLOBAL	-	0,520

Com base nos resultados da Figura 3 e da Tabela 5, concluímos que não existem evidências da violação da suposição de riscos proporcionais. Sendo assim, o modelo semiparamétrico de Cox é adequado para ajustar os dados do tempo de aleitamento materno.

## 1.2 Comparação entre modelo de tempo de vida acelerado e modelo de Cox

Seja o modelo de Cox definido como  $\log \lambda(t) = \alpha \log t + \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$  e o modelo de tempo de vida acelerado expresso da forma  $\log T = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \sigma u$ . Em particular, de acordo com Allison (1995), existe uma relação entre os parâmetros dos modelos de Cox e Weibull, que é dada por:

$$\hat{\beta}_j = \frac{-\beta_j}{\sigma}, j = 0, \dots, k$$

onde  $\sigma$  é o parâmetro de escala da distribuição Weibull.

Estudos de simulação foram conduzidos com o objetivo de corroborar a relação existente entre os modelos Weibull e Cox na modelagem de dados de sobrevivência (Allisson, 1995). Os estudos foram realizados no software R versão 3.6.0, considerando-se a ausência e presença de censuras, para tamanhos amostrais iguais a 25, 50, 100 e 500.

Para geração dos tempos de sobrevida usando o modelo Weibull considerou-se duas estratégias: (i) uso do pacote *simsurv*; e (ii) expressão do tempo através da suposição da distribuição Weibull proposto por Bender(2005), que é dada por:

$$t = \left( \frac{\log u}{\lambda e^{x\beta}} \right)^{\frac{1}{\gamma}},$$

onde  $\lambda = \frac{1}{\alpha}$  e a variável aleatória  $u \sim \text{Uniforme}(0,1)$ . Em ambas as estratégias houve a inclusão de uma covariável binária  $X \sim \text{Bernoulli}(0.5)$ ,  $\log(\text{taxa de risco}) = \beta = -0,5$ ,  $\lambda = 0,5$  e  $\gamma = 1,5$ .

Após a geração dos dados, foram ajustados o modelo de tempo de vida acelerado e o modelo de Cox no software R para permitir a comparação das estimativas obtidas pelos dois modelos. As estimativas para o modelo de Cox foram obtidas usando-se a função *coxph* do pacote *survival*. Para estimação dos parâmetros do modelo de tempo de vida acelerado considerou-se o uso de funções em pacotes distintos do software R:

- (1) função *survreg* do pacote *survival*,
- (2) função *WeibullReg* do pacote *SurvRegCensCov*.

De acordo com Haile (2015), a função *WeibullReg* juntamente com a função *ConvertWeibull*, ajusta o modelo de regressão Weibull e converte as estimativas para a escala do modelo de Cox utilizando a relação proposta por Allison (1995). Considerando-se estas estratégias disponíveis no software R, as estimativas apresentadas do modelo Weibull foram as obtidas através do pacote *survreg*, enquanto as estimativas do modelo de Cox foram obtidas através do pacote *coxph*, e também considerou-se as informações sobre *exp(coef)* da função *WeibullReg*, que aproxima a estimativa do modelo Weibull para razão de taxas (HR) do modelo de Cox.

Foram realizadas 1000 simulações para cada uma das estratégias adotadas. Para a realização destas simulações, foi fixada uma semente *set.seed(123456+m)*, onde *m* varia de 0 a 999. Para ambas as simulações o valor do parâmetro  $\beta$  está na escala do modelo de Cox. O valor do parâmetro de escala  $\sigma$  foi obtido por  $1/\gamma$ , ou seja,  $\sigma = 0.67$ . Utilizou-se a relação fornecida por Alisson (1995) para converter as estimativas de  $\beta$  para a escala do modelo com distribuição Weibull. Considerando as especificações do estudo de

simulação,  $\beta$  é aproximadamente igual a 0,33 (valor do parâmetro do modelo Weibull).

As estimativas médias das simulações para o cenário sem censura são apresentadas na Tabela 6. Os resultados dos demais cenários encontram-se no apêndice 1 (Tabela 14). De acordo com os resultados das simulações, à medida que o valor de  $n$  decresce e que há aumento do percentual de censura, as estimativas dos parâmetros afastam-se da relação preconizada para  $\hat{\beta}$  e  $\beta$ . Para valores grandes de  $n$ , mesmo quando há um aumento na porcentagem de censura, as estimativas obtidas pelos dois modelos são próximas. A relação teórica utilizando proposta por Alisson, utilizando o valor fixo do parâmetro de escala, mostra-se mais eficiente do que a função WeibullReg que também utiliza a fórmula proposta por Alisson, porém utiliza os valores estimados para o parâmetro de escala em cada amostra do estudo de simulação.

**Tabela 6:** Estimativa média dos parâmetros dos modelos Weibull e de Cox e estimativas segundo a relação teórica entre os dois modelos para o cenário sem censura.

Modelo	Gerando dados com pacote <i>simsurv</i>				Gerando dados com expressão do tempo			
	n=500	n=100	n=50	n=25	n=500	n=100	n=50	n=25
<b>Weibull</b>								
Estimativa [ <i>survreg</i> ]	0.3342	0.3251	0.3371	0.3319	0.3325	0.3295	0.3387	0.3441
Relação teórica*	0.3347	0.3313	0.3490	0.3582	0.3343	0.3361	0.3486	0.3695
<b>Cox</b>								
Estimativa [ <i>coxph</i> ]	-0.5021	-0.4970	-0.5235	-0.5373	-0.5015	-0.5042	-0.5230	-0.5543
Estimativa log(HR) [ <i>WeibullReg</i> ]	-0.5030	-0.4983	-0.5250	-0.5454	-0.5010	-0.5084	-0.5267	-0.5649
Relação teórica**	-0.5013	-0.4877	-0.5056	-0.4978	-0.4987	-0.4942	-0.5080	-0.5161

\*com base na estimativa do modelo de Cox; \*\* com base na estimativa do modelo Weibull.

## 2. Mediação

Uma variável mediadora (M) auxilia no entendimento de como duas variáveis (X e Y) estão relacionadas de modo sequencial, tal que uma variável X (independente) causa/afeta a variável M (mediadora) que, por sua vez, causa/afeta a variável Y (dependente). Este comportamento é denominado mediação. As seguintes equações dos modelos lineares de regressão capturam o efeito total (Eq. 01) e o efeito mediador (Eq. 02 e 03) considerando a presença de um único mediador:

$$Y = i_1 + cX + e_1 \quad (1)$$

$$Y = i_2 + c'X + bM + e_2 \quad (2)$$

$$M = i_3 + aX + e_3 \quad (3)$$

Na equação (1) o parâmetro  $c$  representa o efeito da relação total entre X e Y sem a consideração de outras variáveis. Na equação (2) o parâmetro  $c'$  representa o efeito direto da relação entre X e Y, controlando-se pelo mediador M, e o parâmetro  $b$  relaciona a variável M à variável Y. Na equação (3), por sua vez, o parâmetro  $a$  relaciona a variável X à variável M (MacKinnon, 2008).

Para situações em que as variáveis dependentes e mediadoras são contínuas, com termos de erro normalmente distribuídos, o procedimento de estimação do efeito mediado e seus erros padrão estão disponíveis e são geralmente precisos (Sobel, 1982; MacKinnon, 1995). Se X é binário, os métodos estatísticos permanecem precisos, mas o tamanho dos efeitos pode ser reduzido se X é uma variável contínua; não se observam dificuldades caso X seja verdadeiramente binário (MacKinnon, 2008, p65).

Os parâmetros dos modelos de regressão linear com erros normais podem ser estimados usando o método de mínimos quadrados ordinários. Para modelos com um mediador, existem duas abordagens estatísticas padrão para a obtenção da estimativa pontual do efeito indireto: (i) considerando o produto dos parâmetros  $a$  e  $b$ , ou (ii) a diferença entre os parâmetros  $c$  e  $c'$ , que representa o efeito de X em Y através da variável M (ou mediado por M).

O erro padrão do efeito mediado pode ser calculado por diversas formas. Os erros padrão baseados em  $\hat{a}\hat{b}$  são chamados de produto do erro padrão dos coeficientes e os erros padrão baseados em  $\hat{c} - \hat{c}'$  são chamados de diferenças do erro padrão dos coeficientes. A estimação do erro-padrão por meio das diferenças é a mais utilizada por conta da facilidade de calcular e generalizar para modelos complicados (McGuigan & Langholtz, 1988), tal que:

$$s_{\hat{c}-\hat{c}'} = \sqrt{s_{\hat{c}}^2 + s_{\hat{c}'}^2 - 2rs_{\hat{c}'}s_{\hat{c}}} \quad (4)$$

onde  $s_{\hat{c}}$  e  $s_{\hat{c}'}$  correspondem ao erro padrão de  $\hat{c}$  e  $\hat{c}'$  e a covariância entre  $\hat{c}$  e  $\hat{c}'$  é definida por  $rs_{\hat{c}}s_{\hat{c}'}$ , que é o quadrado médio do resíduo (QMR) dividido pelo tamanho da amostra multiplicado pela variação da variável independente ( $QMR/N*s_x^2$ ). A utilização deste estimador não é recomendada nos casos em que variável independente é binária. (MacKinnon, 1995).

A metodologia proposta por McGuigan e Langholtz (1988) só está definida para situações em que as variáveis dependentes e mediadoras são contínuas, e em modelos com termos de erro normalmente distribuídos. Isto ocorre devido à necessidade da utilização do valor do quadrado médio dos resíduos para realizar o cálculo do erro padrão da diferença dos coeficientes.

O erro padrão mais comumente usado para o método do produto  $\hat{a}\hat{b}$  é a fórmula derivada de Sobel (1982), sendo definida por

$$s_{1\hat{a}\hat{b}} = \sqrt{\hat{a}^2 s_b^2 + \hat{b}^2 s_a^2} \quad (5)$$

onde  $s_{\hat{a}}$  e  $s_{\hat{b}}$  correspondem ao erro padrão de  $\hat{a}$  e  $\hat{b}$ , respectivamente. Quando os coeficientes de regressão e os erros padrão são pequenos, como geralmente são, é muito mais fácil que os erros de arredondamento afetem a precisão do cálculo manual do erro padrão. Uma fórmula computacionalmente mais fácil é baseada nos valores da estatística t (estatística de Wald) para os efeitos de  $\hat{a}$  e  $\hat{b}$  é definida por

$$s_{2\hat{a}\hat{b}} = \frac{\hat{a}\hat{b} \sqrt{t_a^2 + t_b^2}}{t_a t_b} \quad (6)$$

Cada uma dessas fórmulas de erro padrão pode ser usada para construir limites de confiança superiores e inferiores para o efeito mediado, com base nas seguintes expressões

$$\text{Limite de confiança inferior} = \text{efeito mediado} - z_{\text{erro do tipo 1}} (s_{\hat{a}\hat{b}})$$

$$\text{Limite de confiança superior} = \text{efeito mediado} + z_{\text{erro do tipo 1}} (s_{\hat{a}\hat{b}})$$

A estimativa do efeito mediado é  $\hat{a}\hat{b} = \hat{c} - \hat{c}'$ ,  $z_{\text{erro do tipo 1}}$  é o valor da estatística z (ou t) para os parâmetros de interesse e  $s_{\hat{a}\hat{b}}$  é a estimativa do erro padrão mediado com base em uma das expressões anteriores para o padrão do efeito mediado. Os valores críticos para a distribuição Normal, em vez da distribuição t-Student, são usados principalmente porque as expressões para os erros padrão são aproximações para amostras grandes. Como  $\hat{a}\hat{b}$  é algebricamente equivalente a  $\hat{c} - \hat{c}'$  nos modelos lineares, esses erros padrão podem ser usados para calcular limites de confiança para  $\hat{c} - \hat{c}'$  e também  $\hat{a}\hat{b}$  (MacKinnon, 2008, p52).

Esses intervalos de confiança podem ser simétricos ou assimétricos, que são mais precisos e podem ser obtidos pelo método do produto ou pelo método da reamostragem. Para testar se um efeito mediado é significativamente diferente de zero, é necessário verificar se o zero está contido ou não no intervalo de confiança. Se o zero não estiver contido no intervalo, então o efeito mediado é estatisticamente significativo (Sobel, 1982; MacKinnon, 1995). Os limites de confiança baseados em valores críticos da distribuição normal tendem a ser muito semelhantes aos limites de confiança assimétricos se uma ou ambas as razões  $\frac{\hat{a}}{s_{\hat{a}\hat{b}}}$  ou  $\frac{\hat{b}}{s_{\hat{a}\hat{b}}}$  forem iguais ou maiores que seis (MacKinnon, 2008, p60).

Na subsecção 2.1 é apresentado um exemplo para ilustrar análise de mediação com os conceitos definidos nesta secção.

## 2.1 Aplicação 2: Mediação presumida da mídia

Considere dados de um estudo cujo objetivo consiste em examinar como a importância do tema (*cond*) causa uma determinada reação (*reaction*) mediada pela influência presumida da mídia (*pmi*). O diagrama apresentado na Figura 4 ilustra essas relações. O efeito total da relação entre a importância do tema (*cond*) e uma determinada reação (*reaction*) é representado pelo parâmetro *c*. O parâmetro *c'* representa o efeito direto da relação entre a importância do tema (*cond*) e sua reação (*reaction*), ajustado para os efeitos do mediador.

**Tabela 7:** Estimação dos parâmetros da análise de mediação sobre a influência da mídia.

Parâmetro	Estimativa	IC <sub>95%</sub> ( <i>ab</i> )
Efeito total ( <i>c</i> )	0,4957	...
Efeito direto ( <i>c'</i> )	0,2544	...
Efeito Indireto ( <i>ab</i> )	0,2413	[0,0094;0,5237]

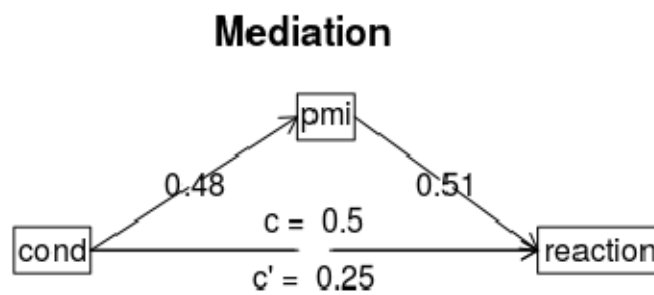
As estimativas apresentadas na Tabela 7 foram obtidas através do software R versão 3.6.0, utilizando a função *mediate* disponível no pacote *psych*, através da qual o intervalo de confiança é obtido através de reamostragens aleatórias com substituição dos dados (Revelle,2017). Com base no resultados da Tabela 7, concluímos que a mediação não é estatisticamente significativa. Quando o tema é importante, espera-se um aumento médio de 0,2544 no escore de reação. O efeito da importância do tema mediado pela influência presumida da mídia na reação do indivíduo é estimado em 0,2413 (95%IC = 0,0094; 0,5237).

Para este mesmo banco de dados também foram calculados erros-padrão usando as diferenças dos erros-padrão dos coeficientes (Eq. 04), o produto dos erros-padrão dos coeficientes (Eq. 05 e Eq. 06), que foram utilizados para construção dos seus respectivos intervalos de 95% de confiança. Baseando-se nos resultados da Tabela 8, concluímos que a mediação não é estatisticamente

significante porque o zero está incluído em todos os intervalos de 95% de confiança.

**Tabela 8:** Intervalos de 95% de confiança para o efeito mediado usando diferentes propostas para estimação dos erros-padrão na análise de mediação sobre a influência da mídia.

Erro Padrão	Estimativa	IC <sub>95%</sub>
$S_{c-c'}$	0,1300	[-0.0134; 0.4961]
$S_{1\hat{a}\hat{b}}$	0,1280	[-0.0096; 0.4922]
$S_{2\hat{a}\hat{b}}$	0,1280	[-0.0096; 0.4922]



**Figura 4:** Diagrama das relações estimadas na análise de mediação sobre a influência da mídia.

A sintaxe no R para as análises apresentadas nesta subsecção encontram-se no Apêndice 2.



### 3. Mediação em Análise de Sobrevivência

Tem havido um uso crescente da análise de mediação com desfechos definidos pelo tempo até o evento (resposta em análise de sobrevivência) e uma literatura ainda limitada sobre os métodos para este tipo de dado. Neste contexto, a variável resposta geralmente não é normalmente distribuída e podem existir eventos não observados (censurados) (Gelfand et al, 2016). As estimativas pontuais e intervalares do efeito mediado com dados de sobrevivência são geralmente obtidas através do ajuste de dois modelos de sobrevivência: modelo de tempo de vida acelerado e modelo de regressão de Cox. No contexto de análise de sobrevivência, alguns autores discutem a necessidade da definição de critérios formais para ajudar pesquisadores, revisores e leitores a reportar ou avaliar seus estudos de mediação (Gelfand et al, 2016; Ochoa, 2020; Lapointe-Shaw et al.,2018; Tein e MacKinnon, 2003).

Avanços recentes na análise de mediação causal formalizaram condições para estimar efeitos diretos e indiretos em diversos contextos, que foram estendidas a vários modelos para análise de dados de sobrevivência, incluindo modelos de tempo de falha acelerado, que são amplamente utilizados em uma gama de aplicações em saúde devido à sua interpretação intuitiva (Fulcher; 2017). Os efeitos mediados no modelo de tempo de vida acelerado (AFT) com a presença de um único mediador podem ser determinados em três equações:

$$Y = \log(T_i) = i_4 + c'X + bM + \sigma v_1 \quad (7)$$

$$Y = \log(T_i) = i_5 + cX + \sigma v_2 \quad (8)$$

$$M = i_6 + aX + e_4 \quad (9)$$

As equações (7) e (8) consideram o modelo AFT, enquanto a equação (9) o modelo de regressão linear para o mediador. A variável resposta  $Y$  é o log do tempo de sobrevivência ( $T_i$ ),  $X$  é a variável de exposição (independente),  $M$  a variável mediadora, os  $i$ 's representam os interceptos, o  $e_4$  representa o termo de erro do modelo de regressão linear para o mediador e os  $\sigma v$ 's representam o termo de erro do modelo de tempo de vida acelerado. Sabe-se que o logaritmo de uma variável aleatória que segue uma distribuição Weibull resulta em uma distribuição de valor extremo com dois parâmetros. Assim, enquanto  $e_4$  segue uma distribuição normal, o erro  $v_z, z = \{1,2\}$  segue uma distribuição de valor extremo de dois parâmetros e  $\sigma$  representa o parâmetro de escala.

Os efeitos mediados nos modelos de Cox, com a presença de um único mediador, podem ser determinados pelo modelo de regressão linear (Eq. 9) e os modelos de Cox (Eqs. 10 e 11).

$$\lambda(t) = \lambda_0(t) \exp\{\beta'X + \gamma M\} \quad (10)$$

$$\lambda(t) = \lambda_0(t) \exp\{\beta'X\} \quad (11)$$

onde  $X$  é a variável de exposição (independente),  $M$  a variável mediadora e  $\lambda_0(t)$  representa a taxa de falha basal.

Os métodos para estimação do efeito mediado definidos na secção 2 - métodos da diferença e do produto de coeficientes - são equivalentes e justificados quando os dois modelos tanto para a variável resposta quanto para o mediador são lineares, com erros seguindo distribuição normal, e sem interação. No entanto, estes métodos diferem e existem questionamentos sobre sua validade quando um ou ambos os modelos é /são não lineares (Vo; 2020).

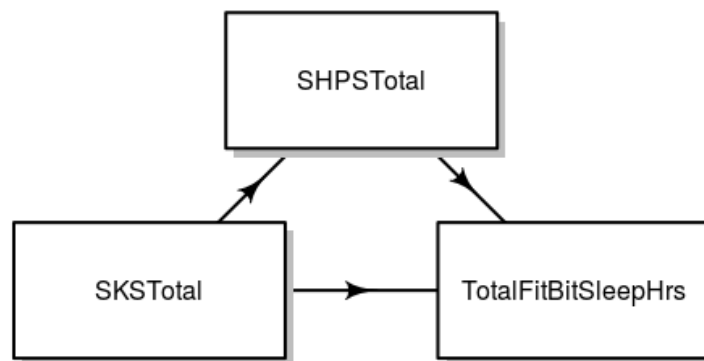
No contexto de análise de sobrevivência, tem-se que, sob a condição de não haver dados censurados, os efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes são idênticos utilizando o modelo de tempo de vida acelerado (*AFT = accelerated failure time*, em inglês). Porém a equivalência entre os dois métodos de estimação de efeitos mediados não é sempre observada utilizando o modelo de regressão de Cox. Estudos anteriores desencorajaram o uso do modelo de riscos proporcionais de Cox para análise de mediação tradicional, pois ele pode fornecer resultados tendenciosos (Tein e MacKinnon, 2003; Ochoa, 2019).

Os resultados disponíveis na literatura levantam a questão de saber se um dos métodos tradicionais para calcular o efeito indireto fornece uma aproximação precisa e imparcial do efeito indireto nos modelos em análise de sobrevivência (Ochoa, 2019). Devido à limitada discussão disponível na literatura sobre estimação do efeito mediado pelos modelos AFT e de Cox para diferentes percentuais de censura, tamanhos amostrais e magnitude do efeito, estudos de simulação foram conduzidos neste trabalho e são descritos na subsecção 3.2.

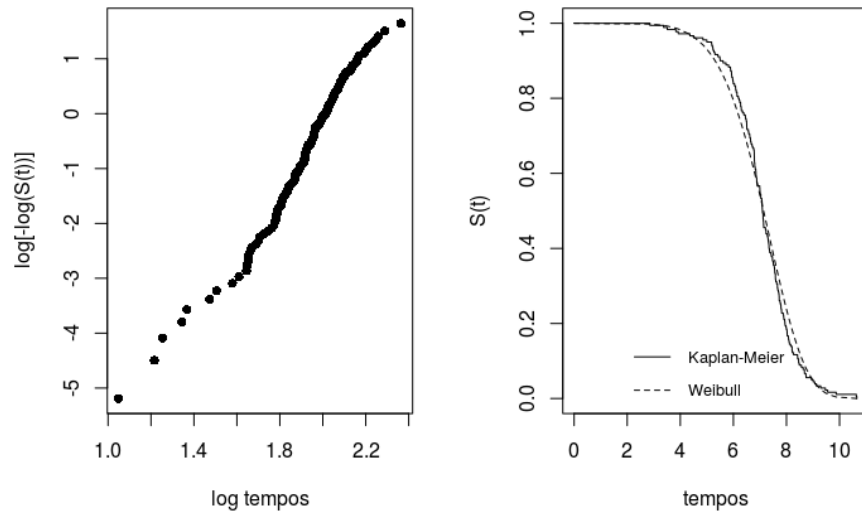
### 3.1 Aplicação 3: Mediação da higiene do sono

Considere os dados de um estudo, sem a presença de censura, cujo objetivo consiste em examinar como o conhecimento da importância e função do sono ( $X$ ) afeta a duração do sono ( $T$ ), em horas, mediada pela influência da higiene do sono ( $M$ ). Este estudo foi conduzido no sudeste dos Estados Unidos, onde 218 universitários de 18 a 25 anos responderam uma série de questionários antes e depois de usar um acelerômetro FitBit Flex por 7 dias. A variável dependente contínua (TotalFitBitSleepHrs) representa a duração do sono e foi coletada através da utilização do acelerômetro. A variável mediadora (SHPSTotal) é também uma

variável contínua e foi obtida por meio da Escala de Prática de Higiene do Sono (SHPS; Lin, Cheng, Yang e Hsu, 2007), que avalia práticas de atividades da vida diária e hábitos de sono que podem impactar negativamente o sono, para medir comportamentos de higiene do sono. O conhecimento da importância e função do sono (SKSTotal) foi definido por informações precisas sobre a ciência e a fisiologia do sono, fatores que influenciam o sono, os requisitos de sono, as características dos distúrbios do sono e os resultados de saúde do sono inadequado. Este conhecimento foi medido usando itens de um questionário interativo do sono, fornecido pelo National Center on Sleep Disorders Research (NIH MedlinePlus, n.d.), onde participantes indicaram a precisão de cada afirmação escolhendo "verdadeiro" ou "falso" e cada resposta correta foi codificada como 1 ponto. O total da soma destes itens forma as pontuações presentes na variável de exposição contínua (Peach,2018). O diagrama apresentado na Figura 5 ilustra essas relações. O efeito total do conhecimento da importância e função do sono (SKSTotal) na duração do sono (TotalFitBitSleepHrs) é representado pelo parâmetro  $c$ , enquanto o parâmetro  $c'$  representa o efeito direto da importância do sono (SKSTotal) na sua duração (TotalFitBitSleepHrs), ajustado para o efeito do mediador (SHPSTotal).



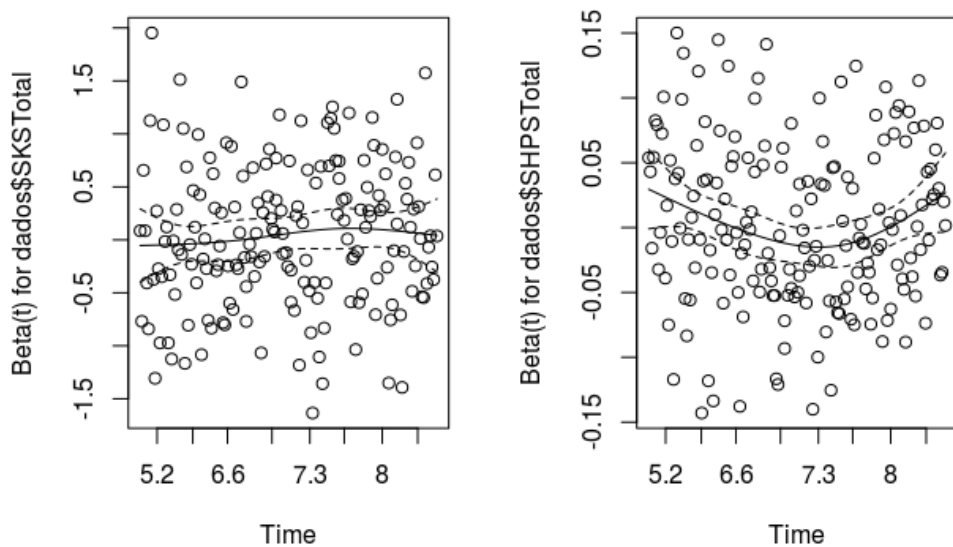
**Figura 5:** Diagrama das relações na análise de mediação sobre a higiene do sono.



**Figura 6:** Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para avaliação do modelo Weibull.

No gráfico à direita da Figura 6, nota-se que as curvas da função de sobrevivência estimadas pelo método de Kaplan-Meier e Weibull, estão bem próximas. O gráfico à esquerda, por sua vez, compara as estimativas obtidas pelos dois métodos, que parece se comportar como uma reta, apresentando um ajuste satisfatório. Deste modo, é possível concluir que o modelo de regressão Weibull é adequado para ajustar os tempos de duração de sono.

Similarmente, verificou-se a suposição de riscos proporcionais para o modelo de Cox através da construção de gráfico dos resíduos padronizados de Schoenfeld (*Figura 7*) e também através da realização de teste de hipóteses de proporcionalidade dos riscos para todas as covariáveis do modelo (*Tabela 9*). Conclui-se, por ambos os métodos, que não existem evidências da violação da suposição de riscos proporcionais. Sendo assim, o modelo semiparamétrico de Cox também é adequado para ajustar os dados do tempo, em horas, de duração do sono.



**Figura 7:** Resíduos padronizados de Schoenfeld do modelo de Cox para análise de mediação sobre a influência da higiene do sono.

**Tabela 9:** Testes da proporcionalidade dos riscos usando resíduos de Schoenfeld obtido no modelo de Cox.

Covariável	Rho	valor $p$
SKSTotal	0,0663	0,372
SHPSTotal	-0.0649	0,348
GLOBAL	-	0,488

Para análise de mediação calculou-se o erro-padrão para o estimador produto dos coeficientes (Eq. 05), que foi utilizado para construção do intervalo de 95% de confiança pelos modelos AFT e Cox. As estimativas para os parâmetros destes modelos estão apresentadas na Tabela 10. Baseando-se nos resultados da Tabela 10, concluímos que a mediação não é estatisticamente significativa porque o zero está incluído no intervalo de 95% de confiança. O aumento no conhecimento da importância e função do sono reduz, em média, 0.0067 no logaritmo do total de horas dormidas no modelo AFT. O efeito do conhecimento da importância e função do sono, mediado pela higiene do sono, é estimado em -0.0007 (95%IC = -0.0028; 0.0014) pelo método do produto no modelo AFT, enquanto a estimativa pontual foi igual a -0.0012 pelo método da diferença. No modelo de Cox, por sua vez, o efeito é praticamente nulo (HR=1,00) (log HR = 0.0022; 95%IC = -0.0100; 0.0146) pelo método do produto.

**Tabela 10:** Estimaco dos parâmetros dos modelos AFT e Cox da anlise de mediao sobre a influncia da higiene do sono em sua durao.

<b>Modelo</b>	<b>Estimativa</b>
Parâmetro	
<b>Weibull</b>	
Efeito total ( <i>c</i> )	-0.0067
Efeito direto ( <i>c'</i> )	-0.0055
Efeito Indireto ( <i>c-c'</i> )	-0.0012
Efeito Indireto ( <i>ab</i> )	-0.0007
<b>Cox</b>	
Efeito total ( <i>c</i> )	0.0391
Efeito direto ( <i>c'</i> )	0.0359
Efeito Indireto ( <i>c-c'</i> )	0.0032
Efeito Indireto ( <i>ab</i> )	0.0022

### 3.2 Comparaco entre o modelo de tempo de vida acelerado e o modelo de Cox

Foram conduzidos estudos de simulaco com o objetivo de comparar a estimaco dos efeitos mediados pelos mtodos do produto e da diferena; e tambm para avaliar o comportamento do procedimento para estimaco do erro-padro dos efeitos mediados usando o mtodo do produto na modelagem de dados de sobrevivncia utilizando os modelos Weibull e de Cox, na presena e ausncia de censuras.

#### i. Especificaces dos estudos de simulaco

Os estudos foram realizados no software R verso 3.6.0, para tamanhos amostrais iguais a 75, 100 e 500. A gerao dos tempos de sobrevida utilizando o modelo Weibull foi feita em duas etapas:

I. A *varivel mediadora* foi simulada por meio do modelo de regresso linear, dado por:  $M = i + aX + e$ , considerando-se duas situaes: (i) varivel independente binria, tal que  $X \sim \text{Bernoulli}(0.5)$  e (ii) varivel independente contnua, sendo  $X \sim N(0,1)$ . Os parâmetros utilizados para a gerao dos dados foram:  $i = 100$ ,  $e \sim N(0,1)$  e  $a = 0.2$  e  $10$ .

II. Os tempos de sobrevida foram gerados atravs da expresso proposta por Bender(2005), assumindo uma distribuo Weibull para  $t$ , tal que:

$$t = \left( \frac{\log u}{\lambda e^{xc' + Mb}} \right)^{\frac{1}{\gamma}},$$

onde a varivel aleatria  $u \sim \text{Uniforme}(0,1)$ . Os parâmetros utilizados foram:  $\gamma = 1$ ;  $b = -0,05$ ; e  $c' = 0,3$ . Diferentes valores de  $\lambda$  foram utilizados a fim de obter diferentes

percentuais de censura. Considerou-se um tempo máximo de acompanhamento do estudo de 120 dias.

Após a geração dos dados, as funções `coxph` e `survreg` do pacote `survival`, respectivamente, do software R foram utilizadas para ajustar o modelo de tempo de vida acelerado e o modelo de Cox. Diversas métricas para comparação das estimativas dos dois modelos foram utilizadas e são descritas a seguir. O viés relativo (VR) percentual foi definido por

$$\left( \frac{\hat{\beta} - \beta}{|\beta|} \right) \cdot 100$$

onde  $\beta$  denota o parâmetro de interesse, definido por  $\beta=ab$ , que assumiu os valores 0,01 e 0,5, de acordo com as diferentes configurações, enquanto  $\hat{\beta} = \hat{a}\hat{b}$  denota a estimativa pontual dos efeitos mediados. As fórmulas derivadas de Sobel (Eq. 05) e baseada na estatística de Wald (Eq. 06) foram utilizadas para estimação do erro-padrão do efeito mediado. As estimativas do erro-padrão obtidos pelas Eqs.05 e 06 foram comparadas com o desvio padrão do efeito mediado estimado utilizando o método do produto para avaliar a performance destes estimadores. A probabilidade de cobertura percentual (CP%) do intervalo de confiança baseado nos erros-padrão definidos anteriormente apresenta o número de vezes em que o efeito mediado estava contido neste intervalo (expressa em porcentagem) dentre as 1000 simulações realizadas para cada uma das configurações. Para a realização destas simulações, foi fixada a semente `set.seed(123456+i)`, onde  $i$  varia de 1 a 1000.

## ii. Resultados

Os resultados das simulações para o cenário sem censura e com censura com  $a = 10$  são apresentados nas Tabelas 11 e 12. Os demais resultados encontram-se no apêndice (Tabelas 14 e 15). As estimativas do erro-padrão ( $s_{2\hat{a}\hat{b}}$ ) calculadas a partir da Eq. 06 não estão tabeladas, pois apresentaram resultados idênticos às estimativas do erro-padrão ( $s_{1\hat{a}\hat{b}}$ ) obtidas a partir da Eq. 05.

As estimativas dos efeitos mediados utilizando os métodos do produto e da diferença dos coeficientes são similares utilizando a variável independente binária e contínua, na ausência e presença censura, quando se utiliza o modelo Weibull. No entanto, a equivalência entre os métodos não é observada com o modelo de Cox, para o qual se verifica que à medida que o tamanho amostral decresce, aumenta a diferença entre as estimativas obtidas pelos dois métodos de estimação do efeito mediado, sobretudo quando a covariável é binária (Tabelas 11 e 12). Os resultados das simulações dos demais cenários, encontram-se nas Tabelas 14 e 15 em anexo no Apêndice 1.

Verifique, por exemplo, na Tabela 12, que para 70% de censura e  $n=500$ , os efeitos mediados pelo método do produto e da diferença, são, respectivamente, iguais a 0.5103 e 0.5098 utilizando o modelo de tempo de vida acelerado; e iguais a -0.5096 e -0.5078 no modelo de Cox. No entanto, mantendo-se o percentual de censura e considerando  $n=75$ , as

correspondentes estimativas são iguais, respectivamente, a 0.6267 e 0.6219 pelo AFT e iguais a -0.6183 e -0.6050 pelo modelo de Cox.

**Tabela 11:** Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados sem censura e  $\alpha = 10$ .

Coeficiente	Weibull			Cox		
	n=500	n=100	n=75	n=500	n=100	n=75
<i>Variável independente</i>						
<i>binária</i>						
$\hat{\alpha}\hat{b}$	0.5130	0.4899	0.5317	-0.5138	-0.4973	-0.5510
$\hat{c} - \hat{c}'$	0.5132	0.4899	0.5313	-0.5107	-0.4892	-0.5392
$VR\%_0(\hat{\alpha}\hat{b})$	2.5954	-2.0197	6.3390	-2.7540	0.5362	-10.2072
$VR\%_0(\hat{c} - \hat{c}')$	2.6346	-2.0222	6.2643	-2.1365	2.1548	-7.8456
$s_{1\hat{\alpha}\hat{b}}$	0.4498	1.0194	1.1770	0.4537	1.0574	1.2387
$dp(\hat{\alpha}\hat{b})$	0.4532	1.0119	1.2357	0.4569	1.0488	1.3053
$CP\%(s_{1\hat{\alpha}\hat{b}})$	94.3	96.2	93.6	95	96.7	94.5
<i>Variável independente</i>						
<i>contínua</i>						
$\hat{\alpha}\hat{b}$	0.5037	0.4980	0.5065	-0.5051	-0.5061	-0.5161
$\hat{c} - \hat{c}'$	0.5038	0.4984	0.5059	-0.5024	-0.4989	-0.5064
$VR\%_0(\hat{\alpha}\hat{b})$	0.7489	-0.3930	1.2956	-1.0271	-1.2114	-3.2248
$VR\%_0(\hat{c} - \hat{c}')$	0.7599	-0.3118	1.1872	-0.4827	0.2144	-1.2848
$s_{1\hat{\alpha}\hat{b}}$	0.4475	1.0242	1.1828	0.4537	1.0635	1.2474
$dp(\hat{\alpha}\hat{b})$	0.4533	1.0285	1.1786	0.4595	1.0667	1.2462
$CP\%(s_{1\hat{\alpha}\hat{b}})$	94.7	94.7	94.4	94.9	94.9	94.6

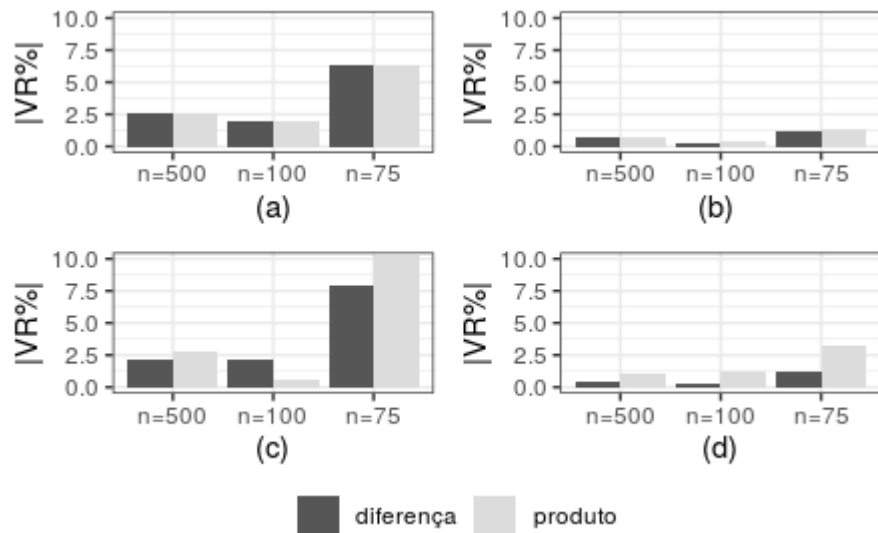


**Tabela 12:** Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados com censura e  $\alpha = 10$ .

Censura%	Weibull			Cox		
	n=500	n=100	n=75	n=500	n=100	n=75
<i>Variável independente binária</i>						
<b>25%</b>						
$\hat{\alpha}\hat{\beta}$	0.5042	0.4930	0.5916	-0.5061	-0.4945	-0.6012
$\hat{c} - \hat{c}'$	0.5038	0.4910	0.5884	-0.5034	-0.4857	-0.5882
$VR\%_0(\hat{\alpha}\hat{\beta})$	0.8395	-1.3999	18.3171	-1.2251	1.1077	-20.2417
$VR\%(\hat{c} - \hat{c}')$	0.7580	-1.8046	17.6828	-0.6846	2.8515	-17.6426
$s_{1\hat{\alpha}\hat{\beta}}$	0.5204	1.1860	1.3737	0.5236	1.2190	1.4287
$dp(\hat{\alpha}\hat{\beta})$	0.5158	1.2071	1.47	0.5214	1.2279	1.5086
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	94.8	95.7	93.8	94.9	96.4	94.3
<b>50%</b>						
$\hat{\alpha}\hat{\beta}$	0.5017	0.4818	0.6148	-0.5015	-0.4939	-0.6231
$\hat{c} - \hat{c}'$	0.5011	0.4792	0.6094	-0.4993	-0.4856	-0.6088
$VR\%_0(\hat{\alpha}\hat{\beta})$	0.3415	-3.6358	22.9538	-0.2924	1.2190	-24.6134
$VR\%(\hat{c} - \hat{c}')$	0.2290	-4.1501	21.8885	0.1476	2.8794	-21.7693
$s_{1\hat{\alpha}\hat{\beta}}$	0.6398	1.4627	1.6968	0.6416	1.4876	1.7446
$dp(\hat{\alpha}\hat{\beta})$	0.669	1.4993	1.8062	0.6713	1.5110	1.8445
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	94.1	95	93.7	94.4	95.2	94.1
<b>70%</b>						
$\hat{\alpha}\hat{\beta}$	0.5103	0.4729	0.6267	-0.5096	-0.4334	-0.6183
$\hat{c} - \hat{c}'$	0.5098	0.4710	0.6219	-0.5078	-0.4264	-0.6050
$VR\%_0(\hat{\alpha}\hat{\beta})$	2.0652	-5.4274	25.3494	-1.9175	13.3159	-23.6526
$VR\%(\hat{c} - \hat{c}')$	1.9574	-5.7967	24.3795	-1.5695	14.7134	-21.0045
$s_{1\hat{\alpha}\hat{\beta}}$	0.8367	1.948	2.2382	0.836	1.9458	2.2771
$dp(\hat{\alpha}\hat{\beta})$	0.8868	2.0887	2.3655	0.8834	2.0458	2.3863
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	93.8	96.1	94.9	93.9	95.5	94.4
<i>Variável independente contínua</i>						
<b>25%</b>						
$\hat{\alpha}\hat{\beta}$	0.4966	0.4940	0.5040	-0.4984	-0.5069	-0.5013
$\hat{c} - \hat{c}'$	0.4962	0.4931	0.5028	-0.4959	-0.4993	-0.4910
$VR\%_0(\hat{\alpha}\hat{\beta})$	-0.6824	-1.2056	0.7989	0.3193	-1.3795	-0.2562
$VR\%(\hat{c} - \hat{c}')$	-0.7617	-1.3845	0.5528	0.8166	0.1350	1.8053
$s_{1\hat{\alpha}\hat{\beta}}$	0.5190	1.1887	1.3800	0.5243	1.2231	1.4369
$dp(\hat{\alpha}\hat{\beta})$	0.5281	1.2135	1.4132	0.5315	1.2429	1.4814
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	94.7	94.5	94.6	94.8	94.4	94.6
<b>50%</b>						
$\hat{\alpha}\hat{\beta}$	0.4846	0.4838	0.5029	-0.4861	-0.4718	-0.4881
$\hat{c} - \hat{c}'$	0.4839	0.4821	0.5015	-0.4838	-0.4643	-0.4774
$VR\%_0(\hat{\alpha}\hat{\beta})$	-3.084	-3.2385	0.5812	2.7823	5.6466	2.3807
$VR\%(\hat{c} - \hat{c}')$	-3.216	-3.5796	0.3092	3.2329	7.1364	4.5276
$s_{1\hat{\alpha}\hat{\beta}}$	0.6382	1.4682	1.7119	0.6425	1.4988	1.7609
$dp(\hat{\alpha}\hat{\beta})$	0.6428	1.5093	1.8621	0.6455	1.5219	1.8976
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	94.7	94.7	92.5	94.9	94.5	93.6
<b>70%</b>						
$\hat{\alpha}\hat{\beta}$	0.4810	0.4517	2.1344	-0.4851	-0.4585	-0.5593
$\hat{c} - \hat{c}'$	0.4803	0.4503	2.1014	-0.483	-0.4507	-0.5462
$VR\%_0(\hat{\alpha}\hat{\beta})$	-3.8022	-9.656	326.8767	2.9726	8.2914	-11.8671
$VR\%(\hat{c} - \hat{c}')$	-3.9359	-9.9414	320.2763	3.3963	9.8577	-9.2398
$s_{1\hat{\alpha}\hat{\beta}}$	0.8268	1.9241	<b>2.2644</b>	0.832	1.9483	2.2932
$dp(\hat{\alpha}\hat{\beta})$	0.8473	1.9886	<b>50.3044</b>	0.8528	2.0059	2.4739
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	94.2	95	93.9	94.1	95	94

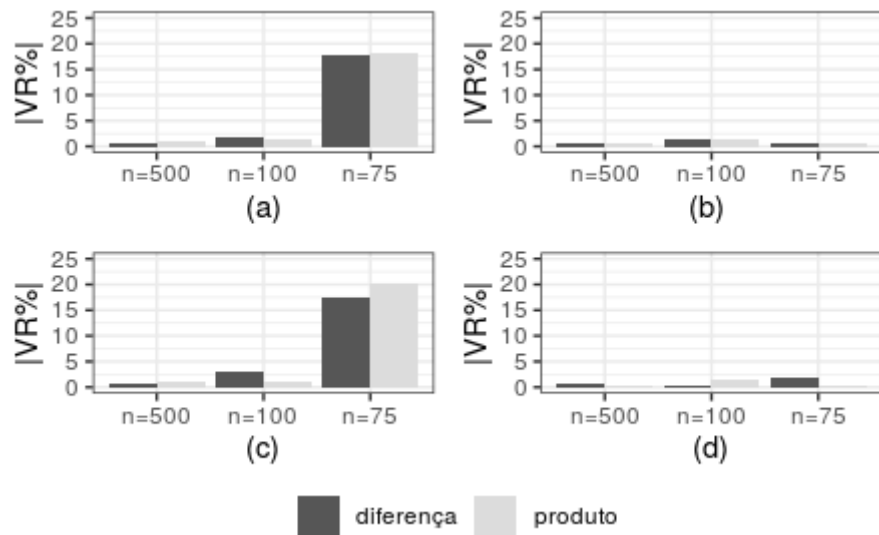
O módulo do viés relativo percentual ( $|VR\%|$ ) aumenta à medida que ocorre um aumento no percentual da censura, como esperado. Quando há um aumento no valor do efeito mediado ( $\hat{\alpha}\hat{\beta} = 0.5$ ), em todas as configurações a variável independente binária apresenta um maior  $|VR\%|$ . Estes comportamentos podem ser observados ao analisar as Figuras 8 e 9, com percentuais de censura respectivamente, de 0% e 50%. As demais figuras que ilustram o comportamento do  $|VR\%|$  nos demais cenários, encontram-se nas Figuras 11-17 no Apêndice 1.

Por exemplo, na Figura 8, considerando-se a ausência de censura, o maior valor do  $|VR\%|$  é de 10,2%, (Figura 8 c) para o modelo de Cox; com método do produto; variável independente binária; e  $n=75$ . No entanto, mantendo-se as mesmas configurações e aumentando o percentual de censura para 50% (Figura 9 c), a correspondente estimativa do  $|VR\%|$  é igual a 24,6%.



(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 8:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na ausência de censura e  $\hat{\alpha}\hat{\beta} = 0,5$ , utilizando os modelos AFT e de Cox.



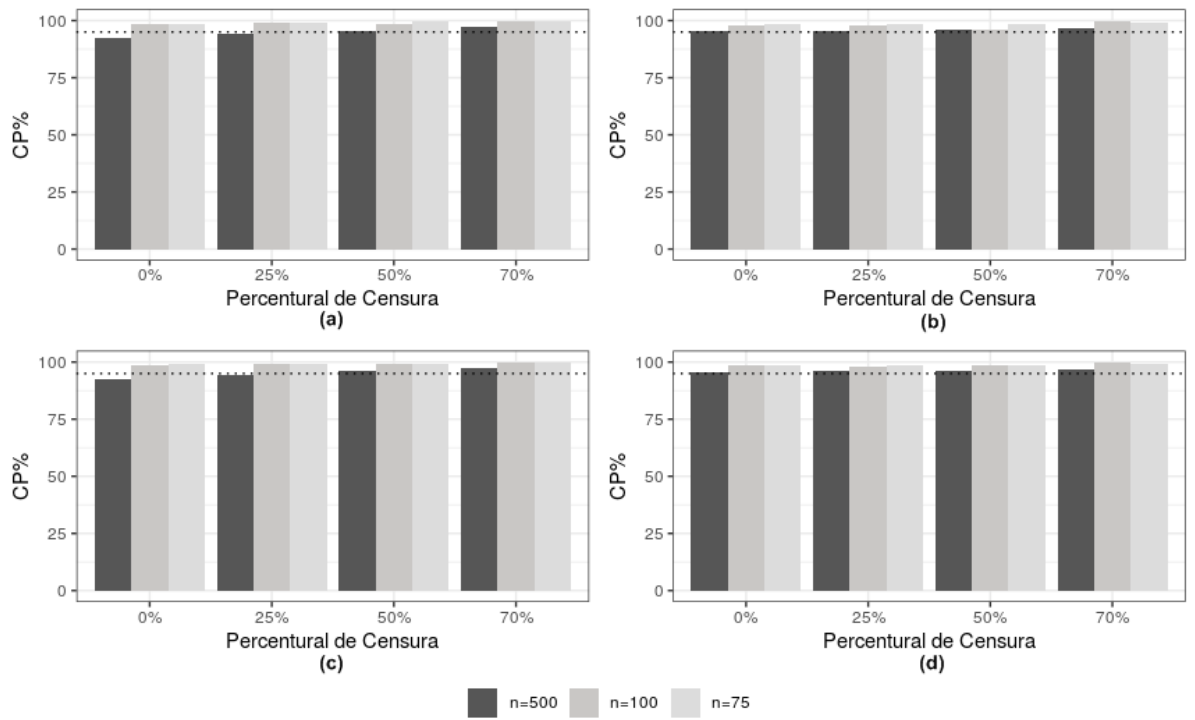
(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 9:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 50% de censura e  $\hat{\alpha} = 0,5$ , utilizando os modelos Weibull e de Cox.

O modelo AFT apresentou alguns problemas ao calcular o erro-padrão baseado na estatística de Wald ( $s_{2_{\hat{\alpha}\hat{\beta}}}$ ), retornando valores faltantes (NaN) para algumas amostras geradas utilizando as seguintes configurações: 70% de censura;  $n=75$ ; variável de exposição binária e contínua; o que, conseqüentemente, afeta o resultado das seguintes avaliações: (i) performance do estimador do erro-padrão; (ii) probabilidades de cobertura.

A Figura 10 mostra resultados referentes às probabilidades de cobertura percentual (CP%) dos intervalos de 95% de confiança construído baseado no erro padrão derivado de Sobel ( $s_{1_{\hat{\alpha}\hat{\beta}}}$ ), com  $\alpha=0,2$ , variando o tamanho amostral, o percentual de censura e o tipo da variável de exposição no modelo. Quando há uma diminuição do tamanho amostral e um aumento no percentual de censura, nota-se um esperado afastamento das probabilidades de cobertura do valor 95% em todos os modelos.

Por exemplo, na Figura 10, para as seguintes configurações: 0%, 25%, 50% e 70% de censura; variável de exposição contínua; e  $n=500$ ; as probabilidades de cobertura são, respectivamente, iguais a 95.6%, 95.5%, 95.8% e 96.7% utilizando o modelo de tempo de vida acelerado; e iguais a 95.4%, 96.1%, 96.2% e 96.9% no modelo de Cox. No entanto, mantendo-se as mesmas configurações e considerando  $n=75$ , as correspondentes estimativas são iguais, respectivamente, a 98.5%, 98.3%, 98.5% e 99.2% pelo AFT e iguais a 98.7%, 98.8%, 98.3% e 98.9% pelo modelo de Cox.



(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 10:** Probabilidade de cobertura percentual (CP%) do intervalo de confiança construído baseado no erro-padrão estimado, com  $\alpha = 0,2$ .

## REFERÊNCIAS BIBLIOGRÁFICAS

- Allison, P.D. (1995) *Survival Analysis Using the SAS System: A Practical Guide*. Editora SAS Institute Inc.
- Bender, R., Augustin, T. and Blettner, M. (2005). *Generating survival times to simulate Cox proportional hazards models*. **Statist. Med.**, 24: 1713-1723. doi:10.1002/sim.2059
- Colosimo, E.A. e Giolo, S.R. (2005). *Análise de Sobrevida Aplicada*. ABE-Projeto Fisher. Editora Edgard Blucher
- Fulcher I, Tchetgen Tchetgen E, Williams P (2017). Mediation analysis for censored survival data under an accelerated failure time model. **Epidemiology**, 28(5):660–666
- Gelfand L.A., MacKinnon D.P., DeRubeis R.J. e Baraldi A.N. (2016). *Mediation Analysis with Survival Outcomes: Accelerated Failure Time vs. Proportional Hazards Models*. **Front. Psychol**, 7:423. doi: 10.3389/fpsyg.2016.00423
- Iacobucci, D., Saldanha, N., Deng X. (2007) *A meditation on mediation: Evidence that structural equations models perform better than regressions*. **Journal of Consumer Psychology**, 17 (2): 139-153.
- Lapointe-Shaw, L., Bouck, Z., Howell, N.A. et al. (2018) *Mediation analysis with a time-to-event outcome: a review of use and reporting in healthcare research*. **BMC Med Res Methodol**, 18, 118.
- MacKinnon D.P (2008) *Introduction to mediation analysis*. New York: Erlbaum
- MacKinnon D.P., Warsi G., Dwyer J.H. (1995). *A simulation study of mediated effect measures*. **Multivariate Behavioral Research**, 30 : 41-62.
- Ochoa, L. B., Rijnhart, J. J. M., Penninx, B. W., Wardenaar, K. J., Twisk, J. W. R. & Heymans, M. W. (2019). *Performance of methods to conduct mediation analysis with time-to-event outcomes*. **Statistica Neerlandica**, 74(1) 72-91
- Peach, H.D., Gaultney, J.F. & Ruggiero, A.R. (2018). *Direct and Indirect Associations of Sleep Knowledge and Attitudes With Objective and Subjective Sleep Duration and Quality via Sleep Hygiene*. **The Journal of Primary Prevention**, 39:555–570 <https://doi.org/10.1007/s10935-018-0526-7>
- Revelle W. (2017). *How to Use the Psych Package for Mediation/Moderation/Regression Analysis*. (<http://personality-project.org/r/psych/HowTo/mediation.pdf>)
- Sarah R. H. (2019). *Weibull AFT Regression Functions* in R. (<https://cran.r-project.org/web/packages/SurvRegCensCov/vignettes/weibull.pdf>)
- Sobel M.E. (1982). *Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models*. **Sociological Methodology**, 13, 290. doi:10.2307/270723
- Tein, J.-Y., MacKinnon, D. P. (2003). *Estimating Mediated Effects with Survival Data*. **New Developments on Psychometrics**, 405-412
- Vo T.T., Superchi C., Boutron I., Vansteelandt S. (2020). *The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review*. **J Clin Epidemiol**, 117:78-88. doi:10.1016/j.jclinepi.2019.10.001

## Apêndice 1

Neste apêndice encontram-se resultados adicionais dos dois estudos de simulação apresentados neste relatório técnico. A Tabela 13 apresenta resultado referente ao estudo realizado para comparar as estimativas obtidas pelos modelos Weibull e Cox em análise de sobrevivência clássica, enquanto que as Tabelas 14 e 15 apresentam os resultados referentes à análise de mediação em análise de sobrevivência. Resultados adicionais deste segundo estudo de simulação são apresentados nas Figuras 11-17.

**Tabela 13:** Estimativa média dos parâmetros dos modelos Weibull e de Cox e estimativas segundo a relação teórica entre os dois modelos para o cenário com censura.

Modelo (Censura%)	Pacote simsurv Estimativa				Expressão do tempo Estimativa			
	n=500	n=100	n=50	n=25	n=500	n=100	n=50	n=25
<b>Weibull (25%)</b>								
Estimativa ( $\beta$ ) – <i>survreg</i>	0.3331	0.3232	0.3456	0.3245	0.3358	0.3277	0.3464	0.3315
Relação teórica	0.3333	0.3282	0.3426	0.3475	0.3363	0.3335	0.3445	0.3588
<b>Cox (25%)</b>								
Estimativa ( $\beta$ ) - <i>coxph</i>	-0.4999	-0.4923	-0.5138	-0.5212	-0.5044	-0.5002	-0.5167	-0.5383
Estimativa log(HR) - <i>WeibullReg</i>	-0.5007	-0.4950	-0.5231	-0.5283	-0.5050	-0.5039	-0.5253	-0.5425
Relação teórica	-0.4997	-0.4848	-0.5184	-0.4868	-0.5037	-0.4915	-0.5196	-0.4972
<b>Weibull (50%)</b>								
Estimativa ( $\beta$ ) – <i>survreg</i>	0.3356	0.3243	0.3359	0.3239	0.3381	0.3276	0.3389	0.3387
Relação teórica	0.3358	0.3297	0.3353	0.3527	0.3392	0.3342	0.3404	0.3643
<b>Cox (50%)</b>								
Estimativa ( $\beta$ ) - <i>coxph</i>	-0.5038	-0.4945	-0.5029	-0.5290	-0.5088	-0.5013	-0.5106	-0.5465
Estimativa log(HR) - <i>WeibullReg</i>	-0.5045	-0.4974	-0.5099	-0.5375	-0.5093	-0.5047	-0.5176	-0.5556
Relação teórica	-0.5034	-0.4865	-0.5038	-0.4858	-0.5072	-0.4914	-0.5084	-0.5080
<b>Weibull (75%)</b>								
Estimativa ( $\beta$ ) – <i>survreg</i>	0.3418	0.3231	0.3860	0.5663	0.3409	0.3334	0.3433	0.5189
Relação teórica	0.3429	0.3270	0.3973	0.6553	0.3415	0.3424	0.3540	0.6019
<b>Cox (75%)</b>								
Estimativa ( $\beta$ ) - <i>coxph</i>	-0.5143	-0.4905	-0.5959	-0.9830	-0.5123	-0.5135	-0.5310	-0.9028
Estimativa log(HR) - <i>WeibullReg</i>	-0.5149	-0.4933	-0.6049	-0.9937	-0.5129	-0.5165	-0.5358	-0.9095
Relação teórica	-0.5127	-0.4847	-0.5789	-0.8495	-0.5113	-0.5001	-0.5150	-0.7784

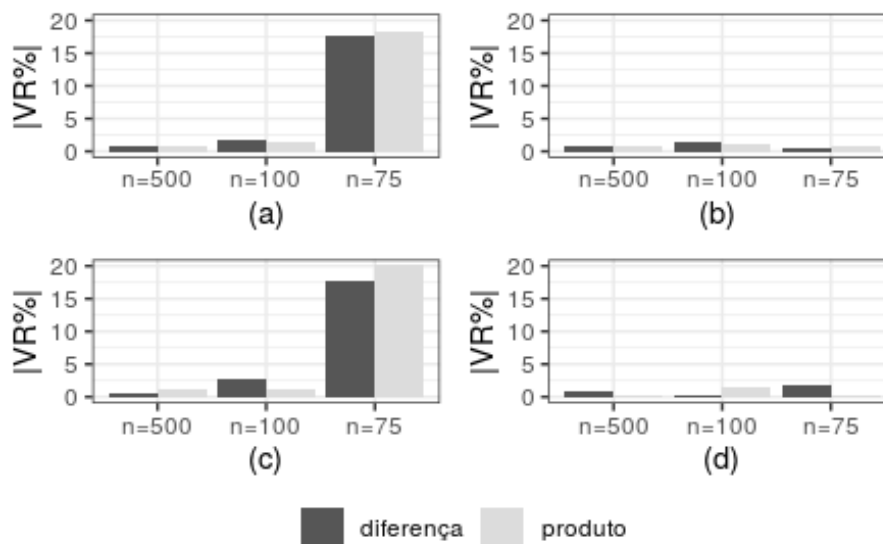
**Tabela 14:** Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados sem censura e  $\alpha = 0,2$ .

Coeficiente	Weibull			Cox		
	n=500	n=100	n=75	n=500	n=100	n=75
<i>Variável independente binária</i>						
$\hat{\alpha}\hat{\beta}$	0.0100	0.0100	0.0107	-0.0100	-0.0099	-0.0113
$\hat{c} - \hat{c}'$	0.0102	0.0100	0.0103	-0.0088	-0.0070	-0.0059
VR%( $\hat{\alpha}\hat{\beta}$ )	-0.0297	-0.2673	6.9673	0.0102	0.8241	-13.2545
VR%( $\hat{c} - \hat{c}'$ )	1.9280	-0.3949	3.2340	11.9496	30.4289	40.6843
$s_{1\hat{\alpha}\hat{\beta}}$	0.0110	0.0328	0.0420	0.0111	0.0341	0.0442
dp( $\hat{\alpha}\hat{\beta}$ )	0.0108	0.0315	0.0400	0.0109	0.0329	0.0425
CP%( $s_{1\hat{\alpha}\hat{\beta}}$ )	92.4	98.4	98.6	92.6	98.4	99.0
<i>Variável independente contínua</i>						
$\hat{\alpha}\hat{\beta}$	0.0102	0.0097	0.0100	-0.0102	-0.0097	-0.0101
$\hat{c} - \hat{c}'$	0.0102	0.0101	0.0095	-0.009	-0.0067	-0.0056
VR%( $\hat{\alpha}\hat{\beta}$ )	1.6341	-2.9870	0.0273	-2.2870	2.8109	-0.8643
VR%( $\hat{c} - \hat{c}'$ )	2.182	1.0724	-5.3904	9.7198	33.0066	43.9378
$s_{1\hat{\alpha}\hat{\beta}}$	0.0094	0.0239	0.0288	0.0096	0.0247	0.0302
dp( $\hat{\alpha}\hat{\beta}$ )	0.0097	0.0235	0.0277	0.0098	0.0241	0.0291
CP%( $s_{1\hat{\alpha}\hat{\beta}}$ )	95.6	97.8	98.5	95.4	98.4	98.7

**Tabela 15:** Comparação da estimação dos efeitos mediados pelos métodos do produto e da diferença dos coeficientes nos modelos Weibull e Cox para dados com censura e  $\alpha = 0,2$ .

Censura%	Weibull			Cox		
	n=500	n=100	n=75	n=500	n=100	n=75
<i>Variável independente binária</i>						
<b>25%</b>						
$\hat{\alpha}\hat{\beta}$	0.0097	0.0099	0.0126	-0.0098	-0.0098	-0.0125
$\hat{c} - \hat{c}'$	0.0096	0.0089	0.0101	-0.0087	-0.0063	-0.0057
VR%( $\hat{\alpha}\hat{\beta}$ )	-2.7418	-0.5619	25.8453	2.0726	1.998	-24.605
VR%( $\hat{c} - \hat{c}'$ )	-4.2493	-11.0524	0.5588	12.6757	37.1455	42.6413
$s_{1\hat{\alpha}\hat{\beta}}$	0.0125	0.0381	0.0493	0.0125	0.039	0.0511
dp( $\hat{\alpha}\hat{\beta}$ )	0.0116	0.0368	0.0466	0.0117	0.0372	0.0479
CP%( $s_{1\hat{\alpha}\hat{\beta}}$ )	94.3	98.9	99.1	94.4	99.3	99
<b>50%</b>						
$\hat{\alpha}\hat{\beta}$	0.0102	0.0105	0.0132	-0.0102	-0.0105	-0.0132
$\hat{c} - \hat{c}'$	0.01	0.0089	0.0097	-0.0094	-0.007	-0.0065
VR%( $\hat{\alpha}\hat{\beta}$ )	1.5058	4.5025	31.6389	-1.5788	-4.9551	-32.1425
VR%( $\hat{c} - \hat{c}'$ )	-0.2918	-11.3987	-3.1788	6.3878	30.3252	35.3322
$s_{1\hat{\alpha}\hat{\beta}}$	0.015	0.0465	0.06	0.0151	0.0473	0.061
dp( $\hat{\alpha}\hat{\beta}$ )	0.0149	0.0451	0.0569	0.0149	0.0454	0.0578
CP%( $s_{1\hat{\alpha}\hat{\beta}}$ )	95.7	98.7	99.4	96.1	99	99.3
<b>70%</b>						
$\hat{\alpha}\hat{\beta}$	0.0103	-5e-04	0.0128	-0.0103	-0.01	-0.0124
$\hat{c} - \hat{c}'$	0.0102	-0.0438	0.0094	-0.0097	-0.008	-0.0058
VR%( $\hat{\alpha}\hat{\beta}$ )	2.9858	-105.4243	27.8611	-2.9877	-0.1906	-23.9219
VR%( $\hat{c} - \hat{c}'$ )	1.5881	-538.1857	-5.9031	2.5872	20.3126	42.3835
$s_{1\hat{\alpha}\hat{\beta}}$	0.0193	<b>0.0665</b>	0.0776	0.0193	0.0609	0.0788
dp( $\hat{\alpha}\hat{\beta}$ )	0.019	0.2559	0.0711	0.0189	0.0573	0.0718

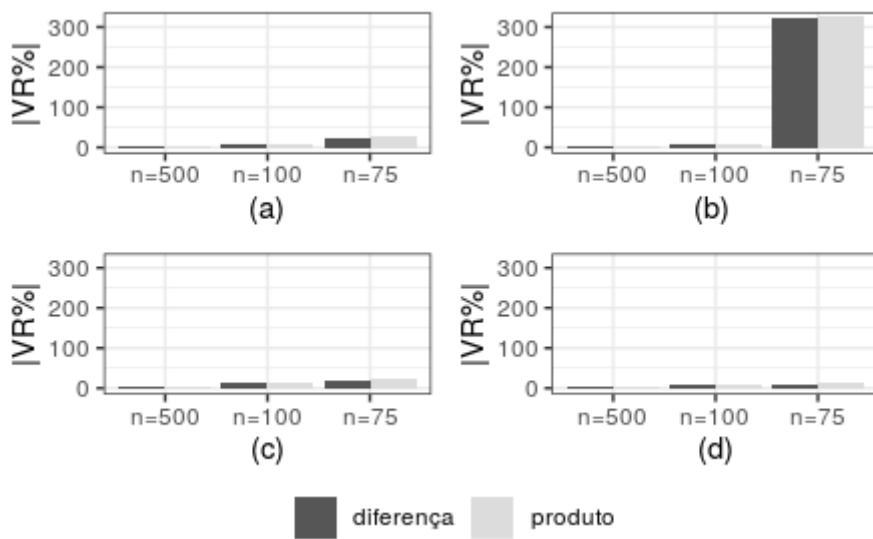
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	97.1	<b>99.5</b>	99.7	97.1	99.9	99.6
<i>Variável independente contínua</i>						
<b>25%</b>						
$\hat{\alpha}\hat{\beta}$	0.0102	0.01	0.01	-0.0103	-0.0098	-0.0099
$\hat{c} - \hat{c}'$	0.01	0.0098	0.009	-0.0093	-0.0069	-0.0054
$VR\%(\hat{\alpha}\hat{\beta})$	2.017	-0.4912	-0.1633	-2.8231	2.2393	0.5524
$VR\%(\hat{c} - \hat{c}')$	-0.03	-1.8427	-10.1706	7.4704	30.943	45.6512
$s_{1\hat{\alpha}\hat{\beta}}$	0.0109	0.0275	0.0336	0.011	0.0282	0.0348
$dp(\hat{\alpha}\hat{\beta})$	0.0111	0.0283	0.0337	0.0111	0.0289	0.0353
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	95.5	97.6	98.3	96.1	98	98.8
<b>50%</b>						
$\hat{\alpha}\hat{\beta}$	0.0097	0.009	0.0095	-0.0098	-0.0087	-0.0095
$\hat{c} - \hat{c}'$	0.0094	0.0084	0.0092	-0.0089	-0.0063	-0.0061
$VR\%(\hat{\alpha}\hat{\beta})$	-3.0463	-10.2787	-5.2095	2.2812	13.2078	4.6536
$VR\%(\hat{c} - \hat{c}')$	-6.4609	-16.4873	-8.0538	10.9515	37.0581	38.6532
$s_{1\hat{\alpha}\hat{\beta}}$	0.0133	0.0338	0.0414	0.0134	0.0342	0.0423
$dp(\hat{\alpha}\hat{\beta})$	0.0133	0.0355	0.0435	0.0134	0.0353	0.0448
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	95.8	95.8	98.5	96.2	98.3	98.3
<b>70%</b>						
$\hat{\alpha}\hat{\beta}$	0.0097	0.008	0.0124	-0.0098	-0.0082	-0.0113 -
$\hat{c} - \hat{c}'$	0.0094	0.0073	0.0303	-0.009	-0.0058	-0.0073
$VR\%(\hat{\alpha}\hat{\beta})$	-3.0449	-20.2861	23.5715	2.2744	17.5977	-12.5828
$VR\%(\hat{c} - \hat{c}')$	-6.3559	-26.7481	202.7431	9.9365	42.2711	26.9125
$s_{1\hat{\alpha}\hat{\beta}}$	0.0171	0.0443	<b>0.0554</b>	0.0172	0.0444	0.0546
$dp(\hat{\alpha}\hat{\beta})$	0.0172	0.0459	0.0612	0.0172	0.0457	0.0576
$CP\%(s_{1\hat{\alpha}\hat{\beta}})$	96.7	99.4	<b>99.2</b>	96.9	99.6	98.9



(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

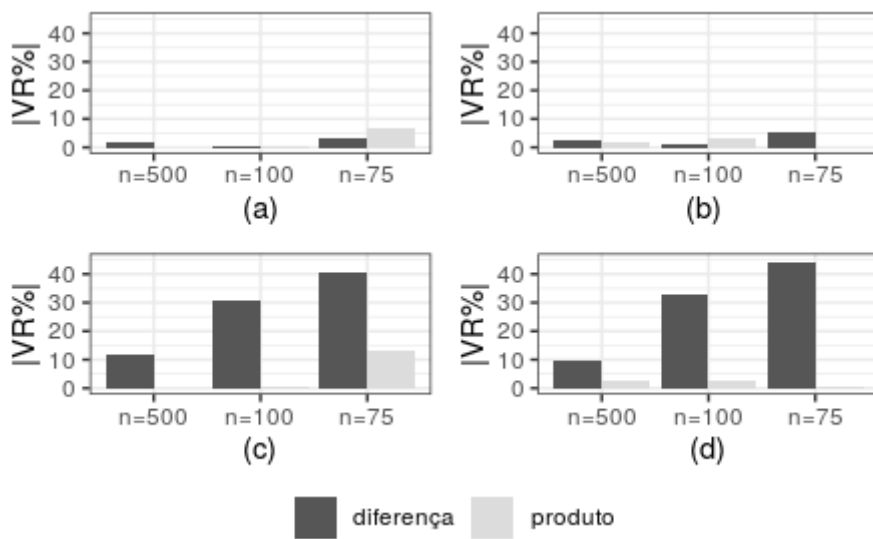
**Figura 11:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 25% de censura e  $\hat{\alpha}\hat{\beta} = 0,5$ , utilizando os modelos Weibull e de Cox.





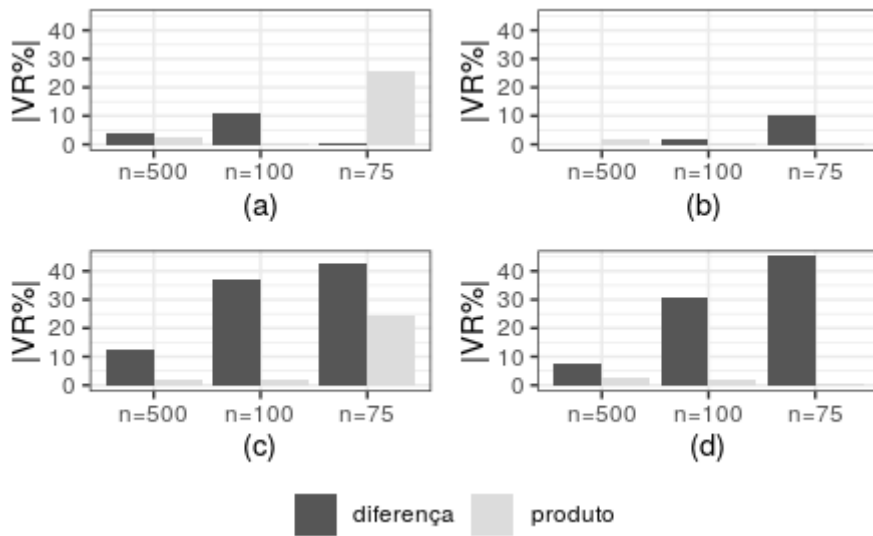
(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 12:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 70% de censura e  $\hat{\alpha}\hat{\beta} = 0,5$ , utilizando os modelos Weibull e de Cox.



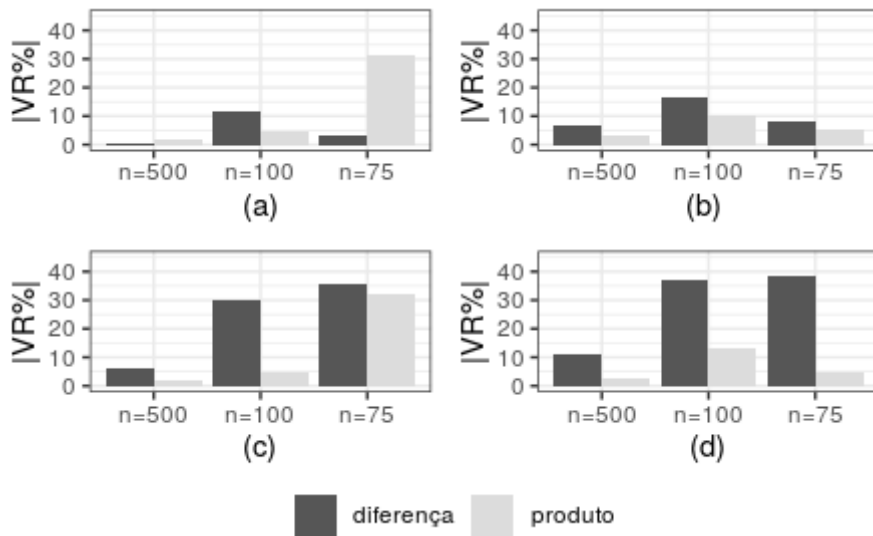
(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 13:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na ausência de censura e  $\hat{\alpha}\hat{\beta} = 0,1$ , utilizando os modelos Weibull e de Cox.



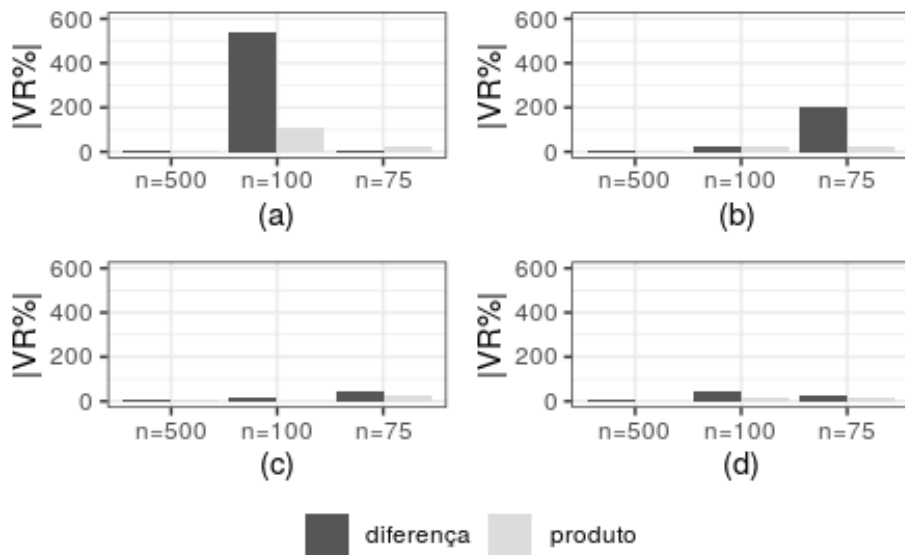
(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 14:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 25% de censura e  $\hat{\alpha}\hat{\beta} = 0,1$ , utilizando os modelos Weibull e de Cox.



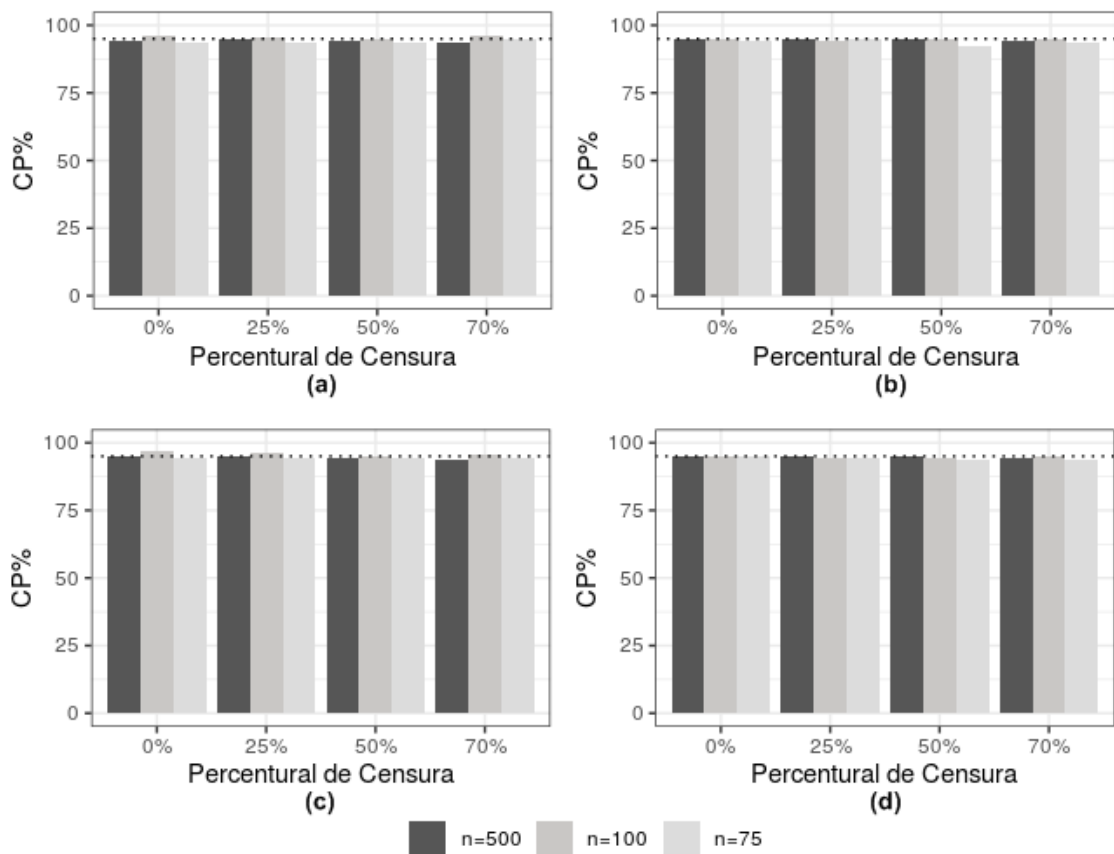
(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 15:** Módulo do viés relativo percentual ( $|VR\%|$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 50% de censura e  $\hat{\alpha}\hat{\beta} = 0,1$ , utilizando os modelos Weibull e de Cox.



(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 16:** Módulo do viés relativo percentual ( $|VR\%$ ) dos efeitos mediados estimados pelos métodos do produto e da diferença dos coeficientes, na presença de 70% de censura e  $\hat{\alpha} = 0,1$ , utilizando os modelos Weibull e de Cox.



(a): Modelo Weibull – Variável Independente Binária; (b): Modelo Weibull – Variável Independente Contínua; (c): Modelo de Cox – Variável Independente Binária; e (d): Modelo de Cox – Variável Independente Contínua.

**Figura 17:** Probabilidade de cobertura percentual (CP%) do intervalo de confiança construído baseado no erro-padrão estimado, com  $\alpha = 10$ .

## Apêndice 2

### Comandos utilizados no pacote estatístico R nas análises de dados

#### 1.A. Análise de Sobrevivência (Subseção 1.1)

##### Figuras

Figura 1: Curvas de sobrevivência

Figura 2: Resíduos padronizados

Figura 3: Resíduos padronizados de Schoenfeld

##### Tabelas

Tabela 1: Teste de igualdade das curvas de sobrevivência

Tabela 3: Ajuste do modelo de regressão log-normal

Tabela 4: Ajuste do modelo de Cox

Tabela 5: Ajuste modelo de regressão Weibull

Tabela 6: Resíduos de Schoenfeld

##### Figura 1: Curvas de sobrevivência

```
ekm <- survfit(Surv(tempo,cens)~V4)
plot(ekm,lty=c(1,4),mark.time = F, xlab = "Tempo até o desmame(meses)", ylab = "S(t)")
text(18.5,0.93,c("Dificuldades para Amamentar"),bty="n", cex=0.85)
legend(15.5,0.9,lty=c(4),c("Sim"),bty="n",cex=0.8)
legend(18.5,0.9,lty=c(1),c("Não"),bty="n",cex=0.8)
```

##### Figura 2: Resíduos de Cox-Snell

```
ajust1 <- survreg(Surv(tempo,cens)~V1+V3+V4+V6, dist = 'lognorm')
xb<-ajust1$coefficients[1]+ajust1$coefficients[2]*V1+ajust1$coefficients[3]*V3+
ajust1$coefficients[4]*V4+ ajust1$coefficients[5]*V6
sigma<-ajust1$scale
res<-(log(tempo)-(xb))/sigma # resíduos padronizados
resid<-exp(res) # exponencial dos resíduos padronizados
ekm<- survfit(Surv(resid,cens)~1)
resid<-ekm$time
sln<-pnorm(-log(resid))
par(mfrow=c(1,2))
plot(ekm$surv,sln, xlab="S(ei*): Kaplan-Meier",ylab="S(ei*): Log-normal padrão",pch=16)
plot(ekm,conf.int=F,mark.time=F,xlab="Resíduos (ei*)",ylab="Sobrevivência estimada",pch=16)
lines(resid,sln,lty=2)
legend(1.3,0.8,lty=c(1,2),c("Kaplan-Meier","Log-normal padrão"),cex=0.8,bty="n")
ei<- -log(1-pnorm(res)) # resíduos de Cox-Snell
```

```

ekm1<-survfit(Surv(ei,cens)~1)
t<-ekm1$time
st<-ekm1$surv
sexp<-exp(-t)

```

Figura 3: Resíduos padronizados de Schoenfeld

```

fit<-coxph(Surv(tempo,cens)~V1+V3+V4+V6,data=desmame,x = T,method="breslow")
plot(cox.zph(fit))

```

Tabela 1: Teste de igualdade das curvas de sobrevivência

```

survdifff(Surv(tempo,cens)~V1, rho = 0) #peso0-logrank
survdifff(Surv(tempo,cens)~V1, rho = 1) #peso1-wilcoxon

```

Tabela 3: Ajuste do modelo de regressão log-normal

```

ajust1 <- survreg(Surv(tempo,cens)~V1+V3+V4+V6, dist = 'lognorm')
summary(ajust1)

```

Tabela 4: Ajuste do modelo de Cox

```

fit<-coxph(Surv(tempo,cens)~V1+V3+V4+V6,data=desmame,x = T,method="breslow")
summary(fit)

```

Tabela 5: Ajuste modelo de regressão Weibull

```

ajust2 <- survreg(Surv(tempo,cens)~V1+V3+V4+V6, dist = 'weibull')
summary(ajust2)

```

Tabela 6: Resíduos de Schoenfeld

```

fit<-coxph(Surv(tempo,cens)~V1+V3+V4+V6,data=desmame,x = T,method="breslow")
cox.zph(fit, transform="identity") ## g(t) = t

```

## 1.B. Análise de Mediação (Subseção 2.1)

Figura 4: Diagrama de mediação

```

install.packages("psych")
library(psych)
data(Tal_Or)
mod4.4 <- mediate(reaction ~ cond + (pmi), data =Tal_Or)

```

Tabela 7: Estimação dos parâmetros de mediação

```

print(mod4.4, digits = 4, short = FALSE) #maior precisão e medidas

```

Tabela 8: Construção dos intervalos de 95% de confiança

```

dados <- Tal_Or
names(dados)
names(dados)[1] <- "X"
names(dados)[2] <- "M"
names(dados)[4] <- "eventtime"

```

```

ajuste1.c <- summary(lm(dados$eventtime~dados$X))
c <- ajuste1.c$coefficients[[2,1]]
sc <- ajuste1.c$coefficients[[2,2]]
ajuste2.cl.b <- summary(lm(dados$eventtime~dados$X+dados$M))
qmr <- ajuste2.cl.b$sigma
cl <- ajuste2.cl.b$coefficients[[2,1]]
scl <- ajuste2.cl.b$coefficients[[2,2]]
b <- ajuste2.cl.b$coefficients[[3,1]]
tb <- ajuste2.cl.b$coefficients[[3,3]]
sb <- ajuste2.cl.b$coefficients[[3,2]]
ajuste3.a <- summary(lm(dados$M~dados$X))
a <- ajuste3.a$coefficients[[2,1]]
ta <- ajuste3.a$coefficients[[2,3]]
sa <- ajuste3.a$coefficients[[2,2]]
a <- lm(dados$M~dados$X)$coefficients[[2]]
sx <- sd(dados$X)
vx <- sx^2

erro.ab.sobel <- sqrt((a^2*sb^2)+(b^2*sa^2))
erro.ab.t <- sqrt(ta^2 + tb^2)*a*b/(ta * tb)
erro.d <- sqrt(scl^2 + sc^2 - conta)

superior.ab.sobel <- a*b + 1.96*erro.ab.sobel
inferior.ab.sobel <- a*b - 1.96*erro.ab.sobel
superior.ab.t <- a*b + 1.96*erro.ab.t
inferior.ab.t <- a*b - 1.96*erro.ab.t
superior.d <- a*b + 1.96*erro.d
inferior.d <- a*b - 1.96*erro.d

```

### 1.C. Análise de Mediação em dados de Sobrevida (Subseção 3.1)

#### Figuras

Figura 5: Diagrama das relações na análise de mediação

Figura 6: Curvas de sobrevivência

Figura 7: Resíduos padronizados de Schoenfeld

Figura 8: Histograma do módulo do viés relativo percentual (|VR%|)

Figura 10: Histograma da probabilidade de cobertura percentual (CP%)

#### Tabelas

Tabela 9: Resíduos de Schoenfeld

Tabela 10: Ajuste do modelo AFT e Cox

Figura 5: Diagrama das relações na análise de mediação

```
data <- c(0, "", 0,
         0, 0, 0,
         "", "", 0)
M<- matrix (nrow=3, ncol=3, byrow = TRUE, data=data)
plot<- plotmat (M, pos=c(1,2),
               name= c( "SHPSTotal", "SKSTotal", "TotalFitBitSleepHrs"),
               box.type = "rect", box.size = 0.16, box.prop=0.5, curve=0)
```

Figura 6: Curvas de sobrevivência

```
ajuste <- survreg(Surv(TotalFitBitSleepHrs,cens)~1,data=dados, dist='weibull')
gamawei=1/ajuste$scale
alphawei=exp(ajuste$coef[1])
ekm <- survfit(Surv(TotalFitBitSleepHrs,cens)~1,data=dados)
st=ekm$surv
time=ekm$time
stk=ekm$surv
stw=exp(-(time/alphawei)^gamawei)
par(mfrow=c(1,2))
#### Lineariza??o da fun??o de sobreviv?ncia (figura 3.8):
plot(log(time),log(-log(st)),pch=16,xlab="log tempos",ylab="log[-log(S(t))]" )
### Figura 3.9 do livro do Colosimo e Giolo
plot(ekm, conf.int=F, xlab="tempos",ylab="S(t)" )
lines(c(0,time),c(1,stw),lty=2)
legend(1.6,0.2, lty=c(1,2),c("Kaplan-Meier", "Weibull"),bty="n", cex=0.8)
```

Figura 7: Resíduos padronizados de Schoenfeld

```
par(mfrow=c(1,2))
plot(cox.zph(ajuste.cox))
```

Figura 8: Histograma do módulo do viés relativo percentual (|VR%|)

```
require(ggplot2)
require(gridExtra)
library(ggpubr)
n <- c("1", "2", "3", "1", "2", "3")
metodo <- c("produto", "produto", "produto", "diferença", "diferença", "diferença")
#vr1 <- c(2.5954,2.0197,6.3390,2.6346,2.0222,6.2643) #WB
#vr2 <- c(0.7489,0.3930,1.2956,0.7599,0.3118,1.1872) #WC
#vr3 <- c(2.7540,0.5362,10.2072,2.1365,2.1548,7.8456) #CB
#vr4 <- c(1.0271,1.2114,3.2248,0.4827,0.2144,1.2848) #CC
var1 <- data.frame(vr1,n,metodo)
var2 <- data.frame(vr2,n,metodo)
var3 <- data.frame(vr3,n,metodo)
var4 <- data.frame(vr4,n,metodo)
gap1 <- aggregate(vr1 ~ n + metodo, data=var1, FUN=mean)
gap2 <- aggregate(vr2 ~ n + metodo, data=var2, FUN=mean)
gap3 <- aggregate(vr3 ~ n + metodo, data=var3, FUN=mean)
gap4 <- aggregate(vr4 ~ n + metodo, data=var4, FUN=mean)
cores <- c("#565656", "#dcdcdc")
par(mfcol=c(2,2))
```

```

fig1 <- ggplot(gap1, aes(x = n, y = vr1, fill = factor(metodo))) +
  geom_col(position = "dodge") + xlab("(a)") +
  coord_cartesian(ylim = c(0, 25))+
  ylab("|VR%|")+
  scale_x_discrete(labels=c("n=500","n=100","n=75"))+
  scale_fill_manual("", values = cores)+
  theme_bw()
fig2 <- ggplot(gap2, aes(x = n, y = vr2, fill = factor(metodo))) +
  geom_col(position = "dodge") + xlab("(b) ") +
  coord_cartesian(ylim = c(0, 25))+
  ylab("|VR%|")+
  scale_x_discrete(labels=c("n=500","n=100","n=75"))+
  scale_fill_manual("", values = cores)+
  theme_bw()
fig3 <- ggplot(gap3, aes(x = n, y = vr3, fill = factor(metodo))) +
  geom_col(position = "dodge") + xlab("(c)") +
  coord_cartesian(ylim = c(0, 25))+
  ylab("|VR%|")+
  scale_x_discrete(labels=c("n=500","n=100","n=75"))+
  scale_fill_manual("", values = cores)+
  theme_bw()
fig4 <- ggplot(gap4, aes(x = n, y = vr4, fill = factor(metodo))) +
  geom_col(position = "dodge") + xlab("(d)") +
  coord_cartesian(ylim = c(0, 25))+
  ylab("|VR%|")+
  scale_x_discrete(labels=c("n=500","n=100","n=75"))+
  scale_fill_manual("", values = cores)+
  theme_bw()
ggarrange(fig1, fig2, fig3, fig4, ncol=2, nrow=2, common.legend = TRUE, legend="bottom")

```

Figura 10: Histograma da probabilidade de cobertura percentual (CP%)

```

library(ggplot2)
library(gridExtra)
library(ggpubr)
n <- c("1","2","3","1","2","3","1","2","3","1","2","3")
cen <- c("0%","0%","0%","25%","25%","25%","50%","50%","50%","70%","70%","70%")
cp1 <- c(92.4,98.4,98.6,94.3,98.9,99.1,95.7,98.7,99.4,97.1,99.5,99.7) #WB
cp2 <- c(95.6,97.8,98.5,95.5,97.6,98.3,95.8,95.8,98.5,96.7,99.4,99.2) #WC
cp3 <- c(92.6,98.4,99.4,94.4,99.3,99.9,96.1,99.9,93.9,97.1,99.9,99.6) #CB
cp4 <- c(95.4,98.4,98.7,96.1,98.9,98.8,96.2,98.3,98.3,96.9,99.6,98.9) #CC
var1 <- data.frame(cp1,n,cen)
var2 <- data.frame(cp2,n,cen)
var3 <- data.frame(cp3,n,cen)
var4 <- data.frame(cp4,n,cen)
gap1 <- aggregate(cp1 ~ cen + n, data=var1, FUN=mean)
gap2 <- aggregate(cp2 ~ cen + n, data=var2, FUN=mean)
gap3 <- aggregate(cp3 ~ cen + n, data=var3, FUN=mean)
gap4 <- aggregate(cp4 ~ cen + n, data=var4, FUN=mean)
cores <- c("#565656","#c9c7c5","#dcdcdc")
par(mfcol=c(2,2))
fig1 <- ggplot(gap1, aes(x = cen, y = cp1, fill = factor(n))) +
  geom_col(position = "dodge") + xlab("Percentual de Censura") +
  geom_abline(intercept = 95, slope = 0, color = "black", lty = 3, lwd = 0.5)+

```



```

coord_cartesian(ylim = c(0, 100))+
ylab("CP%")+
scale_fill_manual("", values = cores, labels = c("n=500", "n=100", "n=75"))+
theme_bw()
fig2 <- ggplot(gap2, aes(x = cen, y = cp2, fill = factor(n))) +
  geom_col(position = "dodge") + xlab("Percentual de Censura") +
  geom_abline(intercept = 95, slope = 0, color = "black", lty = 3, lwd = 0.5)+
  coord_cartesian(ylim = c(0, 100))+
  ylab("CP%")+
  scale_fill_manual("", values = cores, labels = c("n=500", "n=100", "n=75"))+
  theme_bw()
fig3 <- ggplot(gap3, aes(x = cen, y = cp3, fill = factor(n))) +
  geom_col(position = "dodge") + xlab("Percentual de Censura") +
  geom_abline(intercept = 95, slope = 0, color = "black", lty = 3, lwd = 0.5)+
  coord_cartesian(ylim = c(0, 100))+
  ylab("CP%")+
  scale_fill_manual("", values = cores, labels = c("n=500", "n=100", "n=75"))+
  theme_bw()
fig4 <- ggplot(gap4, aes(x = cen, y = cp4, fill = factor(n))) +
  geom_col(position = "dodge") + xlab("Percentual de Censura") +
  geom_abline(intercept = 95, slope = 0, color = "black", lty = 3, lwd = 0.5)+
  coord_cartesian(ylim = c(0, 100))+
  ylab("CP%")+
  scale_fill_manual("", values = cores, labels = c("n=500", "n=100", "n=75"))+
  theme_bw()
ggarrange(fig1, fig2, fig3, fig4, ncol=2, nrow=2, common.legend = TRUE, legend="bottom")

```

Tabela 9: Resíduos de Schoenfeld

```
cox.zph(ajuste.cox, transform="identity")
```

Tabela 1-: Ajuste do modelo AFT e Cox

```

### Modelo Weibull
``{r warning=FALSE}
ajuste.m <- lm(SHPSTotal~SKSTotal ,data = dados)
a <- ajuste.m$coefficients[[2]]
ajuste.w<-survreg(Surv(TotalFitBitSleepHrs,cens)~dados$SKSTotal + dados$SHPSTotal,data=dados,dist
= 'weibull')
ajuste2.w<-survreg(Surv(TotalFitBitSleepHrs,cens)~dados$SKSTotal ,data=dados,dist = 'weibull')
b.w <- ajuste.w$coefficients[[3]]
c.linha.w <- ajuste.w$coefficients[[2]]
c.w <- ajuste2.w$coefficients[[2]]
produto.w <- a*b.w
dif.w <- c.w - c.linha.w
afastamento.w <- abs(produto.w-dif.w)
result.w <- data.frame(produto.w,dif.w,afastamento.w);result.w
gamawei=1/ajuste.w$scale
alphawei=exp(ajuste.w$coef)
sa.estimado <- summary(ajuste.m)$coefficients[[2,2]]
sb.w.estimado <- summary(ajuste.w)$table[[3,2]]
erro.ab.sobel.w.e <- sqrt((a^2*sb.w.estimado^2)+(b.w^2*sa.estimado^2))
ic.inf.sobel.w <- produto.w - 1.95*erro.ab.sobel.w.e
ic.sup.sobel.w <- produto.w + 1.95*erro.ab.sobel.w.e

```

```

ic.w <- data.frame(ic.inf.sobel.w,ic.sup.sobel.w);ic.w
...

### Modelo de Cox
``{r warning=FALSE}
ajuste.m <- lm(SHPSTotal~SKSTotal,data = dados)
a <- ajuste.m$coefficients[[2]]
ajuste.cox<-coxph(Surv(TotalFitBitSleepHrs,cens)~dados$SKSTotal + dados$SHPSTotal,data=dados,x =
T,method="breslow")
ajuste2.cox<-coxph(Surv(TotalFitBitSleepHrs,cens)~dados$SKSTotal ,data=dados,x =
T,method="breslow")
b.cox <- ajuste.cox$coefficients[[2]]
c.linha.cox <- ajuste.cox$coefficients[[1]]
c.cox <- ajuste2.cox$coefficients[[1]]
produto.cox <- a*b.cox
dif.cox <- c.cox - c.linha.cox
afastamento.cox <- abs(produto.cox-dif.cox)
result.cox <- data.frame(produto.cox,dif.cox,afastamento.cox);result.cox
sa.estimado <- summary(ajuste.m)$coefficients[[2,2]]
sb.cox.estimado <- summary(ajuste.cox)$coefficients[[2,3]]
erro.ab.sobel.cox.e <- sqrt((a^2*sb.cox.estimado^2)+(b.cox^2*sa.estimado^2))
ic.inf.sobel.cox <- produto.cox - 1.95*erro.ab.sobel.cox.e
ic.sup.sobel.cox <- produto.cox + 1.95*erro.ab.sobel.cox.e
ic.cox <- data.frame(ic.inf.sobel.cox,ic.sup.sobel.cox);ic.cox
...

```

### Apêndice 3

#### Código das simulações utilizando o software R

##### Estudos de Simulação

##### *2.A. Estudo de Simulação em Análise de Sobrevivência (Subseção 1.2)*

Simulação 1: Relação entre os modelos Weibull e Cox na modelagem de dados de sobrevivência, utilizando o pacote *simsurv*

Simulação 2: Relação entre os modelos Weibull e Cox na modelagem de dados de sobrevivência, utilizando a expressão do tempo Bender (2005)

##### *2.B. Estudo de Simulação em Análise de Mediação em Sobrevivência (Subseção 3.1)*

Simulação 3: Estimação dos efeitos mediados na modelagem de dados de sobrevivência utilizando os modelos Weibull e de Cox, na ausência de censuras

Simulação 4: Estimação dos efeitos mediados na modelagem de dados de sobrevivência utilizando os modelos Weibull e de Cox, na presença de censuras

## Simulação 1: Relação entre os modelos Weibull e Cox na modelagem de dados de sobrevivência, utilizando o pacote simsurv

```
simulando <- function(n,k,c){
  m=0
  dados <- NULL
  cox <- NULL
  aft <- NULL
  b.cox.a1 <- NULL
  beta.cox.e <- NULL
  b.cox.a <- NULL
  b.aft.a1 <- NULL
  beta.aft.e <- NULL
  b.aft.a <- NULL
  beta.cox.wr <- NULL
  beta.cox.wr.g <- NULL
  saida.censura <- NULL
  escala.estimada <- NULL
  while (m<k) {
    set.seed(123456+m)
    id <- 1:n
    trt = rbinom(n, 1, 0.5)
    covs <- data.frame(id,trt)
    gamma = 1.5
    escala = 1/gamma
    s1 <- simsurv(dist = "weibull", lambdas = 0.1, gammas = gamma, betas = c(trt = -0.5), x=covs, maxt =
NULL)
    dados <- data.frame(s1,trt)
    Q3 <- quantile(dados$eventtime, probs = 1-c)[[1]]
    C = rep(Q3,n)
    for (i in 1:length(dados$eventtime)){
      ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$status[i] <- 0)
      ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$eventtime[i] <- Q3)
    }
    saida.censura <- c(saida.censura,Q3)

#Ajustando os modelos
ajust.aft <- survreg(Surv(dados$eventtime,dados$status)~dados$trt,dist = "weibull")
beta.aft <- ajust.aft$coefficients[[2]]
ajust.cox <- coxph(Surv(dados$eventtime,dados$status)~dados$trt, x = T, method = "breslow")
beta.cox <- ajust.cox$coefficients[[1]]
escala.estimada <- c(escala.estimada, ajust.aft$scale)
ajust.aft.wr <- WeibullReg(Surv(dados$eventtime,dados$status)~dados$trt)
beta.cox.wr <- log(ajust.aft.wr$HR[1])
beta.cox.wr.g <- c(beta.cox.wr.g,beta.cox.wr)

# de cox para aft
for (i in 1:length(beta.aft)) {
  b.aft.a1[i] <- -(escala)*(beta.cox[i])
}
beta.aft.e <- c(beta.aft,beta.aft.e)
b.aft.a <- c(b.aft.a1,b.aft.a)
```

```

# de aft para cox
for (i in 1:length(beta.cox)) {
  b.cox.a1[i] <- -(beta.aft[i]/escala)
}
beta.cox.e <- c(beta.cox,beta.cox.e)
b.cox.a <- c(b.cox.a1,b.cox.a)
m = m +1
}
cox <- cbind(beta.cox.e,b.cox.a,beta.cox.wr.g)
aft <- cbind(beta.aft.e,b.aft.a)
return(list("estimates.cox"=cox,"estimates.aft"=aft,"censura"=saida.censura, "escala" =
escala.estimada))
}

```

Simulação 2: Relação entre os modelos Weibull e Cox na modelagem de dados de sobrevivência, utilizando a expressão do tempo Bender (2005)

```

simulando <- function(n,k,c){
  m=0
  dados <- NULL
  cox <- NULL
  aft <- NULL
  b.cox.a1 <- NULL
  beta.cox.e <- NULL
  b.cox.a <- NULL
  b.aft.a1 <- NULL
  beta.aft.e <- NULL
  b.aft.a <- NULL
  beta.cox.wr <- NULL
  beta.cox.wr.g <- NULL
  saida.censura <- NULL
  escala.estimada <- NULL
  while (m<k) {
    set.seed(123456+m)
    lambda = 0.1
    gamma = 1.5
    coveff = -0.5
    escala = 1/gamma
    trt <- rbinom(n , size = 1, prob = 0.5)
    u <- runif(n)
    eventtime <- (-log(u) / (lambda * exp(trt * coveff)))^(1 / gamma)
    id <- c(1:n)
    Q3 <- quantile(eventtime, probs = 1-c)[[1]]
    status <- rep(1,n)
    C = rep(Q3,n)
    dados <- data.frame(id,eventtime,status,trt)
    for (i in 1:length(eventtime)){
      ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$status[i] <- 0)
      ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$eventtime[i] <- Q3)
    }
    saida.censura <- c(saida.censura,Q3)

  }
  #Ajustando os modelos

```

```

ajust.aft <- survreg(Surv(dados$eventtime,dados$status)~dados$trt,dist = "weibull")
beta.aft <- ajust.aft$coefficients[[2]]
ajust.cox <- coxph(Surv(dados$eventtime,dados$status)~dados$trt, x = T, method = "breslow")
beta.cox <- ajust.cox$coefficients[[1]]
escala.estimada <- c(escala.estimada,ajust.aft$scale)
ajust.aft.wr <- WeibullReg(Surv(dados$eventtime,dados$status)~dados$trt)
beta.cox.wr <- log(ajust.aft.wr$HR[1])
beta.cox.wr.g <- c(beta.cox.wr.g,beta.cox.wr)

# de cox para aft
for (i in 1:length(beta.aft)) {
  b.aft.a1[i] <- -(escala)*(beta.cox[i])
}
beta.aft.e <- c(beta.aft,beta.aft.e)
b.aft.a <- c(b.aft.a1,b.aft.a)

# de aft para cox
for (i in 1:length(beta.cox)) {
  b.cox.a1[i] <- -(beta.aft[i]/escala)
}
beta.cox.e <- c(beta.cox,beta.cox.e)
b.cox.a <- c(b.cox.a1,b.cox.a)
m = m + 1
}
cox <- cbind(beta.cox.e,b.cox.a,beta.cox.wr.g)
aft <- cbind(beta.aft.e,b.aft.a)
return(list("estimates.cox"=cox,"estimates.aft"=aft,"censura"=saida.censura, "escala" =
escala.estimada))
}

```

Simulação 3: Estimação dos efeitos mediados na modelagem de dados de sobrevivência utilizando os modelos Weibull e de Cox, na ausência de censuras

```

banco <- function(n,lambda){
i3 = 100
a = 10
e3 <- rnorm(n,0,1)
X <- rbinom(n = n, size = 1, prob = 0.5) # p/ contínua X <- rnorm(n,0,1)
M <- i3 + a*X + e3
gamma=1
escala <- 1/gamma
c.linha.cox = -0.3
b.cox = -0.05
b.w <- -escala*b.cox
c.linha.w <- -escala*c.linha.cox
c.cox = a*b.cox + c.linha.cox
c.w = a*b.w + c.linha.w
coeficientes <- data.frame(a,b.w,c.w,c.linha.w,b.cox,c.cox,c.linha.cox)
u <- runif(n)
id <- c(1:n)
status <- rep(1,n)
eventtime <- (-log(u) / (lambda * exp((X * c.linha.cox)+(M * b.cox))))^(1 / gamma)
dados <- data.frame(id,eventtime,status,X,M)
return(list("dados"=dados,"coef"=coeficientes))
}

```

```

}

simulando <- function(n,k,lambd){

  resultado.w <- NULL
  resultado.cox <- NULL
  cp.sobel.w <- NULL
  cp.sobel.cox <- NULL
  cp.t.w <- NULL
  cp.t.cox <- NULL
  maxi <- NULL

  for (i in 1:k) {
    set.seed(123456+i)
    dados <- banco(n,lambd)$dados
    maxi <- c(max(dados$eventtime),maxi)
    ajuste.m <- lm(dados$M~dados$X)
    a.estimado <- ajuste.m$coefficients[[2]]
    produto.w <- banco(n,lambd)$coef[[1]]*banco(n,lambd)$coef[[2]]
    diferenca.w <- banco(n,lambd)$coef[[3]] - banco(n,lambd)$coef[[4]]
    produto.cox <- banco(n,lambd)$coef[[1]]*banco(n,lambd)$coef[[5]]
    diferenca.cox <- banco(n,lambd)$coef[[6]] - banco(n,lambd)$coef[[7]]
    ajuste.w <- survreg(Surv(dados$eventtime,dados$status)~(dados$X+dados$M),dist = "weibull")
    ajuste2.w <- survreg(Surv(dados$eventtime,dados$status)~(dados$X),dist = "weibull")
    ajuste.cox <- coxph(Surv(dados$eventtime,dados$status)~(dados$X+dados$M), x = T, method =
"breslow")
    ajuste2.cox <- coxph(Surv(dados$eventtime,dados$status)~(dados$X), x = T, method = "breslow")
    b.w.estimado <- ajuste.w$coefficients[[3]]
    c.w.estimado <- ajuste2.w$coefficients[[2]]
    c.linha.w.estimado <- ajuste.w$coefficients[[2]]
    b.cox.estimado <- ajuste.cox$coefficients[[2]]
    c.linha.cox.estimado <- ajuste.cox$coefficients[[1]]
    c.cox.estimado <- ajuste2.cox$coefficients[[1]]
    produto.w.e <- a.estimado*b.w.estimado
    diferenca.w.e <- c.w.estimado-c.linha.w.estimado
    produto.cox.e <- a.estimado*b.cox.estimado
    diferenca.cox.e <- c.cox.estimado-c.linha.cox.estimado
    vr.produto.w <- ((produto.w.e - produto.w)/abs(produto.w))*100
    vr.diferenca.w <- ((diferenca.w.e - diferenca.w)/abs(diferenca.w))*100
    vr.produto.cox <- ((produto.cox.e - produto.cox)/abs(produto.cox))*100
    vr.diferenca.cox <- ((diferenca.cox.e - diferenca.cox)/abs(diferenca.cox))*100
    sa.estimado <- summary(ajuste.m)$coefficients[[2,2]]
    ta.estimado <- summary(ajuste.m)$coefficients[[2,3]]
    sb.w.estimado <- summary(ajuste.w)$table[[3,2]]
    sb.cox.estimado <- summary(ajuste.cox)$coefficients[[2,3]]
    zb.w.estimado <- summary(ajuste.w)$table[[3,3]]
    zb.cox.estimado <- summary(ajuste.cox)$coefficients[[2,4]]
    erro.ab.sobel.cox.e <- sqrt((a.estimado^2*sb.cox.estimado^2)+(b.cox.estimado^2*sa.estimado^2))
    erro.ab.sobel.w.e <- sqrt((a.estimado^2*sb.w.estimado^2)+(b.w.estimado^2*sa.estimado^2))
    erro.ab.t.cox.e <- sqrt(ta.estimado^2 + zb.cox.estimado^2)*a.estimado*b.cox.estimado/(ta.estimado
* zb.cox.estimado)
    erro.ab.t.w.e <- sqrt(ta.estimado^2 + zb.w.estimado^2)*a.estimado*b.w.estimado/(ta.estimado
* zb.w.estimado)
    resultado.w1<-
data.frame(produto.w.e,diferenca.w.e,vr.produto.w,vr.diferenca.w,erro.ab.sobel.w.e,erro.ab.t.w.e)

```

```

    resultado.cox1<-
data.frame(produto.cox.e,diferenca.cox.e,vr.produto.cox,vr.diferenca.cox,erro.ab.sobel.cox.e,erro.ab.t
.cox.e)
    resultado.w <- rbind(resultado.w,resultado.w1)
    resultado.cox <- rbind(resultado.cox,resultado.cox1)
    ic.inf.sobel.w <- produto.w.e - 1.95*erro.ab.sobel.w.e
    ic.sup.sobel.w <- produto.w.e + 1.95*erro.ab.sobel.w.e
    ic.inf.sobel.cox <- produto.cox.e - 1.95*erro.ab.sobel.cox.e
    ic.sup.sobel.cox <- produto.cox.e + 1.95*erro.ab.sobel.cox.e
    ic.inf.t.w <- produto.w.e - 1.95*erro.ab.t.w.e
    ic.sup.t.w <- produto.w.e + 1.95*erro.ab.t.w.e
    ic.inf.t.cox <- produto.cox.e - 1.95*erro.ab.t.cox.e
    ic.sup.t.cox <- produto.cox.e + 1.95*erro.ab.t.cox.e
    ifelse(ic.inf.sobel.w<=produto.w && ic.sup.sobel.w>= produto.w,cp.sobel.w <-
c(1,cp.sobel.w),cp.sobel.w <- c(0,cp.sobel.w))
    ifelse(ic.inf.sobel.cox<= produto.cox && ic.sup.sobel.cox>= produto.cox,cp.sobel.cox <-
c(1,cp.sobel.cox),cp.sobel.cox <- c(0,cp.sobel.cox))
    ifelse(ic.inf.t.w<= produto.w && ic.sup.t.w>= produto.w,cp.t.w <- c(1,cp.t.w),cp.t.w <- c(0,cp.t.w))
    ifelse(ic.inf.t.cox<= produto.cox && ic.sup.t.cox>= produto.cox,cp.t.cox <- c(1,cp.t.cox),cp.t.cox <-
c(0,cp.t.cox))
  }

  m.p.w <- round(mean(resultado.w$produto.w.e),4)
  m.d.w <- round(mean(resultado.w$diferenca.w.e),4)
  m.vr.p.w <- round(mean(resultado.w$vr.produto.w),4)
  m.vr.d.w <- round(mean(resultado.w$vr.diferenca.w),4)
  m.sobel.w <- round(mean(resultado.w$erro.ab.sobel.w.e),4)
  m.t.w <- round(mean(resultado.w$erro.ab.t.w.e),4)
  dp.ab.w <- round(sd(resultado.w$produto.w.e),4)
  p.cp.sobel.w <- mean(cp.sobel.w)*100
  p.cp.t.w <- mean(cp.t.w)*100
  m.p.cox <- round(mean(resultado.cox$produto.cox.e),4)
  m.d.cox <- round(mean(resultado.cox$diferenca.cox.e),4)
  m.vr.p.cox <- round(mean(resultado.cox$vr.produto.cox),4)
  m.vr.d.cox <- round(mean(resultado.cox$vr.diferenca.cox),4)
  m.sobel.cox <- round(mean(resultado.cox$erro.ab.sobel.cox.e),4)
  m.t.cox <- round(mean(resultado.cox$erro.ab.t.cox.e),4)
  dp.ab.cox <- round(sd(resultado.cox$produto.cox.e),4)
  p.cp.sobel.cox <- mean(cp.sobel.cox)*100
  p.cp.t.cox <- mean(cp.t.cox)*100
  saida.w<-
data.frame(m.p.w,m.d.w,m.vr.p.w,m.vr.d.w,m.sobel.w,m.t.w,dp.ab.w,p.cp.sobel.w,p.cp.t.w)
  saida.cox<-
data.frame(m.p.cox,m.d.cox,m.vr.p.cox,m.vr.d.cox,m.sobel.cox,m.t.cox,dp.ab.cox,p.cp.sobel.cox,p.cp.t
.cox)
  return(list("result.w"=saida.w,"result.cox"=saida.cox,"max"=maxi))
}

```

Simulação 4: Estimação dos efeitos mediados na modelagem de dados de sobrevivência utilizando os modelos Weibull e de Cox, na presença de censuras

```

banco <- function(n,tm,lambda){
  i3 = 100
  a = 10
  e3 <- rnorm(n,0,1)

```



```

X <- rbinom(n = n, size = 1, prob = 0.5) # p/ continua
X <- rnorm(n,0,1)
M <- i3 + a*X + e3
gamma=1
escala <- 1/gamma
c.linha.cox = -0.3
b.cox = -0.05
b.w <- -escala*b.cox
c.linha.w <- -escala*c.linha.cox
c.cox = a*b.cox + c.linha.cox
c.w = a*b.w + c.linha.w
coeficientes <- data.frame(a,b.w,c.w,c.linha.w,b.cox,c.cox,c.linha.cox)
u <- runif(n)
id <- c(1:n)
status <- rep(1,n)
eventtime <- (-log(u) / (lambda * exp((X * c.linha.cox)+(M * b.cox))))^(1 / gamma)
dados <- data.frame(id,eventtime,status,X,M)
C <- runif(n,0,tm)
for (i in 1:length(eventtime)){
  ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$status[i] <- 0)
  ifelse(dados$eventtime[i] <= C[i], dados$status[i] <- 1 ,dados$eventtime[i] <- C[i])
}
return(list("dados"=dados,"coef"=coeficientes))
}

simulando <- function(n,k,lambda,tm){
  censura <- NULL
  resultado.w <- NULL
  resultado.cox <- NULL
  cp.sobel.w <- NULL
  cp.sobel.cox <- NULL
  cp.t.w <- NULL
  cp.t.cox <- NULL

  for (i in 1:k) {
    set.seed(123456+i)
    dados <- banco(n,tm,lambda)$dados
    if (table(dados$status)[[1]]==n) {
      remove(dados)
      k=k+1
    }
    else{
      censura <- c(censura,table(dados$status)[[1]]/n*100)
      ajuste.m <- lm(dados$M~dados$X)
      a.estimado <- ajuste.m$coefficients[[2]]
      produto.w <- banco(n,tm,lambda)$coef[[1]]*banco(n,tm,lambda)$coef[[2]]
      diferenca.w <- banco(n,tm,lambda)$coef[[3]] - banco(n,tm,lambda)$coef[[4]]
      produto.cox <- banco(n,tm,lambda)$coef[[1]]*banco(n,tm,lambda)$coef[[5]]
      diferenca.cox <- banco(n,tm,lambda)$coef[[6]] - banco(n,tm,lambda)$coef[[7]]
      ajuste.w <- survreg(Surv(dados$eventtime,dados$status)~(dados$X+dados$M),dist = "weibull")
      ajuste2.w <- survreg(Surv(dados$eventtime,dados$status)~(dados$X),dist = "weibull")
      ajuste.cox <- coxph(Surv(dados$eventtime,dados$status)~(dados$X+dados$M), x = T, method =
"breslow")
      ajuste2.cox <- coxph(Surv(dados$eventtime,dados$status)~(dados$X), x = T, method = "breslow")
      b.w.estimado <- ajuste.w$coefficients[[3]]
      c.w.estimado <- ajuste2.w$coefficients[[2]]

```

```

c.linha.w.estimado <- ajuste.w$coefficients[[2]]
b.cox.estimado <- ajuste.cox$coefficients[[2]]
c.linha.cox.estimado <- ajuste.cox$coefficients[[1]]
c.cox.estimado <- ajuste2.cox$coefficients[[1]]
produto.w.e <- a.estimado*b.w.estimado
diferenca.w.e <- c.w.estimado-c.linha.w.estimado
produto.cox.e <- a.estimado*b.cox.estimado
diferenca.cox.e <- c.cox.estimado-c.linha.cox.estimado
vr.produto.w <- ((produto.w.e - produto.w)/abs(produto.w))*100
vr.diferenca.w <- ((diferenca.w.e - diferenca.w)/abs(diferenca.w))*100
vr.produto.cox <- ((produto.cox.e - produto.cox)/abs(produto.cox))*100
vr.diferenca.cox <- ((diferenca.cox.e - diferenca.cox)/abs(diferenca.cox))*100
sa.estimado <- summary(ajuste.m)$coefficients[[2,2]]
ta.estimado <- summary(ajuste.m)$coefficients[[2,3]]
sb.w.estimado <- summary(ajuste.w)$table[[3,2]]
sb.cox.estimado <- summary(ajuste.cox)$coefficients[[2,3]]
zb.w.estimado <- summary(ajuste.w)$table[[3,3]]
zb.cox.estimado <- summary(ajuste.cox)$coefficients[[2,4]]
erro.ab.sobel.cox.e <- sqrt((a.estimado^2*sb.cox.estimado^2)+(b.cox.estimado^2*sa.estimado^2))
erro.ab.sobel.w.e <- sqrt((a.estimado^2*sb.w.estimado^2)+(b.w.estimado^2*sa.estimado^2))
erro.ab.t.cox.e <- sqrt(ta.estimado^2 +
zb.cox.estimado^2)*a.estimado*b.cox.estimado/(ta.estimado * zb.cox.estimado)
erro.ab.t.w.e <- sqrt(ta.estimado^2 + zb.w.estimado^2)*a.estimado*b.w.estimado/(ta.estimado *
zb.w.estimado)
resultado.w1<-
data.frame(produto.w.e,diferenca.w.e,vr.produto.w,vr.diferenca.w,erro.ab.sobel.w.e,erro.ab.t.w.e)
resultado.cox1<-
data.frame(produto.cox.e,diferenca.cox.e,vr.produto.cox,vr.diferenca.cox,erro.ab.sobel.cox.e,erro.ab.t
.cox.e)
resultado.w <- rbind(resultado.w,resultado.w1)
resultado.cox <- rbind(resultado.cox,resultado.cox1)
ic.inf.sobel.w <- produto.w.e - 1.95*erro.ab.sobel.w.e
ic.sup.sobel.w <- produto.w.e + 1.95*erro.ab.sobel.w.e
ic.inf.sobel.cox <- produto.cox.e - 1.95*erro.ab.sobel.cox.e
ic.sup.sobel.cox <- produto.cox.e + 1.95*erro.ab.sobel.cox.e
ic.inf.t.w <- produto.w.e - 1.95*erro.ab.t.w.e
ic.sup.t.w <- produto.w.e + 1.95*erro.ab.t.w.e
ic.inf.t.cox <- produto.cox.e - 1.95*erro.ab.t.cox.e
ic.sup.t.cox <- produto.cox.e + 1.95*erro.ab.t.cox.e
ifelse(ic.inf.sobel.w<=produto.w && ic.sup.sobel.w>= produto.w,cp.sobel.w <-
c(1,cp.sobel.w),cp.sobel.w <- c(0,cp.sobel.w))
ifelse(ic.inf.sobel.cox<= produto.cox && ic.sup.sobel.cox>= produto.cox,cp.sobel.cox <-
c(1,cp.sobel.cox),cp.sobel.cox <- c(0,cp.sobel.cox))
ifelse(ic.inf.t.w<= produto.w && ic.sup.t.w>= produto.w,cp.t.w <- c(1,cp.t.w),cp.t.w <- c(0,cp.t.w))
ifelse(ic.inf.t.cox<= produto.cox && ic.sup.t.cox>= produto.cox,cp.t.cox <- c(1,cp.t.cox),cp.t.cox <-
c(0,cp.t.cox))
}
}

m.p.w <- round(mean(resultado.w$produto.w.e),4)
m.d.w <- round(mean(resultado.w$diferenca.w.e),4)
m.vr.p.w <- round(mean(resultado.w$vr.produto.w),4)
m.vr.d.w <- round(mean(resultado.w$vr.diferenca.w),4)
m.sobel.w <- round(mean(resultado.w$erro.ab.sobel.w.e),4)
m.t.w <- round(mean(resultado.w$erro.ab.t.w.e),4)

```

```

dp.ab.w <- round(sd(resultado.w$produto.w.e),4)
p.cp.sobel.w <- mean(cp.sobel.w)*100
p.cp.t.w <- mean(cp.t.w)*100
m.p.cox <- round(mean(resultado.cox$produto.cox.e),4)
m.d.cox <- round(mean(resultado.cox$diferenca.cox.e),4)
m.vr.p.cox <- round(mean(resultado.cox$vr.produto.cox),4)
m.vr.d.cox <- round(mean(resultado.cox$vr.diferenca.cox),4)
m.sobel.cox <- round(mean(resultado.cox$erro.ab.sobel.cox.e),4)
m.t.cox <- round(mean(resultado.cox$erro.ab.t.cox.e),4)
dp.ab.cox <- round(sd(resultado.cox$produto.cox.e),4)
p.cp.sobel.cox <- mean(cp.sobel.cox)*100
p.cp.t.cox <- mean(cp.t.cox)*100
saida.w<-
data.frame(m.p.w,m.d.w,m.vr.p.w,m.vr.d.w,m.sobel.w,m.t.w,dp.ab.w,p.cp.sobel.w,p.cp.t.w)
saida.cox<-
data.frame(m.p.cox,m.d.cox,m.vr.p.cox,m.vr.d.cox,m.sobel.cox,m.t.cox,dp.ab.cox,p.cp.sobel.cox,p.cp.t.
cox)
p.c <- mean(censura)
return(list("result.w"=saida.w,"result.cox"=saida.cox,"perc.c"=p.c))
}

```