# Análise de sobrevivência e confiabilidade

## Modelos de regressão

**Prof. Paulo Cerqueira Jr**
**Programa de Pós-Graduação em Matemática e Estatística - PPGME**
**Instituto de Ciências Exatas e Naturais - ICEN**

https://github.com/paulocerqueirajr

# Introdução

# Introdução

- Estudos clínicos/industriais/sociais, em geral, envolvem covariáveis que podem estar relacionadas com o tempo de sobrevivência.
- Usualmente, o objetivo do estudo está relacionado a estas covariáveis.
- As covariáveis podem ser:
  - em estudos clínicos: gênero, idade, pressão arterial, presença de diabetes, etc;
  - em estudo industriais: temperatura, dureza do material, voltagem, etc.

# Introdução

## Técnicas não-paramétricas

- As técnicas não-paramétricas são limitadas na presença de covariáveis.
- De forma a utilizar técnicas não-paramétricas necessitamos dividir em estratos de acordo com as categorias dessas covariáveis.
- Isto gera um número grande de estratos que podem conter poucas, ou talvez nenhuma observação e, portanto, impossibilita a incorporação de covariáveis, por exemplo, contínuas.
- Covariáveis são acomodadas, naturalmente, em um modelo com `estrutura regressão`.
- Modelos do tipo GAM (Generalized Additive Models) pode ser uma alternativa não-paramétrica para estas limitações.

# Introdução

## Modelagem

- Tem-se duas classes de modelos de regressão para análise de dados de sobrevivência:

1. Modelos paramétricos ou de tempos de vida acelerados:

$$T = \exp\left(X^{'}\beta\right)T^{'} \text{ ou } S(t \mid x) = S_0\left[t/\exp(\beta x)\right]$$

2. Modelo semi-paramétrico ou de modelo de taxas de falha proporcionais ou, simplesmente, modelo de Cox:

$$h(t \mid X) = h_0(t)\exp\left(X^{'}\beta\right) \text{ ou } S(t \mid x) = \left[S_0(t)\right]^{\exp(\beta x)}$$

- Em que:

$X^{'}\beta$ : Preditor linear.
$h_0(t)$ : é a função de taxa de falha basal.

# Modelo de tempos de vida acelerados

# Modelo de tempos de vida acelerados (MTVA)

- O MTVA pode ser estendido para a situação em que $p$ covariáveis são medidas para cada indivíduos no estudo.
- A função de sobrevivência e taxa de falha são:

$$S(t \mid x) = S_0\left[t/\exp(\mathbf{X}^{'}\beta)\right] \text{ e } h(t) = \exp(\mathbf{X}^{'}\beta)h_0\left[t/\exp\left(\mathbf{X}^{'}\beta\right)\right].$$

$S_0$ e $h_0$ são funções de base quando $x = 0$.

- De forma alternativa, os tempos de sobrevivência assume que a relação entre $\log T$ e $X$ é linear e pode ser escrito como

$$\log T = \mathbf{X}^{'}\beta + \sigma\epsilon,$$

$\beta = \{\beta_k : k = 0, 1, \ldots, p\}$ e $X^{'} = \{1, x_1, x_2, \ldots, x_p\}, \sigma > 0$ é um parâmetro de escala e $\epsilon$ é o termo aleatório com função densidade g e função de sobrevivência G.

# Modelo de tempos de vida acelerados (MTVA)

- As distribuições usuais para o tempo até a falha é escala-forma.
- Desta forma, modelamos $Y = \log T$ que é um modelo locação-escala.

**Características Gerais:**

- O modelo paramétrico específica um efeito multiplicativo das covariáveis em $T$.
- O papel de $X$ é acelerar ou desacelerar o tempo até a falha.
- Os parâmetros do modelo são estimados pelo método de máxima verossimilhança.
- A adequação do modelo ajustado é realizada utilizando os resíduos:

$$\hat{\nu}_i = \frac{Y_i - X_i'\hat{\beta}}{\hat{\sigma}}, \quad i = 1, \ldots, n.$$

# Modelo de tempos de vida acelerados (MTVA)

- Covariáveis agem multiplicativamente na escala do tempo.
- Modelo interpretado em termos da `velocidade de progressão do tempo` de acompanhamento.
- Forma do Modelo para uma única covariável representada por dois grupos (0 e 1):

$$S_1(t) = S_0(\phi t).$$

- Se $\phi < 1$, o grupo 1 tem o tempo acelerado quando comparado com o 0.
- Se $\phi < 1$ e o evento for morte, o grupo 1 é melhor que o 0 no sentido que caminha para o evento de forma mais devagar.

# Modelo de tempos de vida acelerados (MTVA)

**Interpretação**

- Forma do Modelo para uma única covariável representada por dois grupos (0 e 1):

$$S_1(t) = S_0(\phi t).$$

  - Interpretação em termos de percentis, em especial tempos medianos:

$$S_1(t^1_{0.5}) = S_0(t^0_{0.5}) = 0.5$$

ou

$$S_1(t^1_{0.5}) = S_0(\phi t^1_{0.5}) \text{ ou } t^0_{0.5} = t^1_{0.5} \rightarrow \frac{t^0_{0.5}}{t^1_{0.5}}.$$

que representa a razão dos tempos medianos.

# Modelo de tempos de vida acelerados (MTVA)

## Modelo de regressão Exponencial

Assumindo a estrutura do MTVA, com $\sigma = 1$, temos

$$\log T_i = \mathbf{X}_i^{'}\beta + \epsilon_i = \mu_i + \epsilon_i,$$

em que $\epsilon_i^{\prime}s$ são iid com distribuição do valor extremo,

$$g(\epsilon) = \exp[\epsilon - \exp(\epsilon)]$$
$$G(\epsilon) = \exp[-\exp(\epsilon)]$$

# Modelo de tempos de vida acelerados (MTVA)

## Modelo de regressão Exponencial

Assim, na escala $\log$, $Y$ tem uma distribuição do valor extremo padrão com parâmetro de locação $X^{'}\beta$:

$$S(y \mid X) = \exp[-\exp(y - X^{'}\beta)]$$

ou na escala original, T tem uma distribuição exponencial com parâmetro de escala $\alpha = \exp(X^{'}\beta)$.

A função de sobrevivência é dada por

$$S(t \mid X) = \exp\left[-\frac{t}{\exp(X^{'}\beta)}\right].$$

# Modelo de tempos de vida acelerados (MTVA)

## Modelo de regressão de Weibull

- De forma equivalente podemos definir o modelo Weibull:

1. $T$ tem uma distribuição Weibull com parâmetro de escala $\alpha = \exp(X^{'}\beta)$ e de forma $\gamma$.

$$S(t \mid X) = \exp\left[-\left(\frac{t}{\exp(X^{'}\beta)}\right)^{\gamma}\right].$$

2. $Y$ tem distribuição do valor extremo padrão com parâmetros de locação $X^{'}\beta$ e escala $\sigma = 1/\gamma$.

$$S(y \mid X) = \exp\left[-\exp\left(\frac{y - X^{'}\beta}{\sigma}\right)\right].$$

# Modelo de tempos de vida acelerados (MTVA)

## Função de verossimilhança

- A função de verossimilhança para uma amostra de tamanho $n$, sob o mecanismo de censura não-informativo, é dada por:

$$
\begin{aligned}
L(\beta, \sigma) &= \prod_{i=1}^{n} f(y_i \mid X_i, \beta, \sigma)^{\delta_i} S(y_i \mid X_i, \beta, \sigma)^{1-\delta_i} \\
&\text{ou} \\
L(\beta, \gamma) &= \prod_{i=1}^{n} f(t_i \mid X_i, \beta, \gamma)^{\delta_i} S(t_i \mid X_i, \beta, \gamma)^{1-\delta_i}
\end{aligned}
$$

em que $\delta_i$ é o indicador de falha para a $i-$ésima observação.

- Valem todas as `propriedades`, para grandes amostras, do EMV e das estatísticas de teste.

# Modelo de tempos de vida acelerados (MTVA)

## Interpretação dos parâmetros

- Importante interpretar na escala de $T$. Observe sempre que usamos uma escala transformada (logarítmica) na modelagem estatística.
- Lembre que:

$$E(\log T) \neq \log E(T)$$

- `Interpretação`: Razão de tempos medianos $= exp(\beta)$.
- `Exemplo`: $\exp(\hat{\beta}) = 2$.

Isto significa que o tempo mediano de um grupo é `duas vezes` o do outro grupo (mantendo fixa as demais covariáveis).

# Modelo de tempos de vida acelerados (MTVA)

## Adequação do modelo

- Uma avaliação da adequação do modelo ajustado é parte fundamental da análise dos dados.
- No modelo de regressão linear usual, uma análise gráfica dos resíduos é usada para esta finalidade.
- Técnicas gráficas, que fazem uso dos diferentes resíduos propostos são, em particular, bastante utilizadas para examinar diferentes aspectos do modelo.
- Um desses aspectos é o de avaliar, por meio dos resíduos, a distribuição dos erros.

# Modelo de tempos de vida acelerados (MTVA)

## Resíduos de Cox-Snell

- Os resíduos de Cox-Snell (1968) são definidos como:

$$\hat{e}_i = \hat{H}(t_i \mid X_i)$$

em que $\hat{H}(\cdot)$ é a função de taxa de falha acumulada do modelo em questão.

- Modelos paramétricos:
  - Exponencial: $\hat{e}_i = \left[t_i \exp\{X_i\hat{\beta}\}\right]$.
  - Weibull: $\hat{e}_i = \left[t_i \exp\{X_i\hat{\beta}\}\right]^{\hat{\gamma}}$.
  - Lognormal: $\hat{e}_i = -\log\left[1 - \Phi\left(\frac{\log(t_i) - X_i\hat{\beta}}{\hat{\sigma}}\right)\right]$.
- Se o modelo for adequado os $\hat{e}_i's$ têm uma distribuição exponencial padrão $(\alpha = 1)$.

# Modelo de tempos de vida acelerados (MTVA)

## Resíduos de Padronizados

- A definição dos resíduos padronizados:

$$\hat{\nu}_i = \frac{y_i - X_i\hat{\beta}}{\hat{\sigma}}, \quad i = 1, \ldots, n.$$

- Os resíduos vêm de uma população homogênea. Observe que os resíduos de observações censuradas, também são censurados.
- Comparação dos resíduos do modelo proposto com o Kaplan-Meier.
- Devemos usar os resíduos na escala original, $\exp(\hat{\nu}_i)$, para compará-los com o `Kaplan-Meier`.

# Modelo de tempos de vida acelerados (MTVA)

## Resíduos martingais

- Os resíduos martingais são definidos por

$$\hat{m}_i = \delta_i - \hat{e}_i, \quad i = 1, \ldots, n.$$

em que $\delta_i$ é a variável indicadora de falha e $\hat{e}_i$ os resíduos de Cox-Snell.

- São uma `ligeira modificação` dos resíduos de Cox-Snell.
- São vistos como uma estimativa do número de falhas em excesso observada nos dados mas não predito pelo modelo.
- Os mesmos são usados, em geral, para examinar a melhor forma funcional (linear, quadrática etc.) para uma dada covariável.

# Modelo de tempos de vida acelerados (MTVA)

## Resíduos Deviance

- Os resíduos deviance nos modelos de regressão paramétricos são definidos por

$$\hat{d}_i = \text{sinal}(\hat{m}_i)[-2\left(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i)\right)]^{1/2}.$$

- São uma tentativa de tornar os resíduos martingale mais simétricos em torno de zero.
- Facilitam, em geral, a detecção de pontos atípicos (outliers).
- Se o `modelo for apropriado`, estes resíduos deveriam apresentar um `comportamento aleatório em torno de zero`.

# Modelo de tempos de vida acelerados (MTVA)

## Testes de significância e critérios de informação

- De forma adicional, pode-se realizar os testes de significância e os critérios de informação.
- O teste da razão de verossimilhanças: avaliar se o modelo em questão é um modelo adequado.
- Critérios de informação de Akaike e bayesiano: Escolher o modelo que apresentar o menor valor de critério.

# Modelo de tempos de vida acelerados (MTVA)

## Dados de Leucemia

- 17 pacientes com leucemia.
- `Resposta:` tempo (semanas) do diagnóstico até a morte do paciente.
- `Objetivo:` modelar a resposta em termos da contagem de glóbulos brancos (WBC) no diagnóstico.
- Covariável `x = log10 WBC`.

# Modelo de tempos de vida acelerados (MTVA)

## Dados de Leucemia

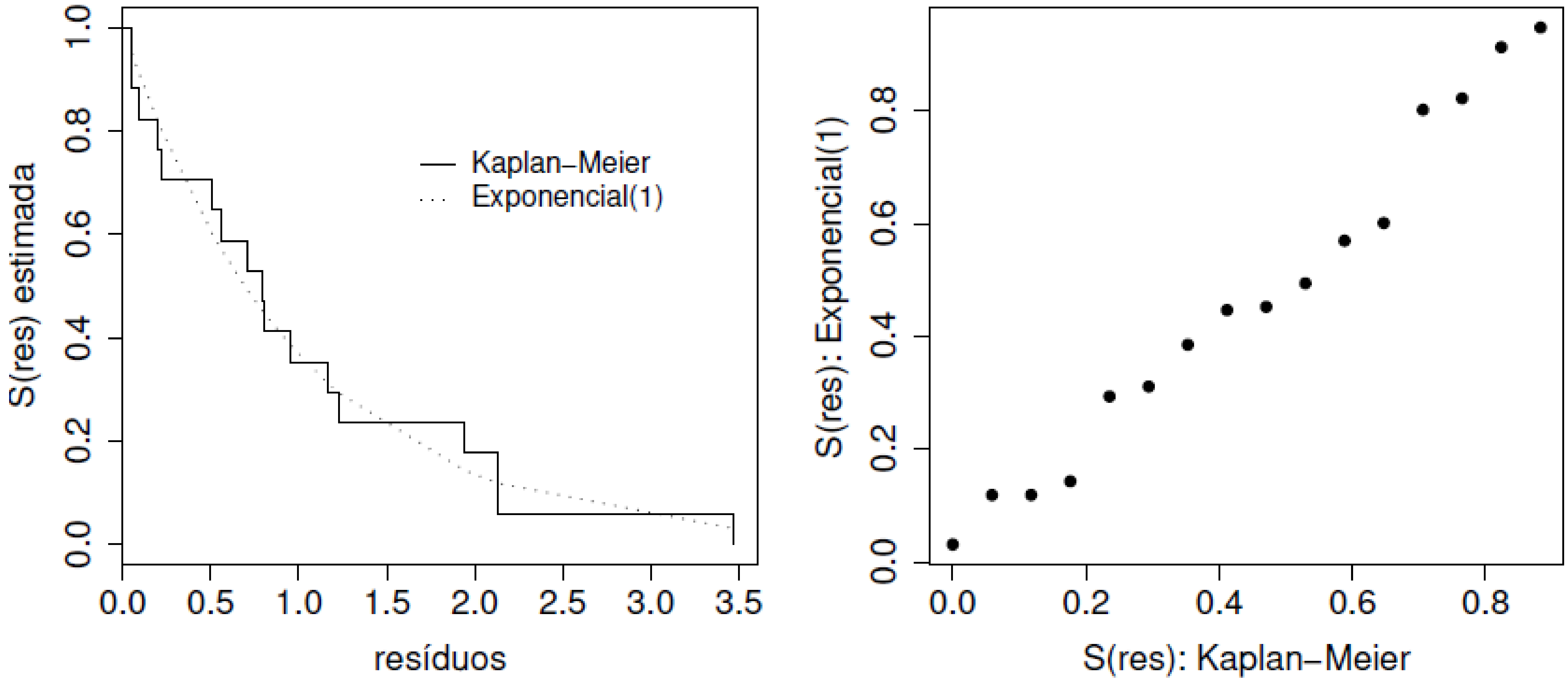
- Este banco de dados foi exaustivamente analisado na literatura utilizando o modelo de regressão exponencial.

| Temos | Exponencial | Weibull | Log-normal |
|---|---|---|---|
| $\hat{\beta}_0$ | 8,48 | 8,44 | |
| $\hat{\beta}_1$ | -1,11 | -1,10 | |
| $\gamma$ | 1 | 1,02 | |
| $\ell(\theta)$ | -83,88 | -83,87 | -83,76 |

- Outros modelos mais flexíveis?
- O ajuste pode ser feito usando a função `survreg()` do pacote `survival`.

# Modelo de tempos de vida acelerados (MTVA)

## Dados de Leucemia




Modelo Exponencial

# Modelo de tempos de vida acelerados (MTVA)

## Interpretação do ajuste

- $X$: $log_{10}$ WBC;
- Interpretação na escala original WBC é multiplicativa;
- Defina $p$: proporção de aumento/redução em WBC;
- Temos pelo ajuste do modelo exponencial que $\hat{\beta}_1 = -1,11$;
- Logo, a razão dos tempos medianos:

$$\hat{\text{RTM}} = \exp\Big(\hat{\beta}_1 \times \log_{10}(p)\Big).$$

Assim, temos:

- $p = 1,1$ (aumento de 10%) - $\hat{\text{RTM}} = \exp(-1,11 \times \log_{10}(1,1)) = 0,96.$
- $p = 1,2$ (aumento de 20%) - $\hat{\text{RTM}} = \exp(-1,11 \times \log_{10}(1,2)) = 0,92.$
- $p = 0,9$ (redução de 10%) - $\hat{\text{RTM}} = \exp(-1,11 \times \log_{10}(0,9)) = 1,05.$

# Modelo de tempos de vida acelerados (MTVA)

## Curvas de sobrevivência

A função de sobrevivência estimada é dada por

$$\hat{S}(t \mid X) = \exp\left\{-\frac{t}{\exp(8,48 - 1,11 * X)}\right\}.$$

Para dois valores de $X = \{3, 4\}$ no tempo $t = 100$, temos

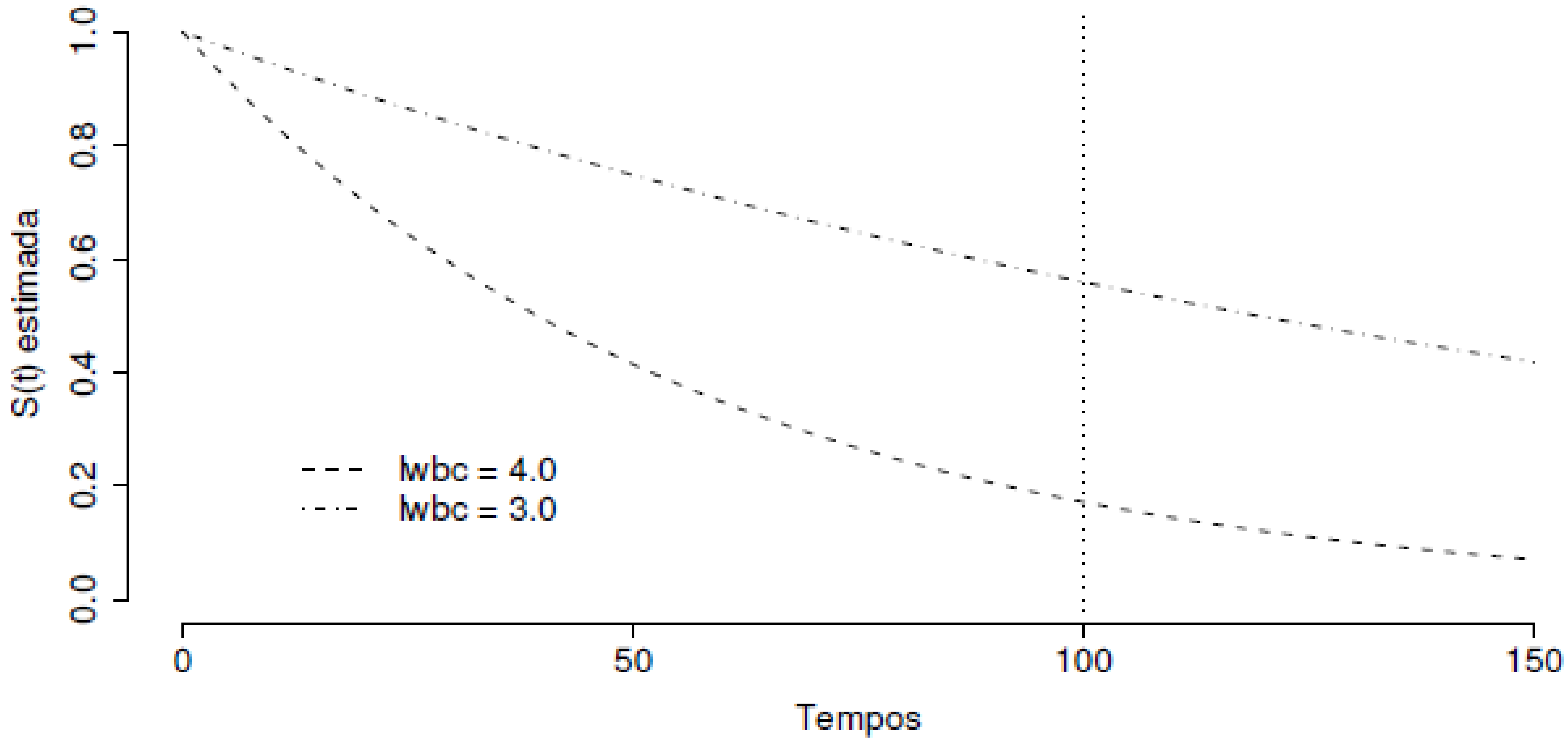
$$\hat{S}(100 \mid X) = \exp\left\{-\frac{t}{\exp(8,48 - 1,11 * 3)}\right\} = 0.559.$$

e

$$\hat{S}(100 \mid X) = \exp\left\{-\frac{t}{\exp(8,48 - 1,11 * 4)}\right\} = 0,172.$$

# Modelo de tempos de vida acelerados (MTVA)

## Curvas de sobrevivência




Modelo Exponencial