CE077 - Análise de Sobrevivência

Modelos Probabilísticos e Modelos de Regressão Paramétricos

Silva, J.L.P.

Março, 2024

Objetivos do Módulo

Objetivos do Módulo

- Apresentar os principais modelos de regressão paramétricos usados em Análise de Sobrevivência.

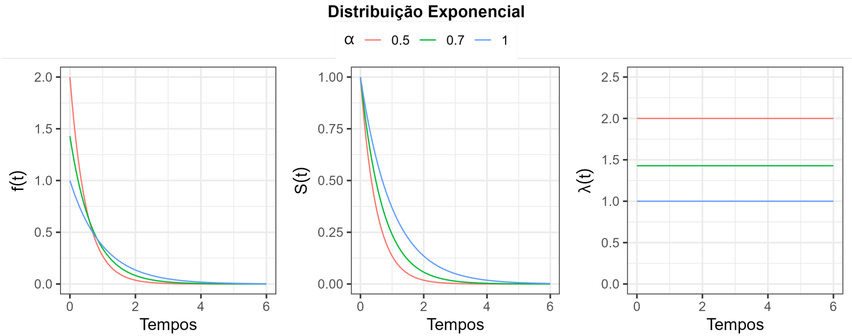
- Especificamente, discutir as distribuições exponencial, Weibull, log-normal, log-logística, gama e gama generalizada.

- Discutir como covariáveis são incorporadas nos modelos, sua estimação e inferência via máxima verossimilhança sob diferentes mecanismos de censura.

- Discutir a escolha/adequação do modelo probabilístico, interpretação dos coeficientes estimados, e exemplificar seu uso através de alguns exemplos de aplicação.

Modelos em Análise de Sobrevivência

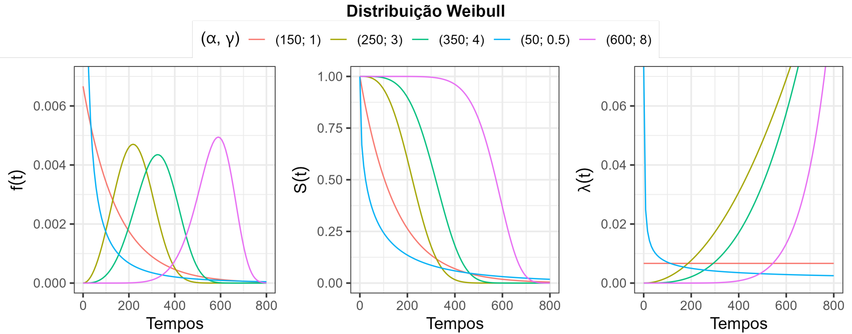
Modelo Exponencial



$$f(t) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\};\ S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)\right\};\ \lambda(t) = \frac{1}{\alpha},\ \alpha > 0$$

$$\text{Média} = \alpha; \quad \text{Variância} = \alpha^2; \quad \text{Percentil } 100p\% = t_p = -\alpha \log(1-p)$$

Modelo Weibull



$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}; \; S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}; \; \lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma - 1},$$

$$\alpha, \gamma > 0.$$

Modelo Weibull

- Quando $\gamma = 1$, tem-se a distribuição exponencial.

- A função $\lambda(t)$ é monótona: estritamente crescente para $\gamma > 1$, estritamente decrescente para $\gamma < 1$ e constante para $\gamma = 1$.

- $E[T] = \alpha\Gamma[1 + (1/\gamma)]$, sendo $\Gamma(k)$ a função gama, definida por $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx$.

- $Var[T] = \alpha^2 \left[\Gamma[1 + (2/\gamma)] - \Gamma[1 + (1/\gamma)]^2\right]$

- Percentis: $t_p = \alpha[-\log(1 - p)]^{1/\gamma}$

Modelo Weibull

É conveniente trabalhar com o logaritmo dos tempos observados, o que leva à distribuição do valor extremo ou de Gumbel.

Assim, se T seja uma distribuição Weibull com parâmetros α e γ, então $Y = \log(T)$ segue uma distribuição do valor extremo com densidade dada por

$$f(y) = \frac{1}{\sigma} \exp\left\{ \left(\frac{y - \mu}{\sigma} \right) - \exp\left\{ \frac{y - \mu}{\sigma} \right\} \right\},$$

com $-\infty < \mu < \infty$ e $\sigma > 0$, os parâmetros de locação e escala, respectivamente. Se $\mu = 0$ e $\sigma = 1$ tem-se a distribuição do valor extremo padrão.

A relação entre os parâmetros é dada por $\gamma = 1/\sigma$ e $\alpha = \exp\{\mu\}$.

Modelo Weibull

As funções de sobrevivência e de taxa de falha são dadas, respectivamente, por:

$$S(y) = \exp\left\{-\exp\left\{\frac{y-\mu}{\sigma}\right\}\right\},$$

e

$$\lambda(y) = \frac{1}{\sigma}\exp\left\{\frac{y-\mu}{\sigma}\right\}.$$

Modelo Weibull

A média e variância são, respectivamente, $E(Y) = \mu - \nu\sigma$ e $Var(Y) = (\pi^2/6)\sigma^2$, em que ν é a constante de Euler-Mascheroni:

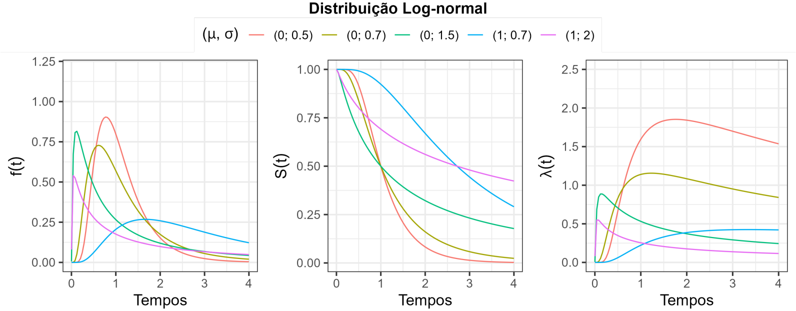
```
- digamma(1)# negativo da derivada da função gama em x=1
```

```
[1] 0.5772157
```

O percentil $100p\%$ é dado por

$$t_p = \mu + \sigma \log\left[-\log(1-p)\right].$$

Modelo Log-normal



$$f(t) = \frac{1}{\sqrt{2\pi} t\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\}; \; S(t) = \Phi\left(\frac{-log(t) + \mu}{\sigma} \right),$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão.

$$\lambda(t) = \frac{f(t)}{S(t)}; \quad -\infty < \mu < \infty, \; \sigma > 0.$$

Modelo Log-normal

- $\lambda(t)$ não é monótona.

- Percentis: $t_p = \exp\{z_p\sigma + \mu\}$, com z_p o $100p\%$ percentil da distribuição normal padrão.

- $E[T] = \exp\left\{\mu + \sigma^2/2\right\}$.

- $Var[T] = \exp\left\{\mu + \sigma^2/2\right\}(\exp(\sigma^2) - 1)$.

- Se T tem distribuição log-normal, então $Y = \log(T)$ tem distribuição normal ou Gaussiana.

Modelo Log-logístico

O modelo log-logístico é uma alternativa aos modelos Weibull e log-normal.

$$f(t) = \frac{\gamma}{\alpha^{\gamma}} t^{\gamma-1} \left(1 + (t/\alpha)^{\gamma}\right)^{-2}, \ t > 0,$$

em que $\alpha > 0$ é o parâmetro de escala e $\gamma > 0$ é o parâmetro de forma.

$$S(t) = \frac{1}{1 + (t/\alpha)^{\gamma}},$$

e

$$\lambda(t) = \frac{\gamma (t/\alpha)^{\gamma-1}}{\alpha \left[1 + (t/\alpha)^{\gamma}\right]}.$$

Modelo Log-logístico

As expressões para a esperança e variância são as seguintes.

$$E(T) = \frac{\pi \alpha Csc(\pi/\gamma)}{\gamma}, \ \gamma > 1$$

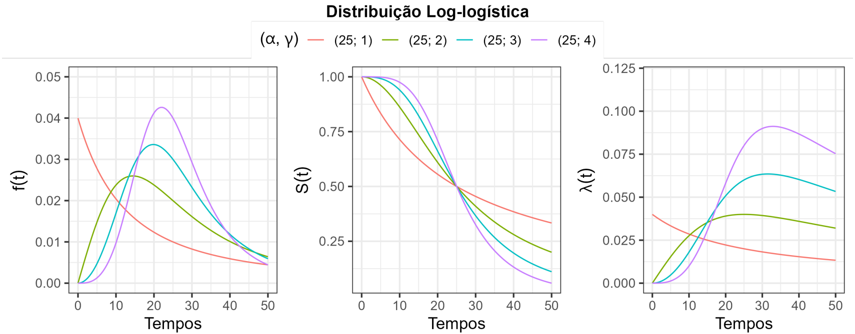
e

$$Var(T) = \frac{2\pi\alpha^2 Csc(2\pi/\gamma)}{\gamma} - E(T)^2,$$

em que $Csc(x) = 1/seno(x)$ é a função cossecante.

$$t_p = \alpha \left[\frac{p}{(1-p)} \right]^{1/\gamma}.$$

Modelo Log-logístico



Modelo Log-logístico

Assim como comentado para a distribuição Weibull, é conveniente trabalhar com o logaritmo dos tempos observados.

Se T seja uma distribuição log-logística com parâmetros α e $\gamma > 0$, então $Y = \log(T)$ segue uma distribuição logística com densidade dada por

$$f(y) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \left(1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right)^{-2},$$

com $-\infty < \mu < \infty$ e $\sigma > 0$, os parâmetros de locação e escala, respectivamente.

Modelo Log-logístico

Para este modelo, temos

$$S(y) = \frac{1}{1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}},$$

e

$$\lambda(y) = \frac{1}{\sigma}\exp\left\{\frac{y-\mu}{\sigma}\right\}\left(1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right)^{-1}.$$

A relação entre os parâmetros é dada por $\gamma = 1/\sigma$ e $\alpha = \exp\{\mu\}$.

Modelo Gama

O modelo gama é bastante utilizado em análise de sobrevivência.

A função de densidade, caracterizada pelos parâmetros de forma k ($k > 0$) e escala α ($\alpha > 0$), é dada por

$$f(t) = \frac{1}{\Gamma(k)\alpha^k} t^{k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, \ t > 0.$$

A função de sobrevivência é dada por

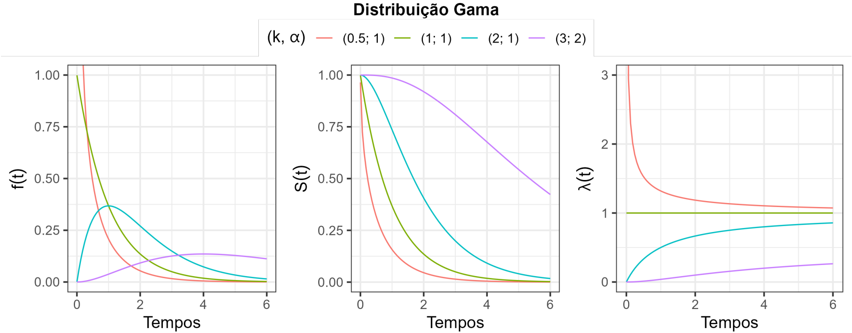
$$S(t) = \int_t^{\infty} \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp\left\{-\left(\frac{u}{\alpha}\right)\right\} du.$$

Para este modelo temos que $E(T) = k\alpha$ e $Var(T) = k\alpha^2$.

Modelo Gama

A função taxa de falha, $\lambda(t) = f(t)/S(t)$, é crescente ou decrescente mas convergindo para um valor constante quanto t vai de 0 a infinito.

Mostramos a seguir as formas de $f(t)$, $S(t)$ e $\lambda(t)$ para o modelo gama.



Modelo Gama Generalizado

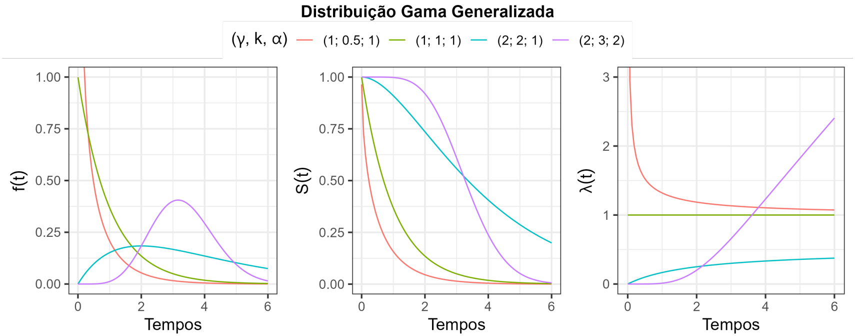
O modelo gama generalizado tem um parâmetro de escala, α, e dois de forma, γ e k, todos positivos:

$$f(t) = \frac{\gamma}{\Gamma(k)\alpha^{\gamma k}} t^{\gamma k - 1} \exp\left\{ - \left(\frac{t}{\alpha} \right)^{\gamma} \right\}.$$

Os principais modelos em análise de sobrevivência são casos particulares da gama generalizada:

- para $\gamma = k = 1$ tem-se $T \sim Exp(\alpha)$.
- para $k = 1$ tem-se $T \sim Weibull(\gamma, \alpha)$.
- para $\gamma = 1$ tem-se $T \sim Gama(k, \alpha)$.
- para $k \to \infty$ tem-se como caso limite a distribuição log-normal.

Modelo Gama Generalizado



Modelos Probabilísticos

Várias outras distribuições – como a log-gama, Rayleigh, normal inversa e Gompertz – podem ser apropriadas para modelar o tempo de falha de produtos, materiais e situações clínicas.

Se corretamente especificados, os modelos paramétricos são bastante eficientes.

A inferência para as quantidades desconhecidas dos modelos é baseada na função de verossimilhança e suas propriedades assintóticas.

Técnicas de adequação, via resíduos, são fundamentais para verificar a adequação dos modelos paramétricos.

Estimação de Parâmetros

Estimação de Parâmetros

Considere inicialmente observações não censuradas $t_1, t_2, \ldots, t_n$ de uma população caracterizada pela sua função de densidade $f(t)$.

A função de verossimilhança para θ é dada por

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

- A contribuição de cada observação não censurada é a sua função de densidade.
- O mesmo não ocorre com uma observação censurada. Temos apenas a informação de que o tempo de falha é maior que o tempo de censura.
- Assim, sua contribuição será a sua função de sobrevivência $S(t)$.

Censura do tipo I

Considere r observações não censuradas $(1, 2, \ldots, r)$ e $(n-r)$ observações censuradas $(r+1, r+2, \ldots, n)$.

Assim, para as r falhas e $n-r$ censuras observadas no término do experimento, temos

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

em que o segundo termo tem a forma $\prod_{i=r+1}^{n} S(t_i; \theta) = [S(c; \theta)]^{n-r}$, dado que as censuras ocorrem em $T = C$.

Censura do tipo II

Temos que r é fixo e somente os r menores tempos são observados. Com base em estatísticas de ordem, segue que

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

em que o terceiro termo tem a forma $\prod_{i=r+1}^{n} S(t_i;\theta) = [S(t_r;\theta)]^{n-r}$, com t_r o maior tempo observado.

Como o primeiro termo não envolve parâmetros de interesse, temos

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

Censura aleatória

Considere os tempos de falha e de censura independentes e seja $g(c)$ e $G(c)$ as funções de densidade e de sobrevivência de C, respectivamente. Para o $i-$ésimo indivíduo, temos:

- Se for observada uma censura, segue que

$$\begin{aligned} P(t_i = t, \delta_i = 0) =& P(C_i = t, T_i > C_i) = P(C_i = t, T_i > t) \\ =& g(t)S(t;\theta). \end{aligned}$$

- Se for observada uma falha

$$\begin{aligned} P(t_i = t, \delta_i = 1) =& P(T_i = t, T_i \leq C_i) = P(T_i = t, C_i \geq t) \\ =& f(t;\theta)G(t). \end{aligned}$$

Censura aleatória

Assim,

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

Sob a suposição de que a censura é não informativa, podemos desprezar os termos $G(t)$ e $g(t)$ por não envolverem θ.

Logo

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

Estimação de Parâmetros

Assim, a função de verossimilhança é a mesma para todos os mecanismos, a menos de constantes, e é dada por

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

ou, equivalentemente,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n} \left[f(t_i;\theta)\right]^{\delta_i} \left[S(t_i;\theta)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\lambda(t_i;\theta)\right]^{\delta_i} S(t_i;\theta). \end{aligned}$$

Estimação de Parâmetros

Os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$ ou equivalentemente o logaritmo de $L(\theta)$.

Eles são encontrados resolvendo-se o sistema de equações:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0.$$

Geralmente são necessários métodos numéricos para solução do sistema de equações.

Na sequência são ilustrados os cálculos para as distribuição exponencial e Weibull.

Exemplo: Distribuição Exponencial

A função de verossimilhança é dada por

$$
\begin{aligned}
L(\alpha) &= \prod_{i=1}^{n} \left[\frac{1}{\alpha} \exp\left\{ - \left(\frac{t_i}{\alpha} \right) \right\} \right]^{\delta_i} \left[\exp\left\{ - \left(\frac{t_i}{\alpha} \right) \right\} \right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{1}{\alpha} \right]^{\delta_i} \exp\left\{ - \left(\frac{t_i}{\alpha} \right) \right\}
\end{aligned}
$$

Tomando logaritmos, segue que

$$
l(\alpha) = - \sum_{i=1}^{n} \delta_i \log(\alpha) - \frac{1}{\alpha} \sum_{i=1}^{n} t_i .
$$

Exemplo: Distribuição Exponencial

Logo,

$$\frac{\partial l(\alpha)}{\partial \alpha} = -\frac{1}{\alpha}\sum_{i=1}^{n} \delta_i + \frac{1}{\alpha^2}\sum_{i=1}^{n} t_i.$$

do qual, igualando a zero e valiando $\alpha = \hat{\alpha}$, obtemos o EMV:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \delta_i} = \frac{\sum_{i=1}^{n} t_i}{r}.$$

Observações:

- O termo $\sum_{i=1}^{n} t_i$ é denominado *tempo total sob teste*.
- Se não houver censuras, $\hat{\alpha}$ será a média amostra, ou seja, $\hat{\alpha} = \bar{t}$.

Distribuição Weibull

A função de verossimilhança é dada por

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

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

Logo,

$$l(\gamma,\alpha) = \sum_{i=1}^{n}\delta_i\log(\gamma) - \sum_{i=1}^{n}\delta_i\gamma\log(\alpha) + (\gamma-1)\sum_{i=1}^{n}\delta_i\log(t_i) - \alpha^{-\gamma}\sum_{i=1}^{n}t_i^\gamma$$

$$= r\log(\gamma) - r\gamma\log(\alpha) + (\gamma-1)\sum_{i=1}^{n}\delta_i\log(t_i) - \alpha^{-\gamma}\sum_{i=1}^{n}t_i^\gamma.$$

Distribuição Weibull

Alternativamente, tomando-se $y_i = \log(t_i)$ e utilizando-se a distribuição do valor do valor extremo, obtemos

$$l(\gamma, \alpha) = -r\log(\sigma) + \sum_{i=1}^{n} \delta_i \left(\frac{y_i}{\sigma}\right) - \frac{r\mu}{\sigma} - \sum_{i=1}^{n} \exp\left\{\frac{(y_i - \mu)}{\sigma}\right\},$$

expressão que é mais simples que aquela obtida para a distribuição Weibull.

Derivando a expressão acima e igualando a zero, obtemos o seguinte sistema de equações:

Distribuição Weibull

$$
\begin{aligned}
\frac{\partial l(\gamma,\alpha)}{\partial \mu} =& \frac{1}{\hat{\sigma}}\left[-r+\sum_{i=1}^{n}\exp\left\{\frac{(y_i-\hat{\mu})}{\hat{\sigma}}\right\}\right]=0\\
\frac{\partial l(\gamma,\alpha)}{\partial \sigma} =& \frac{1}{\hat{\sigma}^2}\left[-r\hat{\sigma}+\sum_{i=1}^{n}\delta_i y_i+r\hat{\mu}+\sum_{i=1}^{n}\exp\left\{\frac{(y_i-\hat{\mu})}{\hat{\sigma}}\right\}(y_i-\hat{\mu})\right]=0
\end{aligned}
$$

Para obtenção dos estimadores de máxima verossimilhança precisamos utilizar um método numérico, como o algoritmo de Newton-Raphson.

Algoritmo de Newton Raphson

Este método utiliza a matriz de derivadas segundas $(\mathcal{F})$ do logaritmo da função de verossimilhança e consiste em iterar a seguinte expressão

$$\hat{\theta}_{(k+1)} = \hat{\theta}_{(k)} - \left[\mathcal{F}(\hat{\theta}_{(k)})\right]^{-1} U(\hat{\theta}_{(k)}),$$

que é baseada em uma expansão de $U(\hat{\theta}_{(k)})$ em série de Taylor em torno de $\hat{\theta}_{(k)}$.

- Partindo de um valor inicial $\hat{\theta}_{(0)}$, geralmente tomado como $\hat{\theta}_{(0)} = 0$, o valor de $\hat{\theta}_{(k+1)}$ é atualizado.
- Geralmente obtém-se convergência em poucos passos, com erro relativo menor do que um valor fixado (e.g. 0,001) entre dois passo consecutivos.

Algoritmo de Newton Raphson

Para o modelo exponencial $\mathcal{F}$ é um escalar igual a

$$\begin{aligned}\mathcal{F}(\alpha) =& \frac{\partial^2 l(\alpha)}{\partial \alpha^2} \\ =& \frac{r}{\alpha^2} - \frac{2\sum_{i=1}^{n} t_i}{\alpha^3}.\end{aligned}$$

Para o modelo Weibull, $\mathcal{F}(\gamma, \alpha)$ é uma matriz simétrica 2×2 com elementos

$$\begin{aligned}\mathcal{F}_{11}(\gamma, \alpha) =& \frac{\partial^2 l(\gamma, \alpha)}{\partial \gamma^2}, \\ \mathcal{F}_{22}(\gamma, \alpha) =& \frac{\partial^2 l(\gamma, \alpha)}{\partial \alpha^2}, \\ \mathcal{F}_{12}(\gamma, \alpha) =& \mathcal{F}_{21}(\gamma, \alpha) = \frac{\partial^2 l(\gamma, \alpha)}{\partial \gamma \partial \alpha}.\end{aligned}$$

Propriedades Importantes

- Assintoticamente, sob certas condições de regularidade, temos $\hat{\theta} \sim N_k\left(\theta, Var(\hat{\theta})\right)$, sendo k a dimensão de $\hat{\theta}$.
- $Var(\hat{\theta}) \approx -\left[E(\mathcal{F}(\theta))\right]^{-1}$, ou seja, a matriz de variância-covariância é aproximadamente o negativo da inversa da esperança da matriz de derivadas segundas de $L(\theta)$.
- Em situações em que a esperança é impossível ou difícil de ser calculada, usa-se simplesmente $[-\mathcal{F}(\theta)]^{-1}$.
- Geralmente, $Var(\hat{\theta})$ depende de θ. Assim, uma estimativa para $Var(\hat{\theta})$ é obtida substituindo-se θ por $\hat{\theta}$.

Intervalos de Confiança e Testes de Hipóteses

Intervalos de Confiança

No caso em que θ é um escalar, um intervalo aproximado de $(1-\alpha)100\%$ é dado por

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta})}.$$

Por exemplo, um intervalo de confiança de 95% para α no modelo exponencial é dado por

$$\hat{\alpha} \pm 1,96\sqrt{\frac{\hat{\alpha}^2}{r}},$$

pois $-\frac{\partial^2 l(\alpha)}{\partial \alpha^2}$ avaliado em $\hat{\alpha}$ é igual a $r/\hat{\alpha}^2$.

Para θ vetor pode-se construir um intervalo de confiança para cada parâmetro separadamente.

Método Delta

Considere agora o caso em que temos interesse em estimar uma função ϕ dos parâmetros. Por exemplo, a mediana da distribuição Weibull é $\phi = g(\gamma, \alpha) = t_{0,5} = \alpha[-\log(1 - 0,5)]^{1/\gamma}$.

O estimador de máxima verossimilhança para ϕ é $\hat{\phi} = g(\hat{\gamma}, \hat{\alpha})$. Assim, basta substituir γ e α por seus respectivos estimadores de máxima verossimilhança.

Para estimar o erro padrão de ϕ usamos o *método delta*, que se baseia em expansão de $g(\hat{\theta})$ em torno de $E(\hat{\theta}) = \theta$ ignorando-se os termos superiores ao de primeira ordem.

Método Delta

Se θ é um escalar e há interesse em $Var[g(\hat{\theta})]$, tem-se

$$g(\hat{\theta}) \approx g(\theta) + (\hat{\theta} - \theta)\left(\frac{dg(\theta)}{d\theta}\right)$$

e, portanto,

$$Var[g(\hat{\theta})] \approx Var(\hat{\theta})\left(\frac{dg(\theta)}{d\theta}\right)^2.$$

Se θ é um vetor, por exemplo $\theta = (\gamma, \alpha)$ e há interesse em $\phi = g(\gamma, \alpha)$, tem-se

$$Var(\hat{\phi}) \approx Var(\hat{\alpha})\left(\frac{\partial\phi}{\partial\alpha}\right)^2 + 2Cov(\hat{\alpha}, \hat{\gamma})\left(\frac{\partial\phi}{\partial\alpha}\right)\left(\frac{\partial\phi}{\partial\gamma}\right) + Var(\hat{\gamma})\left(\frac{\partial\phi}{\partial\gamma}\right)^2.$$

Método Delta - Exemplo Log-normal

Considere, por exemplo, que desejamos construir intervalos de confiança para $E[T]$ para o modelo log-normal, para o qual $E[T] = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}$.

Para isso, obtemos

$$\begin{aligned}\widehat{Var}(\hat{E}[T]) \approx & \widehat{Var}(\hat{\mu})\left[\exp\left\{\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right\}\right]^2 + \widehat{Var}(\hat{\sigma})\left[\hat{\sigma}\exp\left\{\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right\}\right]^2 \\ & + 2\widehat{Cov}(\hat{\mu}, \hat{\sigma})\left[\exp\left\{\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right\}\right]\left[\hat{\sigma}\exp\left\{\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right\}\right]\end{aligned}$$

e procedemos com o cálculo usual utilizando os resultados assintóticos relatados anteriormente.

Testes de Hipóteses

Dado um modelo com um vetor $\theta = (\theta_1, \ldots, \theta_p)^T$, podemos estar interessados em testar hipóteses relacionadas a este vetor ou a um subconjunto deles.

Três testes são em geral utilizados para esta finalidade:

- Teste de Wald: baseado na distribuição assintótica de $\hat{\theta}$, é uma generalização do teste t de Student.
- Teste da Razão da Verossimilhança: baseado na função de verossimilhança, envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição e sob H_0.
- Teste Escore: obtido através da função escore.

Teste de Wald

Baseado na distribuição assintótica de $\hat{\theta}$, é uma generalização do teste t de Student. Dada a hipótese nula

$$H_0 : \theta = \theta_0,$$

a estatística de teste é dada por

$$W = (\hat{\theta} - \theta_0)^T [-\mathcal{F}(\theta_0)](\hat{\theta} - \theta_0),$$

que, sob H_0, segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade.

No caso em que θ é escalar, temos

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\widehat{Var}(\hat{\theta})}.$$

Teste da Razão da Verossimilhança

Baseado na função de verossimilhança, envolve a comparação da verossimilhança maximizada sem restrição e sob a hipótese nula.

A estatística de teste é dada por

$$TRV = -2\log\left[\frac{L(\theta_0)}{L(\hat{\theta})}\right] = 2[\log L(\hat{\theta}) - L(\theta_0)],$$

que sob, H_0, segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade.

Teste Escore

Baseado na função escore, a estatística de teste é dada por

$$S = U^T(\theta_0)[\mathcal{F}(\theta_0)]^{-1}U(\theta_0),$$

em que $U(\theta_0)$ é a função escore $U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta}$ avaliada em θ_0 e $F(\theta_0)$ é a matriz de variância-covariância de $U(\theta)$ também avaliada em θ_0.

Para amostras grandes, H_0 é rejeitada a um nível $100\alpha\%$ de significância, se $S > \chi^2_{p,1-\alpha}$.

Testes de Hipóteses

As três estatísticas de testes podem ser adaptadas para o caso em que já interesse em testar somente um subconjunto de θ.

Pode-se construir intervalos de confiança com base nas estatísticas da razão da verossimilhança e escore.

Por exemplo, a partir da razão de verosimilhança, tem-se que $\left\{\theta | TRV(\theta) < \chi^2_{p,1-\alpha}\right\}$ é um intervalo de confiança de nível $(1-\alpha)100\%$ para θ.

Escolha do Modelo Probabilístico

Seleção/Adequação de Modelos

A escolha do modelo é parte extremamente importante no processo de análise paramétrica de dados de tempos de vida.

Discutiremos as seguintes abordagens:

- Método gráfico:
  - Modelo candidato *versus* Kaplan-Meier.
  - Linearização do modelo.
- Teste de hipóteses: utilizar o TRV para comparar o modelo proposto com a gama generalizada.

Método Gráfico

Comparação direta da sobrevivência estimada do modelo proposto (MP) com o Kaplan-Meier (KM)

Comparamos num mesmo gráfico as funções $\hat{S}_{KM}(t)$ com $\hat{S}_{MP}(t)$. O modelo mais adequado é aquele cuja curva de sobrevivência estimada mais se aproxima da estimativa de Kaplan-Meier.

Comparação da função de sobrevivência linearizada com o tempo

O modelo proposto é adequado se for observada uma relação linear entre a função transformada e o tempo.

A seguir, são apresentados alguns exemplos de linearização da função de sobrevivência.

Linearização no modelo exponencial

Para o modelo exponencial, temos

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

Assim,

$$-\log[S(t)] = \left(\frac{1}{\alpha}\right)t,$$

o que indica que o gráfico $-\log[\hat{S}(t)]$ versus t deve ser aproximadamente linear, passando pela origem, se este modelo for apropriado.

$\hat{S}(t)$ é o estimador de Kaplan-Meier.

Linearização no modelo Weibull

Para o modelo Weibull, temos

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

Deste modelo, obtemos a linearização

$$\log\left[-\log[S(t)]\right] = -\gamma\log(\alpha) + \gamma\log(t).$$

O gráfico de $\log\left[-\log[\hat{S}(t)]\right]$ versus $\log(t)$ deve ser aproximadamente linear, passando pela origem, se este modelo for apropriado.

Se a inclinação foi igual a 1 temos indicação de que o modelo exponencial é adequado.

Linearização no modelo Log-normal

Para o modelo log-normal, a função de sobrevivência

$$S(t) = \Phi\left(\frac{-log(t) + \mu}{\sigma}\right)$$

pode ser linearizada tomando

$$\Phi^{-1}\left(S(t)\right) = \frac{-log(t) + \mu}{\sigma},$$

em que Φ^{-1} denota os percentis da distribuição normal padrão.

O gráfico de $\Phi^{-1}(\hat{S}(t))$ versus $\log(t)$ deve ser aproximadamente linear, com intercepto μ/σ e inclinação $-1/\sigma$, se este modelo for adequado.

Teste de Hipóteses

As hipóteses a serem testadas são:

$$H_0: \text{ O modelo de interesse é adequado}$$

versus uma hipótese alternativa vaga, de que o modelo não é adequado.

Usamos a estatística da razão de verossimilhanças em modelos encaixados:

$$TRV = -2\log\left[\frac{L(\hat{\theta}_M)}{L(\hat{\theta}_G)}\right] = 2\left[\log L(\hat{\theta}_G) - \log L(\hat{\theta}_M)\right],$$

em que $L(\hat{\theta}_G)$ e $L(\hat{\theta}_M)$ correspondem aos valores da verossimilhança sob o ajuste do modelo generalizado e o modelo candidato, respectivamente.

Teste de Hipóteses

Sob H_0, a estatística de teste segue aproximadamente uma distribuição χ^2 com número de graus de liberdade igual a diferença do número de parâmetros ($\hat{\theta}_G$ e $\hat{\theta}_M$) dos modelos sendo comparados.

Por exemplo, podemos testar o ajuste da distribuição gama generalizada que apresenta os modelos exponencial, Weibull, log-normal e gama como casos particulares.

Exemplo de Aplicação

Exemplo de Aplicação

Aplicação 1: pacientes com câncer de bexiga.

Introdução aos Modelos de Regressão Paramétricos

Modelos de Regressão Paramétricos

Como sabemos, estudos em geral envolvem covariáveis que podem estar relacionadas com o tempo de sobrevivência, que devem ser incluídas na análise.

A forma mais eficiente de acomodar covariáveis é através de modelos de regressão.

Devido à sua flexibilidade, o modelo de regressão semiparamétrico de Cox (Cox, 1972) é utilizado com frequência na área médica.

Discutiremos agora uma classe alternativa de modelos, os modelos paramétricos.

Modelos de Regressão Paramétricos

Os modelos paramétricos, também denominados *modelos de tempo de vida acelerados*, são mais eficientes, porém menos flexíveis que os modelos semiparamétricos.

Os modelos paramétricos são utilizados com mais frequência na área industrial do que na médica, pois os estudos industriais podem ser planejados e as fontes de perturbação controladas.

Na sequência, discutiremos a formulação do modelo, assumindo uma distribuição como a exponencial ou Weibull, sua estimação por máxima verossimilhança, técnicas de adequação do modelo, e interpretação das quantidades estimadas através de aplicações reais.

Modelos Paramétricos

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

Covariáveis são incluídas por meio da estrutura de regressão. O modelo paramétrico especifica um efeito multiplicativo das covariáveis em T, sendo o papel de X o de acelerar ou desacelerar o tempo até a falha.

Os parâmetros são estimados via máxima verossimilhança e a adequação do modelo ajustado é feita utilizando os resíduos.

Modelos Lineares para Dados de Sobrevivência

Modelo de Regressão Exponencial

O modelo mais simples é o modelo exponencial. Considere o modelo

$$T = \exp\{\beta_0 + \beta_1 x\}\,\varepsilon, \tag{1}$$

em que ε segue uma distribuição exponencial com média unitária, ou seja, $(f(\varepsilon) = \exp\{-\varepsilon\})$.

O modelo é linearizável fazendo

$$Y = \log(T) = \beta_0 + \beta_1 x + \nu,\ \text{com } \nu = \log(\varepsilon). \tag{2}$$

O modelo é semelhante ao modelo normal, com a exceção da distribuição dos erros que não é normal, mas segue uma distribuição do valor extremo padrão $(f(\nu) = \exp\{\nu - \exp\{\nu\}\})$.

Note que x atua linearmente em Y e multiplicativamente em T.

Modelo de Regressão Exponencial

A função de sobrevivência para Y condicional a x é dada por

$$S(y|x) = \exp\left\{-\exp\left\{y - (\beta_0 + \beta_1 x)\right\}\right\}.$$

A função de sobrevivência para T, condicional em x, é dada por

$$S(t|x) = \exp\left\{-\left(\frac{t}{\exp\left\{\beta_0 + \beta_1 x\right\}}\right)\right\}.$$

Para estimação e inferência sobre $\theta = (\beta_0, \beta_1)$, o vetor de parâmetros do modelo, o método de máxima verossimilhança é utilizado.

Modelo de Regressão Exponencial

Dada a independência entre as observações, a função de verossimilhança para o modelo linear (2) é dada por

$$L(\theta) = \prod_{i=1}^{n} \left[f(y_i|x_i)\right]^{\delta_i} \left[S(y_i|x_i)\right]^{(1-\delta_i)}, \tag{3}$$

ou, para modelos da forma (1), por

$$L(\theta) = \prod_{i=1}^{n} \left[f(t_i|x_i)\right]^{\delta_i} \left[S(t_i|x_i)\right]^{(1-\delta_i)}. \tag{4}$$

Modelo de Regressão Exponencial

É necessário substituir as funções de densidade e sobrevivência por aquelas da distribuição do valor extremo em (3) ou da exponencial em (4).

Fazendo-se isto em (3) e tomando o logaritmo, obtemos

$$\log L(\theta) = \sum_{i=1}^{n} \left[\delta_i(y_i - \beta_0 - \beta_1 x_i) - \exp\left\{y_i - \beta_0 - \beta_1 x_i\right\}\right]. \qquad (5)$$

Os estimadores de máxima verossimilhança são os valores de $\theta = (\beta_0, \beta_1)$ que maximizam a função $\log L(\theta)$ em (5), e são obtidos por métodos numéricos.

Modelo de Regressão Weibull

Uma forma de generalizar o modelo (2) é incluir um parâmetro extra de escala.

O modelo linear (2), na presença de p covariáveis, passa a ser

$$Y = \log(T) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \sigma\nu = x'\beta + \sigma\nu, \qquad (6)$$

em que $x' = (1, x_1, \ldots, x_p)$ e $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$.

No modelo Weibull, $\log(T)$ tem uma distribuição do valor extremo com parâmetro de escala σ.

Modelo de Regressão Weibull

A função de sobrevivência para Y condicional em x é dada por

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

enquanto para T condicional em x, temos

$$S(t|x) = \exp\left\{-\left(\frac{t}{\exp\left\{x'\beta\right\}}\right)^{1/\sigma}\right\}.$$

Modelo de Tempo de Vida Acelerado

Extensões do modelo (6) consideram outras distribuições para ν ou T:

- Distribuições adequadas para T incluem a log-normal, a gama e a log-logística.
- De forma correspondente, a distribuição para ν é normal, log-gama e logística.

O modelo é conhecido como *modelo de tempo de vida acelerado* e o papel das covariáveis é acelerar ou desacelerar o tempo de vida. Na escala original temos:

$$T = \exp\left\{x'\beta\right\}\exp\left\{\sigma\nu\right\}.$$

Uma generalização pode ser obtida acrescentando mais um parâmetro de forma. A gama generalizada é um exemplo.

Adequação do Modelo Ajustado

Adequação do Modelo Ajustado

Diversos resíduos têm sido propostos na literatura para avaliar o ajuste dos modelos de regressão paramétricos em análise de sobrevivência:

- *Resíduos de Cox-Snell* (1968) e *resíduos padronizados*: úteis para examinar o ajuste global do modelo;
- *Resíduos martingal*: úteis para determinar a forma funcional (linear, quadrática etc.) de uma covariável, em geral contínua, sendo incluída no modelo de regressão;
- *Resíduos deviance*: auxiliam a examinar a acurácia do modelo para cada indivíduo sob estudo.

Resíduos de Cox-Snell

São dados por

$$\hat{e}_i = \hat{\Lambda}(t_i|x), \tag{7}$$

em que $\hat{\Lambda}(\cdot)$ é a função taxa de falha acumulada para o modelo ajustado.

Por exemplo, para os modelos exponencial, Weibull e log-normal, temos

$$\begin{aligned}
\text{Exponencial: } \hat{e}_i =& t_i \exp\left\{-x_i'\hat{\beta}\right\}, \\
\text{Weibull: } \hat{e}_i =& \left[t_i \exp\left\{-x_i'\hat{\beta}\right\}\right]^{\hat{\gamma}}, \\
\text{Log-normal: } \hat{e}_i =& -\log\left[1 - \Phi\left(\frac{\log(t_i) - x_i'\hat{\beta}}{\hat{\sigma}}\right)\right].
\end{aligned}$$

Os resíduos vêm de uma população homogênea e devem seguir uma distribuição exponencial padrão se o modelo for adequado.

Resíduos de Cox-Snell

O gráfico de $\hat{e}_i$ versus $\hat{\Lambda}(\hat{e}_i)$ deve ser aproximadamente uma reta com inclinação 1, quando o modelo exponencial for adequado, pois $\hat{\Lambda}(\hat{e}_i) = -\log(\hat{S}(\hat{e}_i))$, em que $\hat{S}(\hat{e}_i)$ é a função de sobrevivência de $\hat{e}_i$ obtida por Kaplan-Meier.

Klein e Moeschberger (2003) observam que tais resíduos devem ser usados com cuidado, pois a distribuição exponencial dos mesmos mantém-se somente quando os verdadeiros valores dos parâmetros são usados em (7).

Desvios da distribuição exponencial podem ocorrer devido à incerteza da estimação de β, sendo maior na cauda direita da distribuição e para amostras pequenas.

Resíduos Padronizados

Os resíduos padronizados são dados por

$$\hat{\nu}_i = \frac{y_i - x_i'\hat{\beta}}{\hat{\sigma}},$$

em que $y_i = \log(t_i)$.

- Observe que os resíduos de observações censuradas também são censurados.
- Podemos utilizar as técnicas apresentadas para população homogênea e comparar os resíduos do modelo proposto com o Kaplan-Meier.
- Equivalentemente, podemos avaliar em um gráfico as probabilidades de sobrevivência dos resíduos $e_i^* = \exp(\hat{\nu}_i)$ comparando-as com o Kaplan-Meier.

Resíduos Martingal

Os resíduos *martingal* são dados por

$$\hat{m}_i = \delta_i - \hat{e}_i,$$

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

Estes resíduos são vistos como uma estimativa do número de falhas em excesso observado nos dados mas não predito pelo modelo.

Eles são geralmente usados para examinar a melhor forma funcional (linear, quadrática, etc) para uma dada covariável no modelo de regressão assumido.

Resíduos Deviance

Os resíduos *deviance* são dados por

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

e buscam tornar os resíduos *martingal* mais simétricos em torno de zero, o que facilita a identificação de valores atípicos.

Se o modelo for adequado, deve-se observar um comportamento aleatório em torno de zero.

Para verificação do ajuste, os resíduos *martingal* ou *deviance* são plotados contra o tempo.

Interpretação dos Coeficientes Estimados

Interpretação dos Coeficientes

A interpretação direta dos coeficientes não é uma tarefa simples.

Uma proposta razoável é de se fazer uso da razão de tempos medianos (Hosmer e Lemeshow, 1999). Considerando uma covariável binária, a razão de tempos medianos é:

$$\frac{t_{0,5}(x=1,\hat{\beta})}{t_{0,5}(x=0,\hat{\beta})} = e^{\hat{\beta}}.$$

Exemplo $e^{\hat{\beta}} = 2$. Isto significa que o tempo mediano (estimado) de um grupo é duas vezes o do outro grupo (mantendo fixas as demais covariáveis).

Interpretação dos Coeficientes

Por exemplo, se T segue uma distribuição Weibull com parâmetros $\exp\{\beta_0 + \beta_1 x\}$ e γ, temos

$$t_{0,5}(x, \hat{\beta}) = [-log(0,5)]^{1/\hat{\gamma}} \exp\left\{\hat{\beta}_0 + \hat{\beta}_1 x\right\},$$

e, assim,

$$\frac{t_{0,5}(x=1, \hat{\beta})}{t_{0,5}(x=0, \hat{\beta})} = \frac{[-log(0,5)]^{1/\hat{\gamma}} \exp\left\{\hat{\beta}_0 + \hat{\beta}_1\right\}}{[-log(0,5)]^{1/\hat{\gamma}} \exp\left\{\hat{\beta}_0\right\}} = e^{\hat{\beta}_1}.$$

Tal resultado vale também para o modelo log-normal.

Modelo de Tempo de Vida Acelerado (AFT)

Para uma variável de tempo até evento T, um AFT tem a seguinte relação.

1. Escala $Y = \log(T)$
$$Y_i = x_i'\beta + W_i,$$
em que $W_i \sim f$ são os termos de erro que definem o modelo *log-linear*.

2. Na escala original temos:
$$\begin{aligned} T &= \exp\left\{x_i'\beta\right\} \exp\left\{W_i\right\} \\ &= \exp\{\eta_i\} T_0. \end{aligned}$$

Por exemplo, W normal leva ao modelo log-normal para T, enquanto W valor-extremo resulta em T com distribuição Weibull.

Como vimos, as covariáveis aceleraram ou desaceleram o tempo de vida.

Modelo de Tempo de Vida Acelerado (AFT)

Da relação $T = \exp\{\eta_i\} T_0$, pode-se obter que

$$S_1(t) = S_0(e^{\eta_i} t),$$

em que $S_0(t)$ é a função de sobrevivência de base e $\varphi = e^{\eta_i}$ é um *fator de aceleração*.

O princípio é que o efeito de uma covariável resulta em "esticar" ou "encolher" a curva de sobrevivência ao longo do eixo do tempo por uma constante relativa φ.

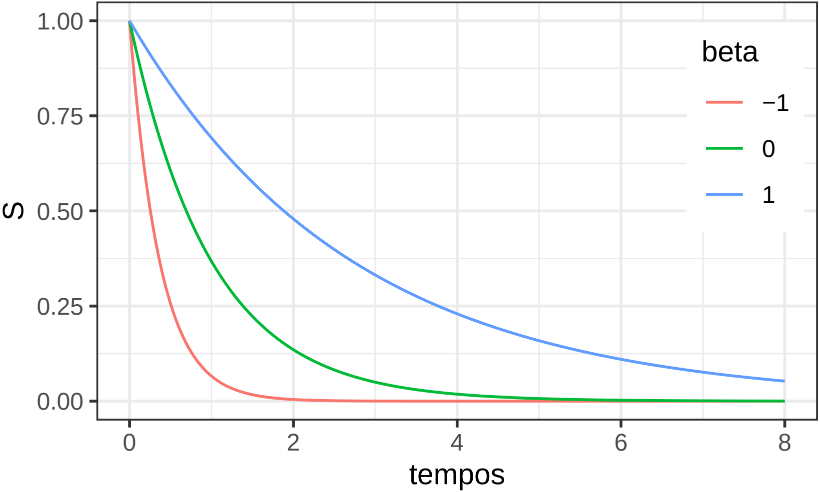
Tal fato será ilustrado para o caso de uma covariável binária x_1 com dois níveis; por exemplo, $x_1 = 0$ para placebo e $x_1 = 1$ para novo tratamento, assumindo uma distribuição Weibull com $\gamma = 1$, $\beta_0 = 0$ e $\beta \pm 1$.

Modelo de Tempo de Vida Acelerado (AFT)

```r
beta_0 <- 0
beta_1pos <- 1
beta_1neg <- -1
tempos <- seq(0, 8, 0.01)
S_0 <- exp(-tempos/exp(beta_0))
S_pos <- exp(-tempos/exp(beta_0 + beta_1pos))
S_neg <- exp(-tempos/exp(beta_0 + beta_1neg))
dat <- data.frame(tempos = rep(tempos, 3),
                  S = c(S_0, S_pos, S_neg),
                  beta = as.factor(rep(c(0, 1, -1),
                        each = length(tempos))))
```

Modelo de Tempo de Vida Acelerado (AFT)

```
p1 <- ggplot(dat, aes(x = tempos, y = S, color = beta)) + geom_line() +
  theme_bw() + theme(legend.position = c(0.9, 0.7)); p1
```



Modelo de Tempo de Vida Acelerado (AFT)

Os tempos medianos para os três grupos são

```
tm_0 <- exp(beta_0)*(-log(1-0.5))
tm_pos <- exp(beta_0 + beta_1pos)*(-log(1-0.5))
tm_neg <- exp(beta_0 + beta_1neg)*(-log(1-0.5))
c(tm_0, tm_pos, tm_neg)
```

```
[1] 0.6931472 1.8841694 0.2549946
```

A razão de tempos medianos vale exp(1), ou seja:

```
c(tm_pos/tm_0, tm_0/tm_neg, exp(1))
```

```
[1] 2.718282 2.718282 2.718282
```

Note que este resultado vale para outros percentis!

Modelo de Tempo de Vida Acelerado (AFT)

A sobrevivência mediana para cada um dos grupos é alcançada em:

- $\beta = 0$ em $t_0 = tm_0$

```
tm_0
```

```
[1] 0.6931472
```

- $\beta = 1$ em $t_{pos} = tm_0 \times \exp(1)$

```
tm_0*exp(1)
```

```
[1] 1.884169
```

- $\beta = -1$ em $t_{neg} = tm_0 \times \exp(-1)$

```
tm_0*exp(-1)
```

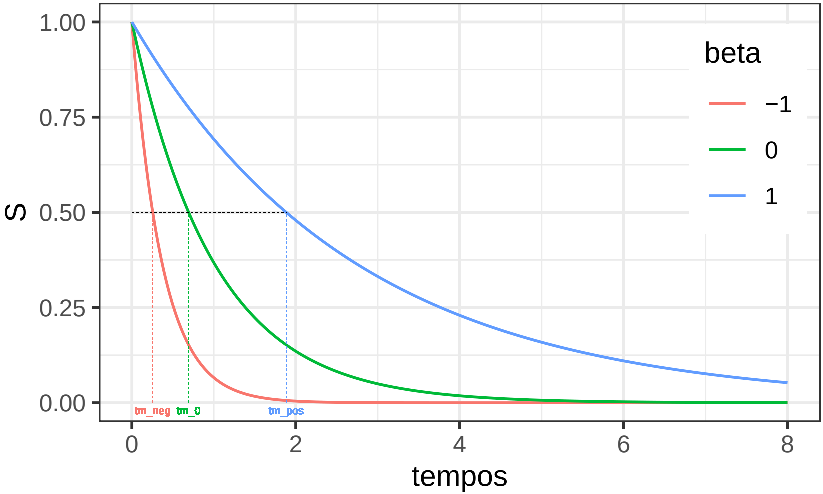
```
[1] 0.2549946
```

Modelo de Tempo de Vida Acelerado (AFT)

```r
p2 <- p1 + geom_segment(aes(x = 0, xend = tm_pos, y = 0.5, yend = 0.5),
                        color = "black", linetype = "dashed",
                        size = 0.1) +
  geom_segment(aes(x = tm_pos, xend = tm_pos, y = 0, yend = 0.5),
               color = "#619CFF", linetype = "dashed", size = 0.1) +
  geom_segment(aes(x = tm_neg, xend = tm_neg, y = 0, yend = 0.5),
               color = "#F8766D", linetype = "dashed", size = 0.1) +
  geom_segment(aes(x = tm_0, xend = tm_0, y = 0, yend = 0.5),
               color = "#00BA38", linetype = "dashed", size = 0.1) +
  geom_text(x = tm_0, y = -0.02, label = "tm_0", color = "#00BA38",
            size = 1.5) +
  geom_text(x = tm_pos, y = -0.02, label = "tm_pos", color = "#619CFF",
            size = 1.5) +
  geom_text(x = tm_neg, y = -0.02, label = "tm_neg", color = "#F8766D",
            size = 1.5)
```

Modelo de Tempo de Vida Acelerado (AFT)

```
p2
```



Exemplos de Aplicação

Exemplos de Aplicação

Aplicação 2: sobrevida de pacientes com leucemia aguda.

Aplicação 3: sobrevida de pacientes com câncer de encéfalo.

Referências

Referências

Colosimo, E. A., & Giolo, S. (2006). *Análise de Sobrevivência Aplicada*. São Paulo: Editora Edgard Blücher.

Colosimo, E. A., *Análise de Sobrevivência Aplicada*. Notas de Aula. Disponível em https://www.est.ufmg.br/~enricoc/metodos_estatisticos_analise_sobrevivencia.htm.

Lee, E. T., & Wang, S. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons.