# Análise de Sobrevivência

Camila Borelli Zeller
Lupércio França Bessegato
Departamento de Estatística/UFJF

## Roteiro Geral





## Modelos de Regressão Paramétricos

## Roteiro do Módulo

## Introdução

## Covariáveis

- Variáveis que podem estar relacionadas com o tempo de sobrevivência
  - √ Ex.: Tempo até a ocorrência de Aids em pacientes infectados pelo HIV
    - Contagem de células CD4 e CD8
    - Fatores de prognóstico importantes

Análise de Sobrevivência -- 2017 6

## Técnicas Não Paramétricas

- Não envolvem nenhuma estrutura paramétrica
- Vantagens:
  - √ Simplicidade e facilidade de aplicação
- Limitação:
  - √ Análise mais detalhada é inviável
    - Não permitem a inclusão direta de covariáveis
  - √ Importância:
    - Descrição dos dados de sobrevivência

Análise de Sobrevivência -- 2017 7

## Estratificação

- Forma simples de fazer análise mais elaborada incluindo covariáveis
  - √ Dividir dados em estratos de acordo com covariáveis
    - Usar as técnicas não paramétricas apropriadas
  - √ Vantagem:
    - Facilidade de cálculo e facilidade de entendimento
  - √ Limitação:
    - Difícil análise com muitas covariáveis
      (gera um número muito grande de estratos)

Análise de Sobrevivência -- 2017 8

## Modelos Paramétricos de Sobrevivência

## Modelos com Covariáveis

- Abordagens para modelas dados de sobrevivência com covariáveis:
  - √ Famílias paramétricas
  - √ Vida acelerada
  - √ Riscos proporcionais
  - √ Chances proporcionais

Análise de Sobrevivência -- 2017 10

## Famílias Paramétricas

- Escolher uma das distribuições paramétricas
- Parâmetros da distribuição dependendo das covariáveis
  - √ Exponencial – modelo log-linear:
    - Parâmetro α: $\log \alpha = \boldsymbol{x}'\boldsymbol{\beta}$
  - √ Weibull – modelo log-linear:
    - Parâmetro α: $\log \alpha = \boldsymbol{x}'\boldsymbol{\beta}$
    - Parâmetro γ: fixo ou $\log \gamma = \boldsymbol{x}'\boldsymbol{\beta}^*$

Análise de Sobrevivência -- 2017 11

- Comentários:
  - √ Se houver k grupos, cada grupo teria sua própria distribuição em uma família
  - √ Abordagem trabalhável, mas que não é parcimoniosa
    - Não leva a interpretações fáceis

Análise de Sobrevivência -- 2017 12

## Modelo de Tempo de Vida Acelerado

$$Y = \ln(T) = \boldsymbol{x}'\boldsymbol{\beta} + \sigma\nu$$

√ υ: termo de erro com distribuição apropriada

- Distribuição para υ:
  - √ Valor extremo (T é Weibull)
  - √ Normal (T é lognormal)
  - √ Log-gama (T é gama)
  - √ Logística (T é log-logística)

Análise de Sobrevivência -- 2017 13

- Modelo de tempo de vida acelerado

$$Y = \ln(T) = \boldsymbol{x}'\boldsymbol{\beta} + \sigma\nu$$

- Modelo na escala original

$$T = \exp\{\boldsymbol{x}'\boldsymbol{\beta}\}\exp\{\sigma\nu\}$$

√ Covariáveis tem efeito de acelerar (ou desacelerar) o tempo de vida

Análise de Sobrevivência -- 2017 14

- Modelo bastante utilizado na prática
  - √ Generalização em termos paramétricos
    - Acréscimo de mais um parâmetro de forma (Ex.: gama generalizada)
  - √ Generalização mais utilizada:
    - Modelo semiparamétrico de Cox (modelos acelerados aparecem como casos particulares)

Análise de Sobrevivência -- 2017 15

- Vida Acelerada – Interpretação:
  - √ Vida acelerada:
    συ : distribuição de referência quando ***x*** = 0
  - √ Escala original
    $T_0$ = exp{συ}: distribuição de referência na escala original
  - √ Probabilidade de indivíduo de referência estar vivo no tempo t:

$$S_0(t) = \mathrm{P}\{T_0 > t\} = \mathrm{P}\left\{\sigma\nu > \ln t\right\} = \mathrm{P}\left\{\nu > \frac{\ln t}{\sigma}\right\}$$

Análise de Sobrevivência -- 2017 16

√ Considerando o efeito das covariáveis

$$T \sim T_0\, e^{\boldsymbol{x}'\boldsymbol{\beta}}$$

– Covariáveis agem multiplicativamente no tempo de sobrevivência

√ Probabilidade de indivíduo com valores de covariáveis ***x*** estar vivo no tempo t:

$$S(t|\boldsymbol{x}) = \mathrm{P}\{T > t|\boldsymbol{x}\} = \mathrm{P}\left\{T_0\, e^{\boldsymbol{x}'\boldsymbol{\beta}} > t\right\}$$
$$= \mathrm{P}\left\{T_0 > t\, e^{-\boldsymbol{x}'\boldsymbol{\beta}}\right\}$$
$$= S_0\left(t\, e^{-\boldsymbol{x}'\boldsymbol{\beta}}\right)$$

Análise de Sobrevivência -- 2017 17

- Probabilidade de indivíduo com valores de covariáveis ***x*** estar vivo no tempo t:

$$S(t|\boldsymbol{x}) = S_0\left(t\, e^{-\boldsymbol{x}'\boldsymbol{\beta}}\right)$$

√ É a mesma probabilidade de indivíduo de referência estar vivo no tempo t $e^{-\boldsymbol{x}'\boldsymbol{\beta}}$

√ O tempo passa mais rapidamente por um fator $e^{-\boldsymbol{x}'\boldsymbol{\beta}}$

– (duas vezes mais rápido, metade mais rápido)

√ Caso o interesse seja probabilidade de estar morto troca-se o sinal dos β's.

Análise de Sobrevivência -- 2017 18

## Função *Baseline*

- Função de interesse de qualquer indivíduo pode ser expressa em termos de função baseline (ou de referência) correspondente:

√ Função de densidade:

$$f(t|\boldsymbol{x}) = f_0\left(t\, e^{\boldsymbol{x}'\boldsymbol{\beta}}\right) e^{\boldsymbol{x}'\boldsymbol{\beta}}$$

√ Função de risco:

$$\lambda(t|\boldsymbol{x}) = \lambda_0\left(t\, e^{\boldsymbol{x}'\boldsymbol{\beta}}\right) e^{\boldsymbol{x}'\boldsymbol{\beta}}$$

– Relação simples entre funções de sobrevivência para diferentes ***x***'s.

(dilatação do eixo do tempo)

Análise de Sobrevivência -- 2017 19

- Exemplo:

√ Multiplicador de 2 para indivíduos com covariáveis ***x***.

– Sobrevivência: $\mathrm{P}\left\{\begin{matrix}\text{indivíduo estar vivo}\\ \text{a qualquer tempo}\end{matrix}\right\} =$

$$\mathrm{P}\left\{\begin{matrix}\text{indivíduo referência estar vivo}\\ \text{com dobro desse tempo}\end{matrix}\right\}$$

– Risco:

▪ Indivíduo está exposto em qualquer tempo ao dobro de risco de indivíduo de referência duas vezes mais velho

Análise de Sobrevivência -- 2017 20

- Pode-se começar com uma dada família para dilatar o tempo e terminar com uma família completamente diferente
  - √ Nem todas as famílias são fechadas à aceleração de tempo
  - √ Exceção:
    - Dilatando uma Weibull produz-se uma outra Weibull



## Riscos Proporcionais

- Abordagem alternativa:
  - √ Assumir que o efeito das covariáveis é aumentar (ou diminuir) o risco por uma quantidade proporcional em todas as durações

$$\begin{aligned}\lambda(t|\mathbf{x}) &= \lambda_0(t)\,\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}\\ &= \lambda_0(t)\,e^{\beta_0}\,e^{\beta_1 x_1}\ldots e^{\beta_p x_p}\end{aligned}$$

  - $\lambda_0$(t): risco para indivíduo com covariáveis zero (risco baseline)
  - exp{$\boldsymbol{x}'$ $\boldsymbol{\beta}$}: risco relativo associado aos valores x das covariáveis



- Risco acumulado:
  - √ Segue a mesma relação

$$\Lambda(t|\mathbf{x}) = \Lambda_0(t)\,e^{\boldsymbol{x}'\boldsymbol{\beta}}$$

- Sobrevivência:

$$\exp\{-\Lambda(t|\boldsymbol{x})\} = \exp\{-\Lambda_0(t)\,\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}\} = \exp\{-\Lambda_0(t)\}^{\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}}$$

$$S(t|\boldsymbol{x}) = [S_0(t)]^{\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}}$$

  - √ Sobrevivência para as covariáveis $\boldsymbol{x}$:
    - Sobrevivência *baseline* elevada a uma potência



- Exemplo:
  - √ Indivíduo exposto ao dobro do risco de indivíduo de referência
    - Probabilidade de indivíduo estar vivo em qualquer tempo é o quadrado da probabilidade de o indivíduo de referência estar vivo na mesma idade
    - Dobro do risco está associado a um determinado valor de covariáveis

- Razão de risco de dois indivíduos:

$$\frac{\lambda_i(t|\boldsymbol{x}_i)}{\lambda_j(t|\boldsymbol{x}_j)} = \frac{\lambda_0(t)\ \exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}{\lambda_0(t)\ \exp\{\boldsymbol{x}_j'\boldsymbol{\beta}\}} = \frac{\exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_j'\boldsymbol{\beta}\}}$$

  √ Função apenas das covariáveis
  - (não depende do tempo)

  √ Diferentes indivíduos com covariáveis $\boldsymbol{x}_i$ e $\boldsymbol{x}_j^*$ possuem riscos proporcionais



- Relação simples em termos de risco traduz-se em relação mais complexa em termos das funções de sobrevivência

  √ Exemplo:
  - Função risco da distribuição de Weibull (baseline)

$$\lambda(t) = \alpha_0\gamma(\alpha t)^{\gamma-1}$$

  - Risco relativo multiplicativo $e^{x'\beta}$:

$$\lambda(t|\boldsymbol{x}) = \alpha_0\gamma(\alpha_0 t)^{\gamma-1}\,e^{\boldsymbol{x}'\boldsymbol{\beta}} = \alpha_0^\gamma\,e^{\boldsymbol{x}'\boldsymbol{\beta}}\,\gamma t^{\gamma-1}$$

$$= \left(\alpha_0\,e^{\frac{\boldsymbol{x}'\boldsymbol{\beta}}{\gamma}}\right)^\gamma \gamma t^{\gamma-1}$$

  - A família Weibull é fechada para a proporcionalidade de risco



- Isso não é verdade para outras distribuições

  √ Se T é log-logística com efeito multiplicativo das covariáveis na função de risco
  - Função de risco baseline multiplicada por $e^{x'\beta}$
  - Função de risco resultante não é da distribuição log-logística



## Riscos Proporcionais e Vida Acelerada

- Os modelos de riscos proporcionais e de vida acelerada coincidem?

  √ Se iniciarmos com um risco e multiplicarmos por um risco relativo e começarmos com outro risco e dilatarmos os tempo, terminaremos com a mesma distribuição?

  √ Formulação da questão:

$$\lambda_0(t)\,e^{\boldsymbol{x}'\boldsymbol{\beta}} = \lambda_0\left(t\,e^{\boldsymbol{x}'\boldsymbol{\beta}}\right)e^{\boldsymbol{x}'\boldsymbol{\beta}},\ \forall \boldsymbol{x}, t\ ?$$

- Formulação da questão:

$$\lambda_0(t)\, e^{\boldsymbol{x}'\boldsymbol{\beta}} = \lambda_0\left(t\, e^{\boldsymbol{x}'\boldsymbol{\beta}}\right) e^{\boldsymbol{x}'\boldsymbol{\beta}},\ \forall \boldsymbol{x}, t\ ?$$

  √ Condição será verdadeira $\forall \boldsymbol{x}$, então será verdadeira para $\boldsymbol{x} = \mathbf{0}$, logo o risco *baseline* deverá ser o mesmo.

  √ Qual é este risco?



- A Weibull é a única distribuição que é fechada para as famílias de vida acelerada e de risco proporcional

  √ Os parâmetros de vida acelerada e de riscos proporcionais $\boldsymbol{\beta}^*$ e $\boldsymbol{\beta}$ são proporcionais entre si, com proporcionalidade constante igual a $\gamma$

  - Iguais para $\gamma = 1$ (exponencial)
  - Neste contexto, o modelo de vida acelerado deve estar definido em termos da probabilidade de morrer, para preservar a interpretação dos coeficientes $\beta$'s como aumento do risco, consistente com os modelos de risco
    - Troca de sinal dos $\beta$'s no Modelo de Vida Acelerada)



- Exemplo:

  √ Modelo exponencial ($\gamma = 1$)

  - O dobro do risco faz o tempo ir duas vezes mais rápido

  √ Modelo Weibull com $\gamma = 2$:

  - O dobro do risco faz o tempo ir somente 40% mais rápido



## Chances Proporcionais

- Abordagem alternativa:

  √ Assumir que o efeito das covariáveis aumenta (ou diminui) as chances do desfecho, para uma dada duração, por uma quantidade proporcional

## Chances Proporcionais

- Abordagem alternativa:
  - √ Assumir que o efeito das covariáveis é aumentar (ou diminuir) as chances do desfecho por uma quantidade proporcional em todas as durações

$$\frac{1 - S(t|\boldsymbol{x})}{S(t|\boldsymbol{x})} = \frac{1 - S_0(t|\boldsymbol{x})}{S_0(t|\boldsymbol{x})}\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}$$

  - $S_0(t)$: função de sobrevivência para indivíduo com covariáveis zero (risco baseline)
  - exp{*x*' β}: multiplicador da chance associado aos valores *x* das covariáveis

Análise de Sobrevivência -- 2017 34

## Função Logit

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$$

- Logit é a inversa do sigmóide (ou função logística)

$$\log(\text{OR}) = \log\left(\frac{\frac{p}{1-p}}{\frac{q}{1-q}}\right) = \text{logit}(p) - \text{logit}(q)$$

Análise de Sobrevivência -- 2017 35

$$\frac{1 - S(t|\boldsymbol{x})}{S(t|\boldsymbol{x})} = \frac{1 - S_0(t)}{S_0(t)}\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}$$

- Tomando o log da chance:

$$\text{logit}\,(1 - S(t|\boldsymbol{x})) = \text{logit}\,(1 - S_0(t)) + \boldsymbol{x}'\boldsymbol{\beta}$$

  - √ Efeito da covariável é linear na escala logit
    - Log chance do desfecho (em geral, morrer)
  - √ Modelo pode ser definido em termos de chance de sobreviver à duração t
    - Trocar sinal dos β's.

Análise de Sobrevivência -- 2017 36

## Modelo de Chances Proporcionais Log-Logístico

- Chance baseline

$$\text{S}_0(t) = \frac{1}{1 + (\alpha t)^\gamma} \qquad \frac{1 - S_0(t)}{S_0(t)} = (\alpha t)^\gamma$$

$$1 - \text{S}_0(t) = \frac{(\alpha t)^\gamma}{1 + (\alpha t)^\gamma}$$

- Multiplicando a chance *baseline* tem-se outro modelo log-logístico, com

$$\alpha^* = \alpha \exp\left\{\frac{\boldsymbol{x}'\boldsymbol{\beta}}{\gamma}\right\}$$

$$\gamma^* = \gamma$$

Análise de Sobrevivência -- 2017 37

- Comentários:
  - √ A família log-logística é fechada para a proporcionalidade de chances
  - √ Isto não é verdade para outras distribuições
- Chances proporcionais e vida acelerada
  - √ Os modelos de chances proporcionais e de vida acelerada coincidem se e somente se a distribuição *baseline* for log-logística



## Estimação dos Parâmetros do Modelo

- Estimação dos efeitos das covariáveis
  (β's da regressão)
  - √ Métodos dos mínimos quadrados
    - Propriedades desejáveis na presença de erros com distribuição normal
    - Inadequado na ausência de normalidade e na presença de censuras
  - √ Método da máxima verossimilhança
    - Adequado na presença de censura e ausência de normalidade



## Estimação de Máxima Verossimilhança

- Todos os modelos podem ser ajustados maximizando a função de verossimilhança apropriada
  - √ Dados: pares ($t_i$, $\delta_i$):
    - $t_i$: tempo até desfecho ou censura
    - $\delta_i$: indicador de desfecho



- Forma geral da função de verossimilhança sob censura geral não-informativa:

$$\mathrm{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} [\lambda(t_i|\boldsymbol{x}_i)]^{\delta_i}\, \mathrm{S}(t_i|\boldsymbol{x}_i)$$

  - √ Em geral, a função de verossimilhança é maximizada numericamente
    - Newton-Raphson ou outro

## Modelo Linear para Dados de Sobrevivência

## Modelos de Regressão para Dados de Sobrevivência

√ Forma de incorporar efeito de covariáveis

- Classes de modelos:
  - √ Modelos paramétricos:
    (modelos de tempo de vida acelerado)
    - Mais eficientes porém menos flexíveis
  - √ Modelos semiparamétricos:
    (modelo de regressão de Cox)
    - São mais flexíveis
    - Fácil a incorporação de covariáveis dependentes do tempo

Análise de Sobrevivência -- 2017 43

## Modelo de Regressão Linear

√ Modelo mais conhecido em Estatística

√ Modelo linear associa resposta com as variáveis explicativas (ou covariáveis)

$$Y = \underbrace{\beta_0 + \beta_1 x}_{\text{Componente determinístico}} + \underbrace{\epsilon}_{\text{Componente estocástico}}$$

- Y: variável resposta
- x: covariável (ou variável explicativa, ou variável independente)
- $\beta_0$, $\beta_1$: parâmetros a serem estimados
- $\varepsilon$: erro aleatório (com distribuição normal)

Análise de Sobrevivência -- 2017 44

- Estrutura de um modelo de regressão:
  - √ Componente aleatório:
    - Descreve probabilisticamente o comportamento da resposta
  - √ Componente determinístico ou estrutural:
    - Descreve a relação entre os parâmetros da distribuição e as covariáveis

Análise de Sobrevivência -- 2017 45

## Modelo de Regressão de Sobrevivência

- Estima efeito das covariáveis sobre o tempo de sobrevivência
  - √ Variável resposta: tempo até desfecho e censura
    - – Em geral, o tipo de resposta e o comportamento da variáveis não permitem a utilização direta do modelo linear

Análise de Sobrevivência -- 2017 46

---

- √ Distribuição da resposta tende a ser assimétrica na direção dos maiores tempos de sobrevivência
  - – Inapropriado o uso da distribuição normal para o comportamento estocástico

Análise de Sobrevivência -- 2017 47

---

- Formas de modelagem estatística em análise de sobrevivência:
  - √ Transformar a resposta para tentar retornar ao modelo linear, ou
  - √ Modificar o modelo, usando
    - – Componente determinístico não linear nos parâmetros
    - – Componente estocástico com distribuição assimétrica
  - √ As duas formas são equivalentes

Análise de Sobrevivência -- 2017 48

---

- Equivalência:
  - – Usar como componente determinístico $\exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$ (transformação logarítmica da resposta)
  - – Distribuição lognormal para os erros (há outras distribuições assimétricas disponíveis)

Análise de Sobrevivência -- 2017 49

## Modelo de Regressão do Risco

- Função de risco de um indivíduo no tempo t, dado o vetor de covariáveis fixas ***x***:

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\ \exp\{\boldsymbol{x}'\boldsymbol{\beta}\}$$

  - √ **β**: vetor de coeficientes a estimar
  - √ $\lambda_0$(t): risco basal de um indivíduo que possui exp(**x'β**)=1



- Ajuste de modelo de regressão paramétrico
  - √ Necessário supor uma distribuição de probabilidade para o tempo de sobrevivência
  - √ Parâmetros da distribuição adotada definem forma dessa distribuição
    - – Não confundir com os parâmetros da regressão



- Modelo de riscos proporcionais
  - √ Efeito multiplicativo exponencial

$$\begin{aligned}\lambda(t|\mathbf{x}) &= \lambda_0(t)\ \exp\{\boldsymbol{x}'\boldsymbol{\beta}\}\\ &= \lambda_0(t)\, \mathrm{e}^{\beta_0}\, \mathrm{e}^{\beta_1 x_1} \ldots \mathrm{e}^{\beta_p x_p}\end{aligned}$$

  - √ Efeito das covariáveis:
    - – Aumentar ou diminuir o risco por quantidade proporcional em todas as durações



- Razão de risco de dois indivíduos:

$$\frac{\lambda_i(t|\boldsymbol{x}_i)}{\lambda_j(t|\boldsymbol{x}_j)} = \frac{\lambda_0(t)\ \exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}{\lambda_0(t)\ \exp\{\boldsymbol{x}_j'\boldsymbol{\beta}\}} = \frac{\exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_j'\boldsymbol{\beta}\}}$$

  - √ Função apenas das covariáveis
    - – (não depende do tempo)
  - √ Diferentes indivíduos com covariáveis $\boldsymbol{x}_i$ e $\boldsymbol{x}_j^*$ possuem riscos proporcionais

## Modelo de Tempo de Vida Acelerada

- Modelo na log-escala

$$Y = \ln(T) = \boldsymbol{x}'\boldsymbol{\beta} + \sigma\nu$$

√ συ : distribuição de referência quando ***x*** = 0

- Modelo na escala original

$$T = \exp\{\boldsymbol{x}'\boldsymbol{\beta}\}\exp\{\sigma\nu\}$$

- Considerando o efeito das covariáveis

$$T \sim T_0\, e^{\boldsymbol{x}'\boldsymbol{\beta}}$$

  – Covariáveis agem multiplicativamente no tempo de sobrevivência

Análise de Sobrevivência -- 2017 54

- Probabilidade de indivíduo com valores de covariáveis ***x*** estar vivo no tempo t:

$$S(t|\boldsymbol{x}) = S_0\left(t\, e^{-\boldsymbol{x}'\boldsymbol{\beta}}\right)$$

√ É a mesma probabilidade de indivíduo de referência estar vivo no tempo t $e^{-x'\beta}$

√ O tempo passa mais rapidamente por um fator $e^{-x'\beta}$

  – (duas vezes mais rápido, metade mais rápido)

√ Caso o interesse seja probabilidade de estar morto troca-se o sinal dos β's.

Análise de Sobrevivência -- 2017 55

## Modelo de Regressão Exponencial

√ Modelo mais simples em Análise de Sobrevivência

  – Componente determinístico: $\exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$
  – Componente aleatório: ε ~ exponencial padrão

√ Modelo de regressão exponencial

$$T = \exp\{\boldsymbol{\beta}'\boldsymbol{x}\}\epsilon$$

  – Relação não linear entre T e x
  – Componente de erro com distribuição assimétrica
  – Função de ligação: logarítmica
  – Resposta com distribuição exponencial

Análise de Sobrevivência -- 2017 56

- Modelo é linearizável

$$\begin{aligned} Y = \ln(T) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \nu \\ &= \boldsymbol{\beta}'\boldsymbol{x} + \nu \end{aligned}$$

√ υ ~ distribuição de valor extremo padrão

$$\nu = \ln(\epsilon), \text{ com } f(\nu) = \exp\{\nu - exp\{\nu\}\}$$

$$\boldsymbol{x}' = (1, x_1, x_2, \ldots, x_p)$$
$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$$

√ Covariável x atua linearmente em Y e multiplicativamente em T

Análise de Sobrevivência -- 2017 57

- Parâmetro α depende das covariáveis

$$\alpha(\boldsymbol{x}) = \exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$$

  - β: estimativa dos efeitos das covariáveis
  - α: parâmetro que define o risco exponencial (assumido constante)
- Função de sobrevivência
  - √ Para Y, condicionada a x

$$S(y|\boldsymbol{x}) = \exp\left\{-\exp\left\{y + \boldsymbol{\beta}'\boldsymbol{x}\right\}\right\}$$

  - √ Para T, condicionada a x:

$$S(t|\boldsymbol{x}) = \exp\left\{-\boldsymbol{\beta}'\boldsymbol{x}\, t\right\}$$



- Modela o risco como uma função das covariáveis
  - √ Parâmetro α depende das covariáveis

$$\alpha(\boldsymbol{x}) = \exp\{\boldsymbol{x}'\boldsymbol{\beta}\}$$

  - √ Risco considerado constante para qualquer tempo t, dado os valores ***x*** para as covariáveis
  - √ Exemplos:
    - Idade aumenta o risco de óbito por insuficiência renal
    - Categoria “usuário de drogas injetáveis” tem risco maior do que a categoria “transmissão sexual” para o tempo de sobrevida com Aids



- Funções para o modelo de regressão exponencial
  - √ Função de risco

$$\lambda(t|\boldsymbol{x}) = \alpha(\boldsymbol{x}) = \exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$$

  - √ Função de sobrevivência:

$$S(t|\boldsymbol{x}) = \exp\{-\alpha(\boldsymbol{x})t\} = \exp\{-\exp\{\boldsymbol{\beta}'\boldsymbol{x}\}t\}$$



## Estimação dos Parâmetros do Modelo

- Estimação dos efeitos das covariáveis (β’s da regressão)
  - √ Método da máxima verossimilhança
    - Adequado na presença de censura e ausência de normalidade

## Exemplo

- Pacientes com insuficiência renal submetidos à hemodiálise
  - √ Período: janeiro/1988 a outubro/2001
  - √ Dados: Sistema Apac
  - √ Coorte com 6.805 pacientes, no Rio de Janeiro
    - – 1.603 óbitos
  - √ Objetivo:
    - – Estimar o efeito da idade no risco de morte



- Hipótese:
  - √ Risco de morrer é constante ao longo do tempo
    (modelo exponencial)
- Modelo: $\lambda(t|idade) = \exp\{\beta_0 + \beta_1 \times idade\}$
- Ajuste do modelo

```
> library(survival)
>
> dialise <- read.csv("dados/dialise.csv", header = T)
> reg.exp <- survreg(Surv(tempo, status)~idade, data = dialise,
+  dist='exponential')
> reg.exp$coefficients
(Intercept)       idade
 6.13552416 -0.03700427
```

$\hat{\beta}_0 = -6,136$

$\hat{\beta}_1 = 0,037$

  - – Importante: sinal do parâmetro está invertido em relação ao utilizado no texto



- Estimativa do risco no tempo t:

$$\hat{\lambda}(t|idade) = \exp\{-6,136 + 0,037 \times idade\}$$

  - √ Para cada ano de vida o risco aumenta
    exp{0,037} = 1,037 vezes
  - √ Comparação indivíduos com 30 e 70 anos,

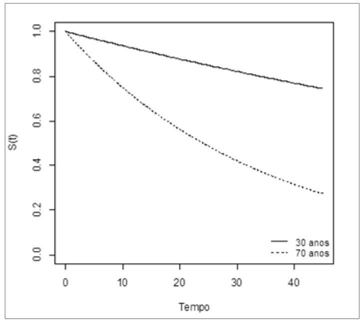
$$\frac{\hat{\lambda}(t|idade = 70)}{\hat{\lambda}(t|idade = 30)} = \frac{\exp\left\{\hat{\beta}_0 + \hat{\beta}_1 \times 70\right\}}{\exp\left\{\hat{\beta}_0 + \hat{\beta}_1 \times 30\right\}}$$

$$= \exp\{0,037 \times 40\} = 4,4$$

    - – Risco de indivíduo em diálise morrer aos 70 anos é 4,4 vezes maior do que o de morrer aos 30
    - – Indivíduos com diferença de 40 anos entre si geram um risco relativo estimado em 4,4 (modelo de riscos proporcionais)



- Curva de sobrevivência (30 e 70 anos)




  - √ Indivíduo com 70 anos progride 4,4 vezes mais rápido ao longo do tempo
    - – acelera a uma taxa constante

## Modelo de Regressão de Weibull

√ Forma de generalizar o modelo exponencial

- Inclusão de parâmetro extra de escala, na presença de covariáveis

$$Y = \ln(T) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \sigma\nu$$
$$= \boldsymbol{\beta}'\boldsymbol{x} + \sigma\nu$$

- com: $\upsilon$ ~ valor extremo padrão, e
- Y = ln(T) ~ Valor extremo com parâmetro de escala $\sigma$ e parâmetro de locação $\mu$.

Análise de Sobrevivência -- 2017 68

- Modelo na escala original

$$T = \exp\{\boldsymbol{\beta}'\mathbf{x}\}\exp\{\sigma\nu\}$$

√ T ~ Weibull ($\alpha$, $\gamma$), com parâmetro $\alpha$ sendo modelado pelas covariáveis

$$\alpha(\boldsymbol{x}) = \exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$$

- **β**': estimativa dos efeitos das covariáveis
- **x**': valores das covariáveis

√ Relação entre os parâmetros da Weibull escala original) e valor extremo (log-escala)

$$\alpha = \exp\{-\mu\}$$
$$\gamma = \frac{1}{\sigma}$$

Análise de Sobrevivência -- 2017 69

- Risco e sobrevivência modelados em função das covariáveis

√ Função de risco

$$\lambda(t|\boldsymbol{x}) = \gamma t^{\gamma-1}\alpha(\boldsymbol{x})^{\gamma}$$
$$= \gamma t^{\gamma-1}[\exp\{(\boldsymbol{\beta}'\boldsymbol{x})\}]^{\gamma}$$

√ Caso $\alpha$ = g(***x***) (modelo em questão)

- Os riscos para diferentes indivíduos são proporcionais

√ Caso $\gamma$ também varie com as covariáveis [$\gamma$= h(***x***)]

- Modelo de regressão deixa de ser de riscos proporcionais

Análise de Sobrevivência -- 2017 70

- Parâmetro $\gamma$:

√ $\gamma > 1$: o risco aumenta no tempo

√ $\gamma < 1$: o risco diminui no tempo

√ $\gamma = 1$: a função de risco é constante (exponencial)

Análise de Sobrevivência -- 2017 71

- Função de sobrevivência
  - √ Para Y, condicionada a x

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

  - √ Para T, condicionada a x:

$$S(t|\boldsymbol{x}) = \exp\{-[\alpha(\boldsymbol{x})\,t]^{\gamma}\} = \exp\{-\left(\exp\{\boldsymbol{\beta}'\boldsymbol{x}\}\,t\right)^{\gamma}\}$$

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



## Exemplo

- Pacientes com insuficiência renal submetidos à hemodiálise
  - √ Período: janeiro/1988 a outubro/2001
  - √ Dados: Sistema Apac
  - √ Coorte com 6.805 pacientes, no Rio de Janeiro
    - – 1.603 óbitos
  - √ Objetivo:
    - – Estimar o efeito da idade no risco de morte



- Hipótese:
  - √ Tempo até óbito tem distribuição de Weibull
- Modelo: $\lambda(t|idade) = \gamma t^{\gamma-1}[\exp\{\beta_0 + \beta_1 \times idade\}]^{\gamma}$
- Ajuste do modelo

```
> dialise <- read.csv("dados/dialise.csv", header = T)
> reg.weib <- survreg(Surv(tempo, status)~idade, data = dialise,
+   dist='weibull')
> summary(reg.weib)
Call:
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
    dist = "weibull")
              Value Std. Error      z        p
(Intercept)  6.7512    0.14693  45.95 0.00e+00
idade       -0.0436    0.00224 -19.50 1.12e-84
Log(scale)   0.1987    0.02083   9.54 1.49e-21
Scale= 1.22
Weibull distribution
Loglik(model)= -7877.7   Loglik(intercept only)= -8104.2
        Chisq= 453.16 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

$$\hat{\beta}_0 = -6,751$$
$$\hat{\beta}_1 = 0,0436$$
$$\hat{\sigma} = \exp\{0,1987\} = 1,22$$

  - – Importante: inverter o sinal do parâmetro



- Análise resultados

```
> summary(reg.weib)
Call:
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
    dist = "weibull")
              Value Std. Error      z        p
(Intercept)  6.7512    0.14693  45.95 0.00e+00
idade       -0.0436    0.00224 -19.50 1.12e-84
Log(scale)   0.1987    0.02083   9.54 1.49e-21
Scale= 1.22
Weibull distribution
Loglik(model)= -7877.7   Loglik(intercept only)= -8104.2
        Chisq= 453.16 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

$$\hat{\beta}_0 = -6,751$$
$$\hat{\beta}_1 = 0,0436$$
$$\hat{\sigma} = \exp\{0,1987\} = 1,22$$

  - √ Teste de significância do parâmetro
    - – Efeito da idade ($\beta_1$) é significativo
    - – Parâmetro de forma ($\gamma$) é significativamente diferente de 1
  - √ Parâmetros da distribuição do tempo: $\hat{\alpha} = \exp\{-\boldsymbol{x}'\boldsymbol{\beta}\}$ $\hat{\gamma} = \dfrac{1}{\hat{\sigma}} = 0,820$

LFB1

√ Parâmetros da regressão: $\hat{\beta}_0 = -6,751$, $\hat{\beta}_1 = 0,0436$, $\hat{\sigma} = \exp\{0,1987\} = 1,22$

√ Parâmetros da distribuição de T: $\hat{\alpha} = \exp\{-\boldsymbol{x}'\boldsymbol{\beta}\}$, $\hat{\gamma} = \frac{1}{\hat{\sigma}} = 0,820$

- Estimativa do risco no tempo t:

$$\hat{\lambda}(t|idade) = (0,82)t^{-0,18}[\exp\{-6,751 + 0,044 \times idade\}]^{0,82}$$

√ Para cada ano de vida o risco aumenta

$(\exp\{0,0436\})^{0,82} = 1,036$ vezes

√ Parâmetro de forma ($\gamma = 0,82$)

- O risco de morrer diminui à medida que aumenta o tempo do paciente em hemodiálise ($\gamma < 1$)
- Possível motivo:
  - Baixo preparo inicial dos pacientes (hemodiálise em emergência) (excesso de óbitos no início do processo)

76

Análise de Sobrevivência -- 2017

√ Comparação entre indivíduos com 30 e 70 anos,

$$\frac{\hat{\lambda}(t|idade=70)}{\hat{\lambda}(t|idade=30)} = \frac{\gamma t^{\gamma-1}[\exp\left\{\hat{\beta}_0 + \hat{\beta}_1 \times 70\right\}]^{\gamma}}{\gamma t^{\gamma-1}[\exp\left\{\hat{\beta}_0 + \hat{\beta}_1 \times 30\right\}]^{\gamma}}$$

$$= \left[\frac{\exp\{-6,7512 + 0,0436 \times 70\}}{\exp\{-6,7512 + 0,0436 \times 30\}}\right]^{0,82}$$

$$= [\exp\{0,0436 \times 40\}]^{0,82} = 4,11$$

- Risco de indivíduo em diálise morrer aos 70 anos é 4,11 vezes maior do que o de morrer aos 30
- Quaisquer dois indivíduos com diferença de 40 anos entre si gera um risco relativo estimado em 4,11

  (modelo de riscos proporcionais)

Análise de Sobrevivência -- 2017

77

## Estratificação

- Utilização de estratos na modelagem

√ Subgrupos populacionais com parâmetros $\gamma$ diferentes nas distribuições de seus tempos

√ Parâmetros regressão comuns em todos grupos

- Mesmo efeito das covariáveis

√ Curva de risco basal variando entre os estratos

- Melhoria na qualidade da estimativa do efeito das covariáveis

√ Nas situações em que for indicado:

- Estimar o efeito das covariáveis para cada estrato

Análise de Sobrevivência -- 2017

78

- Causa da insuficiência: doença congênita

√ Padrão de sobrevida pode ser muito diferente

(mais transplantes em pacientes mais jovens)

- Análise dessa hipótese

√ Utilizar doença congênita no modelo

- Aditivamente, como estrato e com interação entre o estrato e a idade

```
> congenita.simples <- survreg(Surv(tempo, status)~idade + congenita,
+ data = dialise, dist='weibull')
> congenita.strat <- survreg(Surv(tempo, status)~idade + strata(congenita),
+     data = dialise, dist='weibull')
> congenita.inter <- survreg(Surv(tempo, status)~idade*strata(congenita),
+ data = dialise, dist='weibull')
```

Análise de Sobrevivência -- 2017

79

**Slide 76**

---

**LFB1** Conferir interpretação exp{0,0436} (livro, pág 156) ou (exp{0,0436})^0,82 (risco relativo, página 156)

Lupercio; 13/11/2014

- Doença congênita incluída aditivamente

```
> summary(congenita.simples)
Call:
survreg(formula = Surv(tempo, status) ~ idade + congenita, data = dialise,
    dist = "weibull")
             Value Std. Error      z        p
(Intercept)  6.7357    0.14647  45.99 0.00e+00
idade       -0.0437    0.00223 -19.58 2.20e-85
congenita    0.9781    0.27457   3.56 3.67e-04
Log(scale)   0.1971    0.02082   9.47 2.82e-21
Scale= 1.22
Weibull distribution
Loglik(model)= -7869.3   Loglik(intercept only)= -8104.2
        Chisq= 469.95 on 2 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

Coeficientes significativos

$$\hat{\beta}_0 = -6,7357$$
$$\hat{\beta}_{\text{idade}} = 0,0437$$
$$\hat{\beta}_{\text{cong}} = -0,9781$$
$$\hat{\sigma}_{\text{global}} = 1,22$$

$$\hat{\gamma}_{\text{global}} = \frac{1}{\hat{\sigma}} = 0,820$$



## Modelo Simples – Comentários

- Compatibilização entre saída do R e texto
  - √ Parâmetro α: $\hat{\alpha} = \exp\{-\boldsymbol{x}'\boldsymbol{\beta}\}$
  - √ Parâmetro γ: $\hat{\gamma} = \frac{1}{\hat{\sigma}} = \frac{1}{\text{scale}}$
- Erro padrão e teste de significância (γ=1)

$$\ln(\text{scale}) = \ln\left(\frac{1}{\hat{\gamma}}\right)$$

- Risco de morrer diminui à medida que aumenta o tempo do paciente em hemodiálise (γ < 1)
  - √ Sem distinção entre os dois grupos



- Estratificação por doença congênita

```
> summary(cong.strat)
Call:
survreg(formula = Surv(tempo, status) ~ idade + strata(congenita),
    data = dialise, dist = "weibull")
              Value Std. Error      z        p
(Intercept)  6.7385    0.14645  46.01 0.00e+00
idade       -0.0434    0.00223 -19.50 1.17e-84
congenita=0  0.2047    0.02087   9.81 1.02e-22
congenita=1 -0.2716    0.11620  -2.34 1.94e-02

Scale:
congenita=0 congenita=1
      1.227       0.762

Weibull distribution
Loglik(model)= -7870.3   Loglik(intercept only)= -8097
        Chisq= 453.49 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

Log scale

Coeficientes significativos

$$\hat{\gamma}_{\text{cong}=0} = \frac{1}{1,227} = 0,814$$
$$\hat{\gamma}_{\text{cong}=1} = \frac{1}{0,762} = 1,312$$

$$\hat{\beta}_0 = -6,7385$$
$$\hat{\beta}_{\text{idade}} = 0,0434$$



## Modelo Estratificado – Comentários

- Efeito de variável de estratificação não é estimado
- Níveis com parâmetros de escala diferentes
  - √ Congênita = 0
    - Risco de morrer diminui c/ tempo de hemodiálise
  - √ Congênita = 1
    - Risco de morrer aumenta c/ tempo de hemodiálise

- Interação entre doença congênita e idade

```
> summary(congenita.inter)
Call:
survreg(formula = Surv(tempo, status) ~ idade * strata(congenita),
    data = dialise, dist = "weibull")
                                    Value Std. Error       z        p
(Intercept)                       6.74462    0.14648  46.046 0.00e+00
idade                            -0.04375    0.00223 -19.602 1.48e-85
idade:strata(congenita)congenita=1  0.00735  0.00446   1.649 9.92e-02
congenita=0                       0.20041    0.02091   9.583 9.45e-22
congenita=1                      -0.11184    0.15546  -0.719 4.72e-01

Scale:
congenita=0 congenita=1
      1.222       0.894

Weibull distribution
Loglik(model)= -7868.1   Loglik(intercept only)= -8097
        Chisq= 457.81 on 2 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 17
n= 6805
```

Interação não significativa

Log scale

$$\hat{\gamma}_{\text{cong}=0} = \frac{1}{1,222} = 0,818$$
$$\hat{\gamma}_{\text{cong}=1} = \frac{1}{0,894} = 1,118$$

$$\hat{\beta}_0 = -6,7446$$
$$\hat{\beta}_{\text{idade}} = 0,04375$$
$$\hat{\beta}_{(\text{idade})\times(\text{congenita})} = 0,040735 = 0$$

Análise de Sobrevivência -- 2017 84

## Modelo com Iteração – Comentários

- Efeito da interação entre idade e a doença congênita não é significativo
- Níveis com parâmetros de escala diferentes
  - √ Congênita = 0
    - – Risco de morrer diminui c/ tempo de hemodiálise
  - √ Congênita = 1
    - – Risco de morrer aumenta c/ tempo de hemodiálise

Análise de Sobrevivência -- 2017 85

- Resumos dos modelos – Saídas do R

| | Modelos | | |
|---|---|---|---|
| | cong.simples | cong.strat | cong.inter |
| Coeficientes (β) | | | |
| Intercept | 6.7357 | 6.7385 | 6.74462 |
| idade | −0.0437 | −0.0434 | −0.04375 |
| congenita | 0.9781 | | |
| idade:strata(congenita)=1 | | | 0.00735 |
| Scale | | | |
| Global | 1,22 | | |
| congenita=0 | | 1.227 | 1.222 |
| Congenita=1 | | 0.762 | 0.894 |

Análise de Sobrevivência -- 2017 86

- Estimação dos parâmetros de forma

$$\hat{\gamma} = \frac{1}{\hat{\sigma}} = \frac{1}{\text{scale}}$$

```
> # Parametros de Forma
> # Modelo simples
> 1/cong.simples$scale
[1] 0.8210882
> # Modelo estratificado
> 1/cong.strat$scale
congenita=0 congenita=1
  0.8148844   1.3121269
> # Modelo estratificado c/ interação
> 1/cong.inter$scale
congenita=0 congenita=1
  0.8183929   1.1183291
```

Análise de Sobrevivência -- 2017 87

- Estimação dos riscos relativos

$\lambda(t|\boldsymbol{x}) = \gamma t^{\gamma-1}[\exp\{(\boldsymbol{\beta}'\boldsymbol{x})\}]^{\gamma}$

$[\exp\{\beta\}]^{\gamma}$

```
> # Risco Relativo
> # Modelo simples
> exp(-cong.simples$coeff[-1])^(1/cong.simples$scale)
    idade congenita
1.0365145 0.4479174
> # Modelo estratificado para congenita=0
> exp(-cong.strat$coeff[-1])^(1/cong.strat$scale[1])
   idade
1.035988
> # Modelo estratificado para congenita=1
> exp(-cong.strat$coeff[-1])^(1/cong.strat$scale[2])
   idade
1.058581
> # Modelo estratificado com interação
> # congenita = 0
> exp(-cong.inter$coeff[2])^(1/cong.inter$scale[1])
   idade
1.036457
> # Modelo estratificado com interação
> # congenita = 1
> exp(-sum(cong.inter$coeff[2:3]))^(1/cong.inter$scale[2])
congenita=1
   1.041548
```

Análise de Sobrevivência -- 2017 88

- Estimativas do risco relativo $[\exp(\beta)]^{\gamma}$

| | Modelos | | |
|---|---|---|---|
| Variáveis | Simples | Estratificado | Interação |
| **Idade** | 1,036 | | |
| Congênita = Não | | 1,034 | 1,036 |
| Congênita = Sim | | 1,058 | 1,041 |
| Congênita | 0,448 | | |
| Forma (g) | 0,821 | | |
| Congênita = Não | | 0,815 | 0,818 |
| Congênita = Sim | | 1,312 | 1,118 |

Análise de Sobrevivência -- 2017 89

## Comentários

- Parâmetro de forma ($\gamma$)
  - √ Bastante diferente em estrato de insuficiência renal causada por doença congênita
  - √ $\gamma_{\text{global}} \approx \gamma_{\text{congenita=0}}$
    - Apenas 2% dos pacientes com doença congênita
- Aumento do risco de morte por ano de vida
  - √ Maior em estrato de doença congênita
  - √ Mantém-se próximos em todos os modelos para estrato sem doença congênita

Análise de Sobrevivência -- 2017 90

- Equações de risco dos modelos estimados
  - √ Modelo simples

$$\hat{\lambda}(t|\boldsymbol{x}) = (0,821)t^{0,821-1}[\exp\{-6,735 + 0,044 \times idade - 0,978 \times congenita\}]^{0,821}$$

  - √ Modelo estratificado
    - Congênita = Não

$$\hat{\lambda}(t|\boldsymbol{x}) = (0,815)t^{0,815-1}[\exp\{-6,739 + 0,0433 \times idade\}]^{0,815}$$

    - Congênita = Sim

$$\hat{\lambda}(t|\boldsymbol{x}) = (1,312)t^{1,312-1}[\exp\{-6,739 + 0,0433 \times idade\}]^{1,312}$$

Análise de Sobrevivência -- 2017 91

√ Modelo estratificado com interação

– Congênita = Não

$$\hat{\lambda}(t|\boldsymbol{x}) = (0,818)t^{0,818-1}[\exp\{-6,745 + 0,044 \times idade\}]^{0,818}$$

– Congênita = Sim

$$\hat{\lambda}(t|\boldsymbol{x}) = (1,118)t^{1,118-1}[\exp\{-6,745 + (0,044 - 0,007) \times idade\}]^{1,118}$$

Análise de Sobrevivência -- 2017 92

## Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 193 dias
    - – Óbitos: 90
    - – Censuras: 103
  - √ Objetivo:
    - – Estimar efeito do tratamento controlado por idade e sexo

Análise de Sobrevivência -- 2017 93

- Modelo de regressão – Exponencial:

```
> modhiv.exp <- survreg(Surv(tempo, status)~idade, data = hiv,
+  dist='exponential')
> modhiv.exp <- survreg(Surv(tempo, status)~idade + sexo + tratam,
+     data = hiv, dist='exponential')
> summary(modhiv.exp)

Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "exponential")
               Value Std. Error      z        p
(Intercept)  6.11597     0.4921 12.429 1.83e-35
idade        0.00829     0.0112  0.739 4.60e-01
sexoM       -0.19455     0.2841 -0.685 4.93e-01
tratam       1.37541     0.1920  7.164 7.83e-13

Scale fixed at 1

Exponential distribution
Loglik(model)= -743.5   Loglik(intercept only)= -774.6
        Chisq= 62.3 on 3 degrees of freedom, p= 1.9e-13
Number of Newton-Raphson Iterations: 5
n= 193
```

Análise de Sobrevivência -- 2017 94

- Modelo de regressão – Weibull:

```
> modhiv.wei <- survreg(Surv(tempo, status)~idade + sexo + tratam
+     data = hiv, dist='weibull')
> summary(modhiv.wei)

Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "weibull")
               Value Std. Error      z        p
(Intercept)  6.06842     0.5674 10.695 1.07e-26
idade        0.00951     0.0130  0.731 4.65e-01
sexoM       -0.23627     0.3277 -0.721 4.71e-01
tratam       1.48608     0.2273  6.538 6.25e-11
Log(scale)   0.14185     0.0862  1.647 9.97e-02

Scale= 1.15

Weibull distribution
Loglik(model)= -742   Loglik(intercept only)= -770.3
        Chisq= 56.64 on 3 degrees of freedom, p= 3.1e-12
Number of Newton-Raphson Iterations: 5
n= 193
```

Análise de Sobrevivência -- 2017 95

## Comandos em R

`m`: objeto que contém resultado do ajuste

> `summary(m)`

√ `Call`: modelo ajustado

√ `Value`: parâmetros estimados

√ `Std. Error`: erro padrão do parâmetro estimado

√ `z`: estatística para estimativa dos parâmetros

√ `p`: p-valor da estatística z

√ `Scale`: Weibull = 1/g; Exponencial = 1

√ `Loglik(model)`: log-verossimilhança modelo completo

√ `Loglik(intercept only)`: log-verossimilhança modelo nulo

√ `Number of Newton-Raphson Iterations`: estimação

√ `n`: número de observações

96

Análise de Sobrevivência -- 2017

- Ajuste dos modelos paramétricos de regressão
  - √ Formula = Surv(temp, status) ~
    - ~ var1 + var2 + ... (covariáveis)
    - ~ 1 (sem covariáveis)
    - ~ var1*var2 (interação entre variáveis)
    - ~ (var1:var2) (somente inclui o termo de interação)

Análise de Sobrevivência -- 2017 97

- Parametrização dos modelos Weibull no R:
  - √ Na presença de covariáveis, os parâmetros de regressão β ficam com o sinal inverso ao que aparece na saída do pacote

Análise de Sobrevivência -- 2017 98

## Seleção de Modelos

## Seleção de Modelos Paramétricos

- Para comparar modelos paramétricos aninhados
  - √ Teste da razão de verossimilhanças
  - √ Teste de Wald



## Teste da Razão de Verossimilhanças

- Usado para comparar modelos aninhados
  - √ Modelo com maior número de parâmetros contém todos os parâmetros do modelo menor
  - √ Estatística de teste: $\mathrm{RV} = 2\,(l_{maior} - l_{menor})$
  - √ Distribuição amostral: $\mathrm{RV} \sim \chi^2_{gl}$
    - – gl: diferença entre o número de parâmetros dos modelos



## Exemplo

- Dados de Aids
  - √ Aplicação:
    - – Comparar modelos exponencial e Weibull
  - √ Modelos:
    - – $H_0$: modhiv.exp = modhiv.wei ($\gamma = 1$)
  - √ Estatística de teste:

$$\mathrm{RV} = 2\,(l_{weibull} - l_{exponencial}) = 2[(-742) - (-743,5)] = 3$$

  - √ Comando R:

```
> anova(modhiv.exp, modhiv.wei)
                 Terms Resid. Df    -2*LL Test Df Deviance   Pr(>Chi)
1 idade + sexo + tratam       189 1486.942      NA       NA         NA
2 idade + sexo + tratam       188 1484.049    =  1 2.893647 0.08892942
```



- Saída do modelo Weibull

```
> modhiv.wei
Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "weibull")
Coefficients:
 (Intercept)        idade        sexoM       tratam
 6.068424160  0.009505648 -0.236267819  1.486083990
Scale= 1.152408

Loglik(model)= -742   Loglik(intercept only)= -770.3
        Chisq= 56.64 on 3 degrees of freedom, p= 3.1e-12
n= 193
```

  - √ Hipóteses:
    - – $H_0$: modhiv.wei não é significativo (modelo nulo)
    - – $H_1$: modhiv.wei é significativo

$$\mathrm{RV} = 2\,[l_{estimado} - l_{nulo}] = 2[(-742) - (-770,3)] = 56,6$$

## Teste de Wald

- Testa a significância individual de covariáveis
  - √ $H_0$: $\beta_j = 0$ vs. $\beta_j \neq 0$
  - √ Estatística de teste: $z_j = \frac{\hat{\beta}_j}{\text{ep}(\hat{\beta}_j)}$
  - √ Distribuição amostral (sob $H_0$): $Z_j \sim N(0, 1)$

Análise de Sobrevivência -- 2017 104

- Modelo de regressão paramétrico – Weibull:

```
> modhiv.wei <- survreg(Surv(tempo, status)~idade + sexo + tratam,
+     data = hiv, dist='weibull')
> summary(modhiv.wei)

Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "weibull")
               Value Std. Error      z        p
(Intercept)  6.06842     0.5674 10.695 1.07e-26
idade        0.00951     0.0130  0.731 4.65e-01
sexoM       -0.23627     0.3277 -0.721 4.71e-01
tratam       1.48608     0.2273  6.538 6.25e-11
Log(scale)   0.14185     0.0862  1.647 9.97e-02

Scale= 1.15

Weibull distribution
Loglik(model)= -742   Loglik(intercept only)= -770.3
        Chisq= 56.64 on 3 degrees of freedom, p= 3.1e-12
```

Único efeito significativo:
- tratamento

  - Número de medicamentos antirretrovirais tem efeito protetor no risco de morte dos pacientes da coorte

Análise de Sobrevivência -- 2017 Prof. Lupércio 105

## Adequação do Modelo Ajustado

## Qualidade do Ajuste

- Avaliação da qualidade do ajuste de um modelo paramétrico
  - √ Medida global de qualidade:
    - Função desvio (ou *deviance*)
    - Ferramenta gráfica

Análise de Sobrevivência -- 2017 108

## Deviance

- Utilizada para obter medida global de qualidade de ajuste do modelo:
  - √ $H_0$: o modelo se ajusta aos dados
  - √ Estatística de teste: $D = 2\,[l_{saturado} - l_{modelo}]$
    - Modelo saturado: ajusta n parâmetros para as n observações (não produz resíduos)
  - √ Parâmetros de perturbação são mantidos fixos
    - Parâmetros comuns aos dois ajustes e sem interesse no estimador
    - Ex. Parâmetro de forma no modelo de Weibull



√ Quanto maior a *deviance* pior o ajuste

√ Distribuição amostral (sob $H_0$): $D \sim \chi^2_{(n-p-1)}$

- Quanto menor a *deviance*, maior o p-valor (melhor a qualidade do ajuste do modelo)

√ Deviance – Comando no R

```
> # Deviance
>
> modhivwei.dev = sum(resid(modhiv.wei, type = "deviance")^2)
> modhivwei.dev
[1] 237.1074
> gl = 193 - 3 - 1
> 1-pchisq(modhivwei.dev, gl)
[1] 0.01004656
```

Teste de qualidade global de ajuste também rejeita $H_0$

- Teste de qualidade global de ajuste rejeita $H_0$
- Ainda assim, serão verificados riscos relativos relacionados os modelos Weibull e exponencial



- Conclusão – Modelo Weibull:
  - √ O modelo foi rejeitado

```
              Value Std. Error      z        p
(Intercept)  6.06842     0.5674 10.695 1.07e-26
idade        0.00951     0.0130  0.731 4.65e-01
sexoM       -0.23627     0.3277 -0.721 4.71e-01
tratam       1.48608     0.2273  6.538 6.25e-11
Log(scale)   0.14185     0.0862  1.647 9.97e-02

Scale= 1.15
```

  - √ Risco relativo:
    - Para cada medicamento acrescentado no tratamento, o risco diminui em 72,5%

$$\exp\{-1,486\}]^{1/1,15} = 0,275$$



- Conclusão – Modelo Exponencial
  - √ O modelo foi rejeitado

```
> summary(modhiv.tratam)
Call:
survreg(formula = Surv(tempo, status) ~ tratam, data = hiv, dist = "exponential")
            Value Std. Error     z         p
(Intercept)  6.23      0.173 36.02 3.41e-284
tratam       1.41      0.187  7.57  3.88e-14
Scale fixed at 1

Exponential distribution
Loglik(model)= -744   Loglik(intercept only)= -774.6
        Chisq= 61.32 on 1 degrees of freedom, p= 4.9e-15
```

  - √ Risco relativo:
    - Para cada medicamento acrescentado no tratamento o risco diminui em 75,6%

$$\exp\{-1,41\} = 0,244$$

- Critério AIC
  - √ Modelos exponencial (tratam)
  - √ Modelo Weibull (idade + sexo + tratam)
  - √ Modelo exponencial (idade + sexo + tratam)

```
> AIC(modhiv.tratam, modhiv.wei, modhiv.exp)
             df      AIC
modhiv.tratam  2 1491.924
modhiv.wei     5 1494.049
modhiv.exp     4 1494.942
```

$$\mathrm{AIC} = 2\left[p - l(\hat{\beta}_0, \hat{\beta}_{\mathrm{tratam}})\right]$$
$$= 2[2 - (-743.9620)]$$
$$= 1.491,924$$

  - – Quanto menor o valor do Critério de Informação de Akaike (AIC), melhor o ajuste do modelo aos dados



## Análise Gráfica do Ajuste

- Ferramenta exploratória
  - √ Auxiliar na escolha de distribuições candidatas
- Gráficos:
  - √ log(Λ(t)) vs. log(t)
  - √ Comparação de curvas de sobrevivência
    - – Estimador de Kaplan-Meier vs. Estimação paramétrica



1. Gráfico de log(Λ(t)) vs. log(t)
   - √ log(Λ(t)) = log(–log(S(t)))
   - √ Avaliação suposição de proporcionalidade dos riscos
   - √ Forma d risco
     - – Weibull: esperam-se retas paralelas
     - – Exponencial: inclinação igual a 1
     - – Curvas coincidentes: parâmetro referente à covariável provavelmente não será significativo

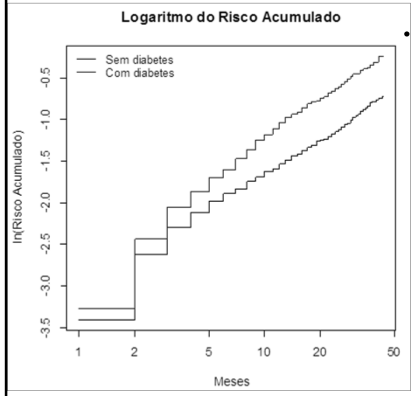


## Exemplo

- Pacientes com insuficiência renal submetidos à hemodiálise
  - √ Período: janeiro/1988 a outubro/2001
  - √ Dados: Sistema Apac
  - √ Coorte com 6.805 pacientes, no Rio de Janeiro
    - – 1.603 óbitos
  - √ Covariável:
    - – Presença de diabetes
  - √ Modelo de regressão de Weibull é candidato?

√ Gráfico do log(Risco acumulado)

– Pacientes com e sem diabetes




- Retas paralelas
  - Modelo de Weibull aparenta ser adequado



2. Comparação de curvas de sobrevivências

√ Estimativa por Kaplan-Meier vs. curvas estimadas parametricamente

√ Quanto mais próximo o modelo paramétrico estiver da curva de sobrevivência do estimador e Kaplan-Meier, melhor o modelo



**Exemplo – Continuação**

- Pacientes com insuficiência renal submetidos à hemodiálise

√ Covariável:

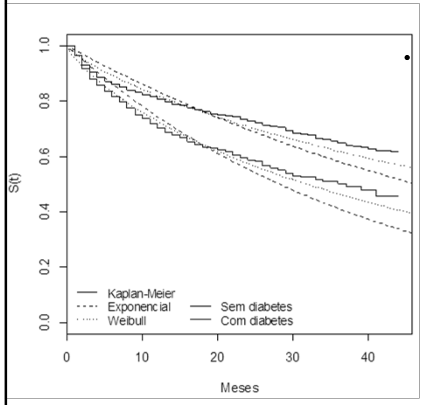
– Presença ou ausência de diabetes

√ Objetivo:

– Comparação de curvas de sobrevivência:
- Estimativa não paramétrica de Kaplan-Meier
- Estimativa do modelo de regressão exponencial
- Estimativa do modelo de regressão de Weibull



√ Comparação de curvas de sobrevivência




- Modelo que mais se aproxima da estimação não paramétrica
  - Modelo de regressão de Weibull

## Comentários

√ Análises gráficas para avaliação do ajuste do modelo são mais indicadas do que testes estatísticos formais
  – Tendem a ter baixo poder para amostras pequenas e rejeitar o modelo para amostras grandes

√ Métodos gráficos devem ser usados para descartar modelos claramente inapropriados
  – E não para demonstrar que um modelo paramétrico é melhor

  (nesse caso o melhor é análise de resíduos)



√ É comum diferentes modelos paramétricos terem ajustes razoáveis para um mesmo conjunto de dados
  – Em geral, em intervalos com concentração de observações, os modelos podem apresentar quantidades de interesse com resultados semelhantes



## Análise de Resíduos

- Definição de uma medida de resíduo não é clara no contexto de sobreviência
- Resíduos para modelos paramétricos:
  √ Resíduos de perturbação para modelos tempos de vida acelerados
- Resíduos para modelos de Cox
  √ Resíduo de Schoenfeld
    – Avaliação do pressuposto de proporcionalidade



- Resíduos para modelos paramétricos e de Cox:
  √ Resíduo de Cox-Snell
  √ Resíduo martingale
  √ Resíduo *deviance*
  √ Resíduo de escore

- Resíduos de perturbação para modelos tempos de vida acelerados
  - √ Avaliam o impacto da retirada de uma observação no ajuste global do modelo
  - √ Tipos:
    - Mede mudança no vetor estimado de parâmetros
    - Mede mudança na resposta predita (em desvios padrão)
    - Avalia efeito sobre parâmetro de forma



- Objetivo
  - √ Identificar pontos influentes do modelo
  - √ Verificar seu impacto sobre
    - Conjunto de parâmetros de regressão: **ldcase**
    - Valores preditos: **ldresp**
    - Parâmetro de forma: **ldshape**
- Comandos em R:
  - √ Para identificar indivíduos muito afastados
    - Usar função **identify**



## Exemplo – Continuação

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 193 dias
    - Óbitos: 90
    - Censuras: 103
  - √ Objetivo:
    - Determinar os resíduos do modelo paramétrico de Weibull
    - Identificar e analisar possíveis pontos influentes

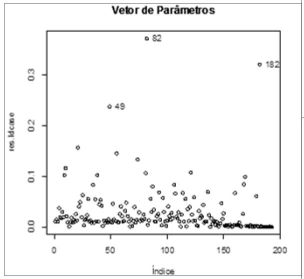


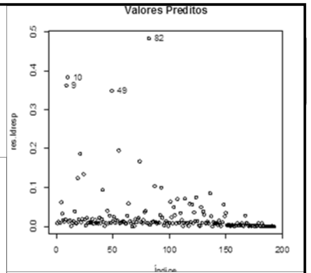
- Modelo de regressão paramétrico – Weibull:

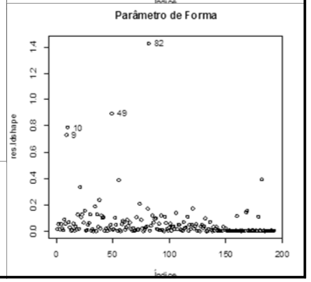
```
> modhiv.wei <- survreg(Surv(tempo, status)~idade + sexo + tratam
+     data = hiv, dist='weibull')
> summary(modhiv.wei)

Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "weibull")
               Value Std. Error      z        p
(Intercept)  6.06842     0.5674 10.695 1.07e-26
idade        0.00951     0.0130  0.731 4.65e-01
sexoM       -0.23627     0.3277 -0.721 4.71e-01
tratam       1.48608     0.2273  6.538 6.25e-11
Log(scale)   0.14185     0.0862  1.647 9.97e-02

Scale= 1.15

Weibull distribution
Loglik(model)= -742   Loglik(intercept only)= -770.3
        Chisq= 56.64 on 3 degrees of freedom, p= 3.1e-12
Number of Newton-Raphson Iterations: 5
n= 193
```

√ Resíduos de perturbação de verossimilhança

– Coorte de Aids – modelo de regressão de Weibull




– Casos identificados:

- 9, 10, 49, 82, 182



- Identificação dos resíduos

```
> # Análise de Resíduos de Perturbação - Modelo de Weibull
>
> res.ldcase <- residuals(modhiv.wei, type = "ldcase")
> res.x = 1:length(res.ldcase)
> plot(res.ldcase, xlab = "Índice")
> title("Vetor de Parâmetros")
> idx.ldcase<-identify(x = res.x, y = res.ldcase, n = 5 )
>
> res.ldresp <- residuals(modhiv.wei, type = "ldresp")
> plot(res.ldresp, xlab = "Índice")
> title("Valores Preditos")
> idx.ldresp<-identify(x = res.x, y = res.ldresp, n = 5)
>
> res.ldshape <- residuals(modhiv.wei, type = "ldshape")
> plot(res.ldshape, xlab = "Índice")
> title("Parâmetro de Forma")
> idx.ldshape<-identify(x = res.x, y = res.ldshape, n = 5)
>
> idx<-as.numeric(levels(as.factor(c(idx.ldcase, idx.ldresp, idx.ldshape))))
> idx
[1]   9  10  49  82 182
```



- Casos identificados nos gráficos:

```
> idx
[1]   9  10  49  82 182
> hiv[idx,c("tempo", "status", "sexo", "idade", "tratam")]
    tempo status sexo idade tratam
9    1563      1    M    44      0
10   1247      1    M    23      0
49   1344      0    M    30      0
82   1272      0    M    22      0
182    16      1    M    42      3
```

√ Casos 9 e 10:

– Sobreviveram um longo tempo

√ Casos 49 e 82:

– Censurados sem receber antirretroviral

√ Caso 182:

– Morreu rapidamente e recebeu 3 medicamentos



- Modelo de regressão de Weibull:

```
> modhiv.wei <- survreg(Surv(tempo, status)~idade + sexo + tratam
+     data = hiv, dist='weibull')
> summary(modhiv.wei)

Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, dist = "weibull")
               Value Std. Error      z        p
(Intercept)  6.06842     0.5674 10.695 1.07e-26
idade        0.00951     0.0130  0.731 4.65e-01
sexoM       -0.23627     0.3277 -0.721 4.71e-01
tratam       1.48608     0.2273  6.538 6.25e-11
Log(scale)   0.14185     0.0862  1.647 9.97e-02

Scale= 1.15

Weibull distribution
Loglik(model)= -742   Loglik(intercept only)= -770.3
        Chisq= 56.64 on 3 degrees of freedom, p= 3.1e-12
Number of Newton-Raphson Iterations: 5
n= 193
```

- Modelo de regressão de Weibull
  - √ Sem o caso 82

```
> summary(survreg( formula = Surv(tempo, status) ~ idade + sexo + tratam,
+    data = hiv, dist = "weibull", subset = -82))
Call:
survreg(formula = Surv(tempo, status) ~ idade + sexo + tratam,
    data = hiv, subset = -82, dist = "weibull")
              Value Std. Error      z        p
(Intercept)  5.7996     0.5760 10.069 7.60e-24
idade        0.0151     0.0133  1.137 2.55e-01
sexoM       -0.2603     0.3231 -0.806 4.20e-01
tratam       1.5490     0.2266  6.836 8.16e-12
Log(scale)   0.1281     0.0857  1.496 1.35e-01
Scale= 1.14
Weibull distribution
Loglik(model)= -739.2   Loglik(intercept only)= -769.7
        Chisq= 61.03 on 3 degrees of freedom, p= 3.5e-13
Number of Newton-Raphson Iterations: 5
n= 192
```



√ Modelo de Weibull – Comparação

| | **Completo** | **Sem indivíduo 82** |
|---|---|---|
| Intercept | 6,0684 (0,5674) | 5,7996 (0,5760) |
| Idade | 0,0095 (0,0130) | 0,0151 (0,0133) |
| sexoM | –0,2363 (0,3277) | –0,2603 (0,3231) |
| tratam | 1,4861 (0,2273) | 1,5490 (0,2266) |
| Scale | 1,1524 | 1,1367 |
| Shape | 0,868 | 0,880 |
| Loglik(model) | –742,02 | –739,21 |

√ Alteração estimativas dos parâmetros (**`ldcase`**)

√ Aumento da log-verossimilhança (**`ldresp`**)

√ Mudança estimativa do parâmetro de forma (**`ldshape`**)



## Exemplos de Aplicação

## Exemplo

- Sobrevida pacientes com leucemia aguda
  - √ Tempo de sobrevivência a diagnóstico (semanas)
  - √ 17 pacientes com leucemia aguda
  - √ Dados sem censura
  - √ Covariável
    - – wbc: contagem de glóbulos brancos na data do diagnóstico
    - – lwbc: $\log_{10}$(WBC)

√ Pacientes com leucemia – Conjunto de dados

```
> temp<-c(65,156,100,134,16,108,121,4,39,143,56,26,22,1,1,5,65)
> cens<-rep(1,17)
> lwbc<-c(3.36,2.88,3.63,3.41,3.78,4.02,4.00,4.23,3.73,3.85,3.97,
+          4.51,4.54,5.00,5.00,4.72,5.00)
> dados<-cbind(temp,cens,lwbc)
> dados<-as.data.frame(dados)
> dados
  temp cens lwbc
1    65    1 3.36
2   156    1 2.88
3   100    1 3.63
4   134    1 3.41
5    16    1 3.78
6   108    1 4.02
7   121    1 4.00
8     4    1 4.23
9    39    1 3.73
10  143    1 3.85
11   56    1 3.97
12   26    1 4.51
13   22    1 4.54
14    1    1 5.00
15    1    1 5.00
16    5    1 4.72
17   65    1 5.00
```

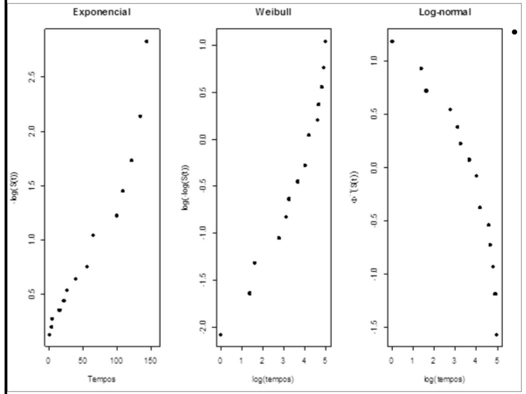
Dados sem censura

– Inviável estratificação pela covariável lwbc



- Modelo de regressão para análise dos dados

√ Gráficos de linearização - sem a covariável: lwbc




- Candidatas:
  - Exponencial
  - Weibull



√ Modelo de regressão linear

– Covariável: $X_1$ = log(WBC)

– Exponencial e Weibull

```
> leuk.exp<-survreg(Surv(temp, cens)~ lwbc,
+ data = dados, dist='exponential')
> summary(leuk.exp)
Call:
survreg(formula = Surv(temp, cens) ~ lwbc,
        data = dados, dist = "exponential")
              Value Std. Error     z        p
(Intercept)   8.48       1.711  4.95 7.27e-07
lwbc         -1.11       0.414 -2.68 7.31e-03
Scale fixed at 1
Exponential distribution
Loglik(model)= -83.9   Loglik(intercept only)= -87
        Chisq= 6.83 on 1 degrees of freedom, p= 0.
Number of Newton-Raphson Iterations: 5
n= 17
> leuk.exp$loglik
[1] -87.28983 -83.87705
```

$$\hat{\beta}_0 = -8,4775$$
$$\hat{\beta}_{\text{lwbc}} = 1,1093$$

```
> leuk.wei<-survreg(Surv(temp, cens)~ lwbc,
+ data = dados, dist='weibull')
> summary(leuk.wei)
Call:
survreg(formula = Surv(temp, cens) ~ lwbc,
            data = dados, dist = "weibull")
               Value Std. Error      z        p
(Intercept)   8.4408      1.709  4.940 7.81e-07
lwbc         -1.0982      0.418 -2.630 8.53e-03
Log(scale)   -0.0216      0.202 -0.107 9.15e-01
Scale= 0.979
Weibull distribution
Loglik(model)= -83.9   Loglik(intercept only)= -87.1
        Chisq= 6.48 on 1 degrees of freedom, p= 0.011
Number of Newton-Raphson Iterations: 6
n= 17
> leuk.wei$loglik
[1] -87.10948 -83.87136
```

Log scale não significativo

$$\hat{\beta}_0 = -8,4408$$
$$\hat{\beta}_{\text{lwbc}} = 1,0982$$
$$\hat{\gamma} = \frac{1}{\text{scale}} = \frac{1}{0,979} = 1,0218$$



- Estimativas dos modelos

| Regressão | |
|---|---|
| Exponencial | Weibull |
| $\beta_0$ = –8,4775 | $\beta_0$ = –8,4408 |
| $\beta_1$ = 1,1093 | $\beta_1$ = 1,0982 |
| $\gamma$ = 1 (fixo) | $\gamma$ = 1,0218 |

√ Estimativa de $\gamma$ muito próxima de 1

– (log scale é não significativo)

– Indicativo de que o modelo é exponencial ($\gamma$=1)

√ Teste da razão da verossimilhança

– Weibull e exponencial são aninhados

$\text{RV} = 2(83,8771 - 83,8714) = 0,0113$ $p = \text{P}\{\chi_1^2 > 0,0113\} = 0,915$

– Resultado fornece indicações favoráveis ao modelo de regressão exponencial

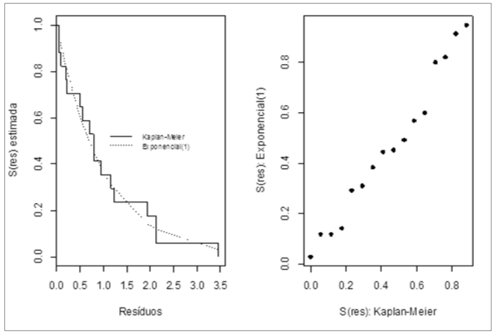
- Avaliação do ajuste
  - √ Resíduos de Cox-Snell – modelo exponencial

$$\hat{e}_i = \hat{\Lambda}(t_i|x_1) = -\ln[S(t_i|x_1)] = \boxed{t_i \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1\}}$$

  - √ Se o modelo for adequado
    - Resíduos seguem distribuição exponencial padrão
    - Curvas de sobrevivência dos resíduos, por Kaplan-Meier e exponencial padrão devem estar próximas
    - Gráfico dos pontos ($S_{KM}$(res), $S_{exp}$(res)) devem se aproximar de uma reta



√ Modelo Exponencial – Análise gráfica dos resíduos Cox-Snell




√ Modelo adequado
- Exponencial padrão parece aceitável



- Teste da razão de verossimilhanças
  - √ $H_0$: $\beta_1 = 0$
  - √ Estatística de teste

$$\boxed{\mathrm{RV} = 2(87.2898 - 83.8771) = 6,825}$$

  - √ Probabilidade de significância

$$\boxed{p = \mathrm{P}\{\chi^2_1 > 6,825\} = 0.0090}$$

  - √ Rejeita-se $H_0$
    - Possível que parte da variação observada nos tempos de sobrevivência possa ser explicada pela contagem de glóbulos brancos

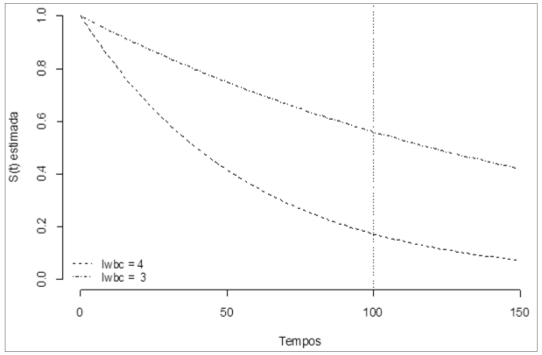


- Modelo de regressão exponencial
  - √ Função de sobrevivência estimada

$$S(t|x_1) = \exp\{-\exp\{-8,4775 + 1.1093 x_1\}\, t\},\ t \geq 0$$

  - √ Na expressão, $\beta_1$ é positivo
    - Quanto maior o valor de $x_1$, menor a probabilidade de sobrevivência estimada

√ Curvas de sobrevivência para `lwbc`




- $S(100 \mid x_1 = 4) = 17\%$ (vivos em 100 semanas)
- $S(100 \mid x_2 = 3) = 56\%$



- Comentários
  - √ Modelo de regressão exponencial ajustou-se satisfatoriamente aos dados
  - √ Conclusão:
    - Tempo estimado diminui à medida que são observadas contagens crescentes de glóbulos brancos, no diagnóstico
  - √ Probabilidade de sobrevivência condicionada ao valor de $x_1$:

$$S(t|x_1) = \exp\{-\exp\{-8,4775 + 1.1093x_1\}\, t\},\ t \geq 0$$



## Exemplo

- Sobrevida pacientes com leucemia aguda
  - √ Sobrevivência a diagnóstico (semanas)
    - 17 pacientes expressando antígeno Calla (Ag+)
    - 16 paciente não expressando antígeno Calla (Ag–)
  - √ Dados sem censura
  - √ Covariáveis
    - $X_1$: lwbc: $\log_{10}$(WBC)
    - $X_2$: grupo (0, se grupo Ag+; 1, se grupo Ag–).
  - √ Escolhido modelo de regressão exponencial para o grupo Ag+.
    - Investigar se o mesmo se aplica ao grupo Ag–



√ Pacientes com leucemia – Conjunto de dados

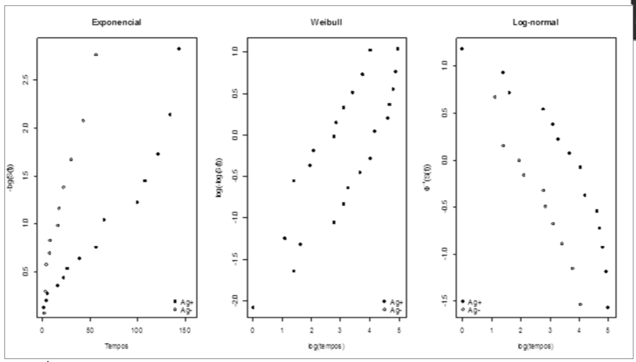
```
> tempo <- c(65,156,100,134,16,108,121,4,39,143,56,26,22,1,1,5,65,
+            56,65,17,7,16,22,3,4,2,3,8,4,3,30,4,43)
> status <- c(rep(1,17),rep(1,16))
> lwbc <-
c(3.36,2.88,3.63,3.41,3.78,4.02,4,4.23,3.73,3.85,3.97,4.51,4.54,5,5,4.72,
+
5,3.64,3.48,3.6,3.18,3.95,3.72,4,4.28,4.43,4.45,4.49,4.41,4.32,4.90,5,5)
>  grupo <- c(rep(0,17),rep(1,16))
> dados<-cbind(tempo, status, lwbc, grupo)
> dados<-as.data.frame(dados)
> dados
```

Dados sem censura

| | tempo | status | lwbc | grupo |
|---|---|---|---|---|
| 1 | 65 | 1 | 3.36 | 0 |
| 2 | 156 | 1 | 2.88 | 0 |
| 3 | 100 | 1 | 3.63 | 0 |
| 4 | 134 | 1 | 3.41 | 0 |
| 5 | 16 | 1 | 3.78 | 0 |
| 6 | 108 | 1 | 4.02 | 0 |
| 7 | 121 | 1 | 4.00 | 0 |
| 8 | 4 | 1 | 4.23 | 0 |
| 9 | 39 | 1 | 3.73 | 0 |
| 10 | 143 | 1 | 3.85 | 0 |
| 11 | 56 | 1 | 3.97 | 0 |
| 12 | 26 | 1 | 4.51 | 0 |
| 13 | 22 | 1 | 4.54 | 0 |
| 14 | 1 | 1 | 5.00 | 0 |
| 15 | 1 | 1 | 5.00 | 0 |
| 16 | 5 | 1 | 4.72 | 0 |
| 17 | 65 | 1 | 5.00 | 0 |

| | tempo | status | lwbc | grupo |
|---|---|---|---|---|
| 18 | 56 | 1 | 3.64 | 1 |
| 19 | 65 | 1 | 3.48 | 1 |
| 20 | 17 | 1 | 3.60 | 1 |
| 21 | 7 | 1 | 3.18 | 1 |
| 22 | 16 | 1 | 3.95 | 1 |
| 23 | 22 | 1 | 3.72 | 1 |
| 24 | 3 | 1 | 4.00 | 1 |
| 25 | 4 | 1 | 4.28 | 1 |
| 26 | 2 | 1 | 4.43 | 1 |
| 27 | 3 | 1 | 4.45 | 1 |
| 28 | 8 | 1 | 4.49 | 1 |
| 29 | 4 | 1 | 4.41 | 1 |
| 30 | 3 | 1 | 4.32 | 1 |
| 31 | 30 | 1 | 4.90 | 1 |
| 32 | 4 | 1 | 5.00 | 1 |
| 33 | 43 | 1 | 5.00 | 1 |

√ Gráficos das linearizações – Grupos Ag+ e Ag–




√ Indicações favoráveis ao modelo de regressão exponencial também para o grupo Ag–



- Modelos de regressão exponencial
  - √ Modelo 1:
    - Nenhuma covariável incluída
  - √ Modelo 2:
    - Incluída apenas a covariável $X_1$: lwbc
  - √ Modelo 3:
    - Incluída apenas a covariável $X_2$: grupos
  - √ Modelo 4:
    - Incluída as covariáveis $X_1$ e $X_2$
  - √ Modelo 5:
    - Incluídas variáveis X1, X2 e X1*X2



```
> leuk.mod1 <-survreg(Surv(temp, status)~ 1,
           data = dados, dist = 'exponential')
> leuk.mod1
Call:
survreg(formula = Surv(temp, status) ~ 1,
           data = dados, dist = "exponential")
Coefficients:
(Intercept)
   3.710611
Scale fixed at 1
Loglik(model)= -155.5   Loglik(intercept only)= -155.5
n= 33
```

Modelo 1:
nenhuma covariável

$$\hat{\beta}_0 = -3,7106$$
$$l(\hat{\boldsymbol{\beta}}) = -155,5$$

```
> leuk.mod2 <- survreg(Surv(temp, status)~ lwbc,
                 data = dados, dist = 'exponential')
> leuk.mod2
Call:
survreg(formula = Surv(temp, status) ~ lwbc,
        data = dados, dist = "exponential")
Coefficients:
(Intercept)        lwbc
  7.3712449  -0.9229052
Scale fixed at 1
Loglik(model)= -150.3   Loglik(intercept only)= -155.5
        Chisq= 10.31 on 1 degrees of freedom, p= 0.0013
n= 33
```

Modelo 2:
$X_1$

$$\hat{\beta}_0 = -7,3712$$
$$\hat{\beta}_1 = 0,9229$$
$$l(\hat{\boldsymbol{\beta}}) = -150,3$$



```
> leuk.mod3 <- survreg(Surv(temp, status)~ grupo,
                 data = dados, dist = 'exponential')
> leuk.mod3
Call:
survreg(formula = Surv(temp, status) ~ grupo,
        data = dados, dist = "exponential")
Coefficients:
(Intercept)       grupo
  4.134696   -1.247802
Scale fixed at 1
Loglik(model)= -149.5   Loglik(intercept only)= -155.5
        Chisq= 11.94 on 1 degrees of freedom, p= 0.00055
n= 33
```

Modelo 3:
$X_2$

$$\hat{\beta}_0 = -4,1347$$
$$\hat{\beta}_2 = 1,2478$$
$$l(\hat{\boldsymbol{\beta}}) = -149,5$$

```
> leuk.mod4 <- survreg(Surv(temp, status)~ lwbc + grupo,
                 data = dados, dist = 'exponential')
> leuk.mod4
Call:
survreg(formula = Surv(temp, status) ~ lwbc + grupo,
                 data = dados, dist = "exponential")
Coefficients:
(Intercept)        lwbc       grupo
  6.8295300  -0.7000326  -1.0180633
Scale fixed at 1
Loglik(model)= -146.5   Loglik(intercept only)= -155.5
        Chisq= 17.81 on 2 degrees of freedom, p= 0.00014
n= 33
```

Modelo 4:
$X_1$ e $X_2$

$$\hat{\beta}_0 = -6,8295$$
$$\hat{\beta}_1 = 0,7000$$
$$\hat{\beta}_2 = 1,0181$$
$$l(\hat{\boldsymbol{\beta}}) = -146,5$$

```
> leuk.mod5 <- survreg(Surv(temp, status)~ lwbc + grupo + lwbc*grupo,
                       data = dados, dist = 'exponential')
> leuk.mod5
Call:
survreg(formula = Surv(temp, status)~lwbc + grupo + lwbc*grupo,
                       data = dados, dist = "exponential")
Coefficients:
(Intercept)        lwbc       grupo  lwbc:grupo
   8.477498   -1.109298   -4.138509    0.755652
Scale fixed at 1
Loglik(model)= -145.7   Loglik(intercept only)= -155.5
        Chisq= 19.58 on 3 degrees of freedom, p= 0.00021
n= 33
```

Modelo 5:
$X_1$ , $X_2$ e $X_1$ *$X_2$

$\hat{\beta}_0 = -8,4775$
$\hat{\beta}_1 = 1,1093$
$\hat{\beta}_2 = 4,1385$
$\hat{\beta}_3 = -0,7557$
$l(\hat{\boldsymbol{\beta}}) = -145,7$



√ Estimativas dos parâmetros e da log-verossimilhança

| Modelo | Covariáveis | Estimativas | Log-verossimilhança |
|---|---|---|---|
| 1 | Nenhuma | $\beta_0 = -3,7106$ | $l_1 = -155,5$ |
| 2 | $X_1$ | $\beta_0 = -7,3712$ | $l_2 = -150,3$ |
| | | $\beta_1 = 0,9229$ | |
| 3 | $X_2$ | $\beta_0 = -4,1347$ | $l_3 = -149,5$ |
| | | $\beta_1 = 1,2478$ | |
| 4 | $X_1$ e $X_2$ | $\beta_0 = -6,8295$ | $l_4 = -146,5$ |
| | | $\beta_1 = 0,7000$ | |
| | | $\beta_2 = 1,0181$ | |
| 5 | $X_1$, $X_2$ e $X_1$*$X_2$ | $\beta_0 = -8,4775$ | $l_5 = -145,7$ |
| | | $\beta_1 = 1,1093$ | |
| | | $\beta_2 = 4,1385$ | |
| | | $\beta_3 = -0,7557$ | |



√ Testes da razão de verossimilhanças:

| Efeito | $H_0$ | RV | GL | p-valor |
|---|---|---|---|---|
| Interação: $X_1$*$X_2$ | $\beta_3 = 0$ | 2(146,5–145,7)=1,6 | 1 | 0,2059 |
| de $X_2$ \| $X_1$ | $\beta_2 = 0$ | 2(150,3–146,5)=7,6 | 1 | 0,0058 |
| de $X_1$ \| $X_2$ | $\beta_1 = 0$ | 2(149,5–146,5)=8,0 | 1 | 0,0047 |

√ Não há evidências estatísticas de que a interação entre $X_1$ e $X_2$ seja significativa

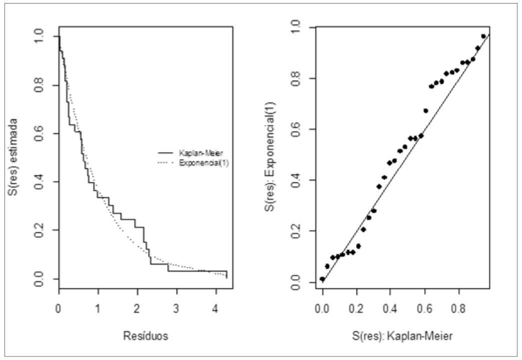
√ Evidências da significância do efeito de $X_1$
(p = 0,0047)

√ Evidências da significância do efeito de $X_2$
(p = 0,0058)



√ Análise gráfica dos resíduos Cox-Snell:

– Modelo exponencial




– Modelo exponencial apresenta ajuste razoável

√ Resultados do modelo de regressão exponencial final

– Modelo 4

```
> summary(leuk.mod4)
Call:
survreg(formula = Surv(temp, status) ~ lwbc + grupo, data = dados,
    dist = "exponential")
              Value Std. Error     z        p
(Intercept)  6.83      1.158  5.90 3.73e-09
lwbc        -0.70      0.286 -2.45 1.44e-02
grupo       -1.02      0.364 -2.80 5.12e-03
Scale fixed at 1
Exponential distribution
Loglik(model)= -146.5   Loglik(intercept only)= -155.5
        Chisq= 17.81 on 2 degrees of freedom, p= 0.00014
Number of Newton-Raphson Iterations: 5
n= 33
```

– Rejeita-se $H_0$: $\beta_1 = \beta_2 = 0$ (p-valor = 0,00014)

– Possível que parte da variação observada na resposta possa ser explicada pela contagem de glóbulos brancos e pelo grupo de antígeno



• Modelo exponencial – Modelo 4

√ Função de sobrevivência estimada

$$S(t|x_1, x_2) = \exp\{-\exp\{-6,8295 + 0,7000x_1 + 1,0181x_2\}\, t\},\ t \geq 0$$

√ $\beta_1 > 0$

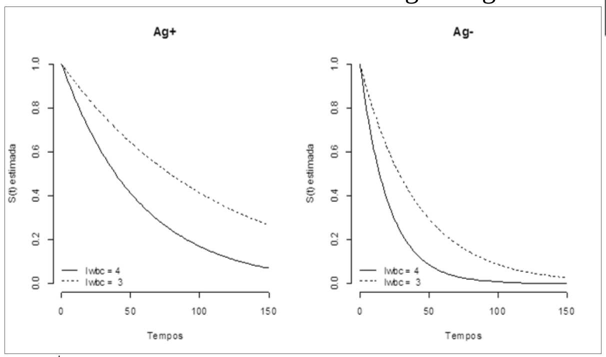
– Quanto maior o valor de $x_1$ (contagem de glóbulos brancos) menor a probabilidade de sobrevivência estimada

√ $\beta_2 > 0$

– Pacientes do grupo Ag– ($x_2$ = 1) apresentam probabilidades de sobrevivência menores do que os pacientes do grupo Ag+ ($x_2$ = 0)



√ Curvas de sobrevivências – Ag+ e Ag–

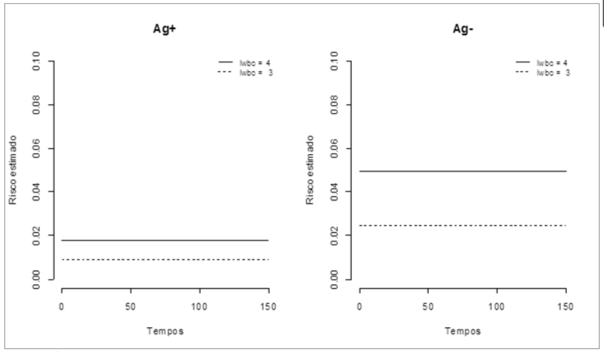



√ Pacientes Ag– com menor sobrevivência

– Menor sobrevivência para pacientes com $x_1$ maior



√ Riscos estimados pelo modelo – Ag+ e Ag–




√ Riscos constantes ao longo do tempo

– Quanto maior a contagem, maior o risco

- Conclusões:
  - √ Quanto maior a contagem de glóbulos brancos no diagnóstico, maior o risco
    - Nos 2 grupos
  - √ Pacientes que apresentam o antígeno Calla ($x_2$=0) têm melhor prognóstico do que os que ainda não experimentaram



## Exemplo

- Análise de dados de aleitamento materno
  - √ Estudo ambulatorial sobre amamentação
  - √ Amostra:
    - 150 mães de crianças com menos de 2 anos
  - √ Resposta:
    - Tempo máximo de aleitamento materno
    - Censura: crianças não acompanhadas até desmame
  - √ Covariáveis: 11
  - √ Objetivo:
    - Identificar fatores de risco ou de proteção para desmame precoce



- Descrição das covariáveis

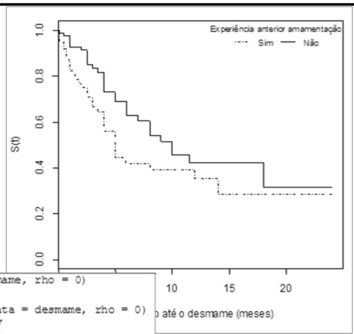
| Código | Descrição | Categorias |
|---|---|---|
| V1 | Experiência anterior amamentação | 0: sim<br>1: não |
| V2 | Número de filhos vivos | 0: ≥ 2<br>1: < 2 |
| V3 | Conceito materno sobre o tempo ideal de amamentação | 0: > 6 meses<br>1: ≤ 6 meses |
| V4 | Dificuldade para amamentar nos primeiros dias pós-parto | 0: não<br>1: sim |
| V5 | Tipo de serviço em que realizou o pré-natal | 0: público<br>1: privado/convênios |
| V6 | Recebeu exclusivamente leite materno na maternidade | 0: sim<br>1: não |
| V7 | A criança teve contato com o pai | 0: sim<br>1: não |
| V8 | Renda per capita (em SM/mês) | 0: ≥ 1 SM<br>1: < 1 SM |
| V9 | Peso ao nascimento | 0: ≥ 2,5 kg<br>1: < 2,5 kg |
| V10 | Tempo de separação mãe-filho pós-parto | 0: ≤ 6 horas<br>1: > 6 horas |
| V11 | Permanência no berçário | 0: não<br>1: sim |

  - √ Possível comparar categorias por K-M
    - Todas as variáveis são dicotômicas



- Análise Exploratória
  - √ Construir estimativas Kaplan-Meier para comparar as duas categorias
  - √ Conduzir testes logrank e Peto para testar
    - $H_0$: $S_1(t) = S_2(t)$
  - √ Todas as covariáveis com p-valor < 0,25 devem ser incluídas na etapa de modelagem estatística
    - Nível de significância modesto recomendado por:
      - Bendel e Afifi (1977) – para regressão linear
      - Constanza e Afifi (1979) – para análise discriminante
      - Mickey e Greenland (1989) – para mudança nos coeficientes de modelo de regressão logística

√ V1: Experiência anterior de amamentação



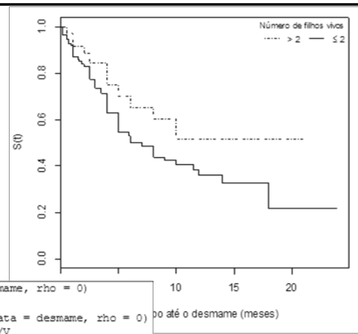

```
> survdiff(Surv(tempo, cens)~ V1, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V1, data = desmame, rho = 0)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V1=0 87       32     39.6      1.46      3.95
V1=1 63       33     25.4      2.28      3.95
 Chisq= 3.9  on 1 degrees of freedom, p= 0.0469
```

```
> survdiff(Surv(tempo, cens)~ V1, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V1, data = desmame, rho = 1)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V1=0 87     22.5     29.6      1.69      5.67
V1=1 63     26.3     19.2      2.61      5.67
 Chisq= 5.7  on 1 degrees of freedom, p= 0.0173
```

166

Análise de Sobrevivência -- 2017

√ V2: Número filhos vivos



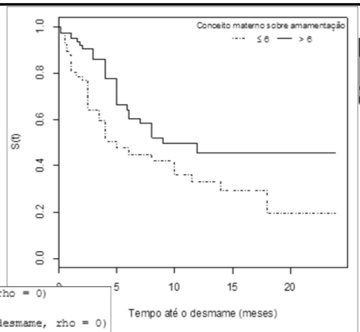

```
> survdiff(Surv(tempo, cens)~ V2, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V2, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V2=0 113       54     48.5     0.628       2.6
V2=1  37       11     16.5     1.843       2.6
 Chisq= 2.6  on 1 degrees of freedom, p= 0.107
```

```
> survdiff(Surv(tempo, cens)~ V2, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V2, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V2=0 113    40.58     36.5     0.457      2.39
V2=1  37     8.24     12.3     1.353      2.39
 Chisq= 2.4  on 1 degrees of freedom, p= 0.122
```

167

Análise de Sobrevivência -- 2017

√ V3:Tempo ideal amamentação



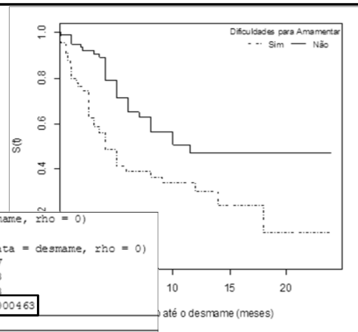

```
> survdiff(Surv(tempo, cens)~ V3, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V3, data = desmame, rho = 0)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V3=0 80       30     39.5      2.27      6.15
V3=1 70       35     25.5      3.51      6.15
 Chisq= 6.1  on 1 degrees of freedom, p= 0.0132
```

```
> survdiff(Surv(tempo, cens)~ V3, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V3, data = desmame, rho = 1)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V3=0 80     21.2     29.6      2.36      7.93
V3=1 70     27.6     19.3      3.62      7.93
 Chisq= 7.9  on 1 degrees of freedom, p= 0.00485
```

Análise de Sobrevivência -- 2017

168

√ V4: Apresentaou dificuldades amamentar




```
> survdiff(Surv(tempo, cens)~ V4, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V4, data = desmame, rho = 0)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V4=0 84       28     41.2      4.24      12.3
V4=1 66       37     23.8      7.36      12.3
 Chisq= 12.3  on 1 degrees of freedom, p= 0.000463
```

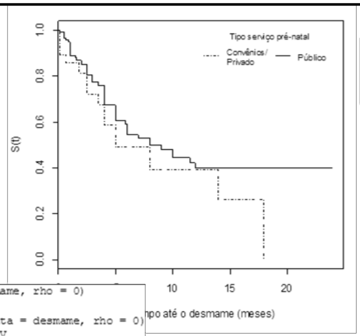
```
> survdiff(Surv(tempo, cens)~ V4, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V4, data = desmame, rho = 1)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V4=0 84     19.3     30.7      4.19      14.9
V4=1 66     29.5     18.1      7.09      14.9
 Chisq= 14.9  on 1 degrees of freedom, p= 0.000114
```

- Mães que não tiveram dificuldade para amamentar apresentam maior tempo até desmame

169

Análise de Sobrevivência -- 2017

√ V5: Tipo de serviço pré-natal

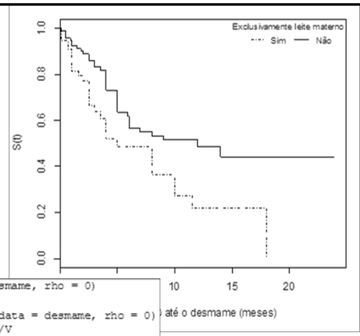



```
> survdiff(Surv(tempo, cens)~ V5, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V5, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V5=0 122       49     52.6     0.249      1.38
V5=1  28       16     12.4     1.059      1.38
 Chisq= 1.4  on 1 degrees of freedom, p= 0.241
```

```
> survdiff(Surv(tempo, cens)~ V5, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V5, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V5=0 122     36.9     39.6     0.182      1.29
V5=1  28     11.9      9.2     0.784      1.29
 Chisq= 1.3  on 1 degrees of freedom, p= 0.257
```

Análise de Sobrevivência -- 2017 170

√ V6: Recebeu somente leite materno

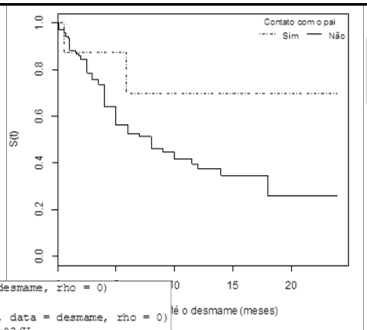



```
> survdiff(Surv(tempo, cens)~ V6, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V6, data = desmame, rho = 0)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V6=0 94       36     45.8      2.08      7.47
V6=1 56       29     19.2      4.96      7.47
 Chisq= 7.5  on 1 degrees of freedom, p= 0.00629
```

```
> survdiff(Surv(tempo, cens)~ V6, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V6, data = desmame, rho = 1)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V6=0 94     26.4     33.9      1.64      7.02
V6=1 56     22.4     15.0      3.72      7.02
 Chisq= 7  on 1 degrees of freedom, p= 0.00806
```

Análise de Sobrevivência -- 2017 171

√ V7: Contato com o pai

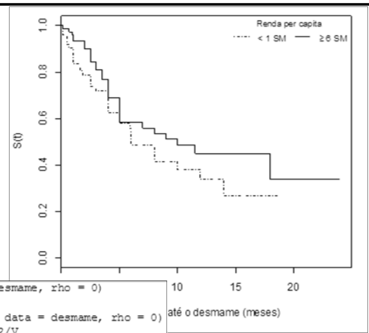



```
> survdiff(Surv(tempo, cens)~ V7, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V7, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V7=0 141       63    60.23     0.128      1.84
V7=1   9        2     4.77     1.611      1.84
 Chisq= 1.8  on 1 degrees of freedom, p= 0.175
```

```
> survdiff(Surv(tempo, cens)~ V7, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V7, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V7=0 141    47.27    45.46    0.0715      1.45
V7=1   9     1.55     3.36    0.9681      1.45
 Chisq= 1.5  on 1 degrees of freedom, p= 0.228
```

Análise de Sobrevivência -- 2017 172

√ V8: Renda per capita

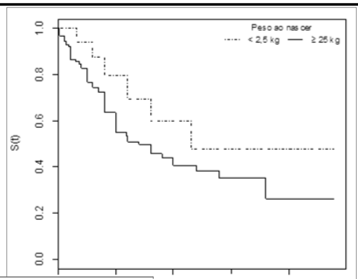



```
> survdiff(Surv(tempo, cens)~ V8, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V8, data = desmame, rho = 0)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V8=0 77       29     34.7     0.932      2.11
V8=1 73       36     30.3     1.066      2.11
 Chisq= 2.1  on 1 degrees of freedom, p= 0.146
```

```
> survdiff(Surv(tempo, cens)~ V8, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V8, data = desmame, rho = 1)
       N Observed Expected (O-E)^2/E (O-E)^2/V
V8=0 77     21.2     25.7     0.796      2.22
V8=1 73     27.6     23.1     0.884      2.22
 Chisq= 2.2  on 1 degrees of freedom, p= 0.136
```

Análise de Sobrevivência -- 2017 173

√ V9:Peso ao nascer



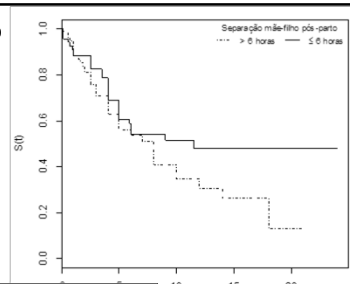

```
> survdiff(Surv(tempo, cens)~ V9, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V9, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V9=0 131       59    55.14     0.271      1.87
V9=1  19        6     9.86     1.513      1.87
 Chisq= 1.9  on 1 degrees of freedom, p= 0.171
```

```
> survdiff(Surv(tempo, cens)~ V9, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V9, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V9=0 131    44.88    41.57     0.264      2.38
V9=1  19     3.93     7.25     1.515      2.38
 Chisq= 2.4  on 1 degrees of freedom, p= 0.123
```

Análise de Sobrevivência -- 2017

174

√ V10: Tempo de separação mãe-filho



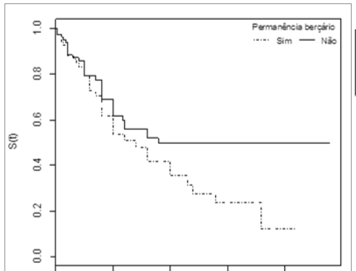

```
> survdiff(Surv(tempo, cens)~ V10, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V10, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V10=0  69       27     33.3      1.20       2.6
V10=1  81       38     31.7      1.26       2.6
 Chisq= 2.6  on 1 degrees of freedom, p= 0.107
```

```
> survdiff(Surv(tempo, cens)~ V10, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V10, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V10=0  69     20.8     24.6     0.592      1.58
V10=1  81     28.1     24.2     0.601      1.58
 Chisq= 1.6  on 1 degrees of freedom, p= 0.208
```

Análise de Sobrevivência -- 2017

175

√ V11: Permanência no berçário




```
> survdiff(Surv(tempo, cens)~ V11, data = desmame, rho = 0)
Call:
survdiff(formula = Surv(tempo, cens) ~ V11, data = desmame, rho = 0)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V11=0  81       31     37.6      1.17      2.93
V11=1  69       34     27.4      1.61      2.93
 Chisq= 2.9  on 1 degrees of freedom, p= 0.0868
```

```
> survdiff(Surv(tempo, cens)~ V11, data = desmame, rho = 1)
Call:
survdiff(formula = Surv(tempo, cens) ~ V11, data = desmame, rho = 1)
        N Observed Expected (O-E)^2/E (O-E)^2/V
V11=0  81     24.1     28.0     0.542      1.67
V11=1  69     24.8     20.9     0.726      1.67
 Chisq= 1.7  on 1 degrees of freedom, p= 0.196
```

Análise de Sobrevivência -- 2017

176

√ Tabela de valores dos testes logrank e Peto

| Variável | | | | *logrank* | | Peto | |
|---|---|---|---|---|---|---|---|
| Código | Descrição | Valor | Qte. | $\chi^2$ | p-valor | $\chi^2$ | p-valor |
| V1 | Experiência anterior amamentação | 0 | 87 | 3,9 | 0,047 | 5,7 | 0,017 |
| | | 1 | 63 | | | | |
| V2 | Número de filhos vivos | 0 | 113 | 2,6 | 0,107 | 2,4 | 0,122 |
| | | 1 | 37 | | | | |
| V3 | Conceito materno sobre o tempo ideal de amamentação | 0 | 80 | 6,1 | 0,013 | 7,9 | 0,005 |
| | | 1 | 70 | | | | |
| V4 | Dificuldade para amamentar nos primeiros dias pós-parto | 0 | 84 | 12,3 | 0,000 | 14,9 | 0,000 |
| | | 1 | 66 | | | | |
| V5 | Tipo de serviço em que realizou o pré-natal | 0 | 122 | 1,4 | 0,241 | 1,3 | 0,257 |
| | | 1 | 28 | | | | |
| V6 | Recebeu exclusivamente leite materno na matemidade | 0 | 94 | 7,5 | 0,006 | 7,0 | 0,008 |
| | | 1 | 56 | | | | |
| V7 | A criança teve contato com o pai | 0 | 141 | 1,8 | 0,175 | 1,5 | 0,228 |
| | | 1 | 9 | | | | |
| V8 | Renda per capita (em SM/mês) | 0 | 77 | 2,1 | 0,146 | 2,2 | 0,136 |
| | | 1 | 73 | | | | |
| V9 | Peso ao nascimento | 0 | 131 | 1,9 | 0,171 | 2,4 | 0,123 |
| | | 1 | 19 | | | | |
| V10 | Tempo de separação mãe-filho pós-parto | 0 | 69 | 2,6 | 0,107 | 1,6 | 0,208 |
| | | 1 | 81 | | | | |
| V11 | Permanência no berçário | 0 | 81 | 2,9 | 0,087 | 1,7 | 0,196 |
| | | 1 | 69 | | | | |

– Todos os p-valores menores que 0,25

Análise de Sobrevivência -- 2017

177

- Comentários:
  - √ 11 covariáveis potencialmente importantes
    - $2^{11}$ = 2048 modelos possíveis
  - √ É impraticável ajustar todos esses possíveis modelos para selecionar aquele que explica melhor a resposta
  - √ Há procedimentos automáticos disponíveis em pacotes estatísticos:

Análise de Sobrevivência -- 2017 178

- Procedimentos automáticos:
  - √ Tipos:
    - Forward
    - Backward
    - Stepwise
  - √ Vantagem
    - Métodos implementados e disponíveis em pacotes
  - √ Desvantagens:
    - Tendem a identificar um particular conjunto de covariáveis em vez de possíveis conjuntos igualmente bons
    - Impede que dois ou mais conjuntos igualmente bons sejam apresentados ao pesquisador para a escolha do mais relevante em sua área de aplicação

Análise de Sobrevivência -- 2017 179

- Seleção de covariáveis:
  - √ Covariáveis com significância clínica devem ser incluídas independente de significância estatística
  - √ Importância clínica deve ser considerada em cada passo de inclusão ou exclusão no processo de seleção de covariáveis
  - √ Evitar ser muito rigoroso ao testar cada nível individual de significância
    - Recomenda-se um valor próximo de 0,10

Análise de Sobrevivência -- 2017 180

- Modelagem separada de cada covariável com a resposta
  - √ Objetivo:
    - selecionar quais covariáveis prosseguirão na análise
  - √ Critério utilizado:
    - Permanecer com covariáveis que apresentarem p-valor inferior a 0,25
  - √ Todas as covariáveis passam por esse critério

Análise de Sobrevivência -- 2017 181

- Estratégia para seleção de covariáveis:

  (Colosimo e Giolo, 2006)

  1. Ajustar todos os modelos contendo uma única covariável
     – Incluir todas as covariáveis significativas ao nível de 0,10
     – Aconselhável usar o Teste da Razão de Verossimilhanças



- Em geral, ajusta-se a gama generalizada para o Teste de Razão de Verossimilhanças

  √ Aninha os modelos exponencial, de Weibull, lognormal e gama

- Ajuste do modelo gama generalizado:

```
> library(flexsurv)
> flexsurvreg(formula, data,
              dist='gengamma')
```



- Ajuste de modelos com uma única covariável

Passo 1

| | Modelo | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
|---|---|---|---|---|---|
| | Nulo | -228,226 | 456,452 | | |
| V1 | Experiência anterior amamentação | -225,579 | 451,158 | 5,294 | 0,0214 |
| V2 | Número de filhos vivos | -226,819 | 453,638 | 2,814 | 0,0935 |
| V3 | Conceito materno sobre o tempo ideal de amamentação | -225,330 | 450,659 | 5,793 | 0,0161 |
| V4 | Dificuldade para amamentar nos primeiros dias pós-parto | -221,811 | 443,621 | 12,831 | 0,0003 |
| V5 | Tipo de serviço em que realizou o pré-natal | -227,335 | 454,670 | 1,782 | 0,1819 |
| V6 | Recebeu exclusivamente leite materno na maternidade | -224,718 | 449,436 | 7,016 | 0,0081 |
| V7 | A criança teve contato com o pai | -227,101 | 454,202 | 2,250 | 0,1336 |
| V8 | Renda per capita (em SM/mês) | -226,736 | 453,473 | 2,979 | 0,0843 |
| V9 | Peso ao nascimento | -226,753 | 453,506 | 2,946 | 0,0861 |
| V10 | Tempo de separação mãe-filho pós-parto | -227,255 | 454,511 | 1,942 | 0,1635 |
| V11 | Permanência no berçário | -227,181 | 454,362 | 2,090 | 0,1483 |

Covariáveis retidas:
V1, V2, V3, V4, V6, V8, V9

Covariáveis excluídas:
V5, V7, V10, V11



2. Ajustar conjuntamente as covariáveis significativas no passo 1
   – Na presença de certas covariáveis, outras podem deixar de ser significativas
   – Ajustam-se modelos reduzidos, excluindo uma única covariável de cada vez
   – Permanecem no modelo somente aquelas que atingirem a significância

- Ajuste conjunto das covariáveis significativas
  - √ Exclusão de uma variável por vez
  - √ Permanecem apenas aquelas que atingirem significância

| Passo 2 | | | | | | |
|---|---|---|---|---|---|---|
| Modelo | | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
| V1 + V2 + V3 + V4 + V6 + V8 + V9 | | -212,475 | 424,950 | | |
| V1+V2+V3+V4+V6+V8 | | -213,129 | 426,258 | 1,307 | 0,2529 |
| V1+V2+V3+V4+V6+V9 | V8 | -214,213 | 428,426 | 3,475 | 0,0623 |
| V1+V2+V3+V4+V8+V9 | V6 | -215,196 | 430,391 | 5,441 | 0,0197 |
| V1+V2+V3+V6+V8+V9 | V4 | -216,698 | 433,397 | 8,447 | 0,0037 |
| V1+V2+V4+V6+V8+V9 | V3 | -214,155 | 428,310 | 3,360 | 0,0668 |
| V1+V3+V4+V6+V8+V9 | | -212,539 | 425,077 | 0,127 | 0,7219 |
| V2+V3+V4+V6+V8+V9 | | -213,100 | 426,199 | 1,249 | 0,2638 |

Covariáveis retidas: V3, V4, V6, V8

Covariáveis excluídas: V1, V2, V9



3. Ajustar modelo com covariáveis retidas no passo 2
   - As covariáveis excluídas no passo 2 retornam ao modelo para confirmar sua significância estatística



- Ajuste de modelo com covariáveis retidas
  - √ Retorno de covariável excluída uma a uma

| Passo 3 | | | | |
|---|---|---|---|---|
| Modelo | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
| V3+V4+V6+V8 | -212,475 | 428,397 | | |
| V3+V4+V6+V8+V1 | -213,221 | 426,442 | 1,955 | 0,1620 |
| V3+V4+V6+V8+V2 | -213,635 | 427,269 | 1,127 | 0,2884 |
| V3+V4+V6+V8+V9 | -213,647 | 427,294 | 1,103 | 0,2936 |

Covariáveis retidas: V3, V4, V6, V8

  - √ As covariáveis excluídas no passo anterior não são significativas



4. Inclusão no modelo de eventuais covariáveis significativas no passo 3
   - As covariáveis excluídas no passo 1 retornam ao modelo para confirmar sua significância estatística

- Ajuste de modelo com covariáveis retidas nos passos 2 e 3
  - √ Retorno de covariáveis excluídas no passo 1

Passo 4

| Modelo | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
|---|---|---|---|---|
| V3+V4+V6+V8 | -212,475 | 428,397 | | |
| V3+V4+V6+V8+V5 | -214,198 | 428,397 | 0,000 | 0,9989 |
| V3+V4+V6+V8+V7 | -213,319 | 426,637 | 1,759 | 0,1847 |
| V3+V4+V6+V8+V10 | -214,072 | 428,143 | 0,253 | 0,6146 |
| V3+V4+V6+V8+V11 | -214,117 | 428,234 | 0,163 | 0,6863 |

Covariáveis retidas: V3, V4, V6, V8

  - √ As covariáveis excluídas no passo 1 não são significativas



5. Inclusão no modelo de eventuais covariáveis significativas no passo 4
   – Testa-se se alguma dela pode ser retirada do modelo



- Inclusão covariáveis significativas em 4
  - √ Testa-se possibilidade da retirada de alguma do modelo

Passo 5

| Modelo | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
|---|---|---|---|---|
| V3+V4+V6+V8 | -212,475 | 428,397 | | |
| V3+V4+V6 | -216,955 | 433,911 | 5,514 | 0,0189 |
| V3+V4+V8 | -216,747 | 433,493 | 5,097 | 0,0240 |
| V3+V6+V8 | -220,753 | 441,506 | 13,109 | 0,0003 |
| V4+V6+V8 | -216,109 | 432,218 | 3,822 | 0,0506 |

Covariáveis retidas: V3, V4, V6, V8

  - √ Todas as covariáveis incluídas nos passos anteriores são significativas



6. Ajusta-se modelo final para os efeitos principais utilizando as covariáveis que não foram excluídas no passo 5
   – Verifica-se a possibilidade de inclusão de termos de interação dupla entre as covariáveis incluídas no modelo

- Modelo final para os efeitos principais
  - √ Verificação inclusão de interações duplas

Passo 6

| Modelo | $l(\theta)$ | $-2l(\theta)$ | TRV | p-valor |
|---|---|---|---|---|
| V3+V4+V6+V8 | -212,475 | 428,397 | | |
| V3+V4+V6+V8+V3*V4 | -213,844 | 427,689 | 0,708 | 0,4002 |
| V3+V4+V6+V8+V3*V6 | -213,295 | 426,590 | 1,807 | 0,1789 |
| V3+V4+V6+V8+V3*V8 | -214,059 | 428,118 | 0,279 | 0,5973 |
| V3+V4+V6+V8+V4*V6 | -213,823 | 427,647 | 0,750 | 0,3865 |
| V3+V4+V6+V8+V4*V8 | -213,826 | 427,652 | 0,744 | 0,3883 |
| V3+V4+V6+V8+V6*V8 | -214,056 | 428,112 | 0,284 | 0,5939 |

Covariáveis retidas:
V3, V4, V6, V8

  - √ Nenhum termo de interação dupla foi significativo



- Modelo Final:
  - √ Efeitos principais identificados no passo 5
    - V3, V4, V6, V8
  - √ Termos de interação identificados no passo 6
    - Nenhum é significativo



- Modelo Final:

| | l(θ) | –2l(θ) | TRV | gl | p-valor |
|---|---|---|---|---|---|
| Nulo | –228,226 | 456,452 | | | |
| V3+V4+V6+V8 | –214,198 | 428,397 | 28,056 | 4 | 0,000 |
| V3+V4+V6+V1 | –215,228 | 430,456 | 25,996 | 4 | 0,000 |

  - √ Possível efeito de multicolienaridade entre:
    - V1: experiência anterior de amamentação
    - V8: renda per capita
    
    (estatísticas de teste não são muito diferentes)
  - √ Decisão sobre permanência de variável baseada em evidências clínicas



- Variáveis prognósticas:
  - √ V1:
    - Experiência anterior de amamentação
  - √ V3:
    - Conceito materno sobre o tempo ideal de amamentação
  - √ V4:
    - Dificuldades de amamentar nos primeiros dias pós-parto
  - √ V6:
    - Recebimento exclusivo de leite materno na maternidade

- Saídas dos modelos Weibull e Lognormal

```
> desm.wei <- survreg(Surv(desmame$tempo, desmame$cens)~ V3 + V4 + V6 + V1,
+ dist = 'weibull')
> desm.wei
Call:
survreg(formula = Surv(desmame$tempo, desmame$cens) ~ V3 + V4 +
    V6 + V1, dist = "weibull")
Coefficients:
(Intercept)          V3          V4          V6          V1
  3.5871202  -0.4926820  -0.6820891  -0.6521603  -0.5336577
Scale= 1.048537
Loglik(model)= -217.9   Loglik(intercept only)= -229.2
        Chisq= 22.57 on 4 degrees of freedom, p= 0.00015
n= 150
```

```
> desm.ln <- survreg(Surv(desmame$tempo, desmame$cens)~ V3 + V4 + V6 + V1,
+ dist = 'lognorm')
> desm.ln
Call:
survreg(formula = Surv(desmame$tempo, desmame$cens) ~ V3 + V4 +
    V6 + V1, dist = "lognorm")
Coefficients:
(Intercept)          V3          V4          V6          V1
  3.2925917  -0.6314176  -0.8242242  -0.6804915  -0.5721767
Scale= 1.438643
Loglik(model)= -215.3   Loglik(intercept only)= -228.9
        Chisq= 27.21 on 4 degrees of freedom, p= 1.8e-05
n= 150
```

Análise de Sobrevivência -- 2017 199

- Busca de modelo mais parcimonioso

| | l(θ) | –2l(θ) | TRV | gl | p-valor |
|---|---|---|---|---|---|
| Gama generalizada | –215,228 | 430,456 | | | |
| Weibull | –217,902 | 435,803 | 5,347 | 1 | 0,0208 |
| Lognormal | –215,337 | 430,673 | 0,218 | 1 | 0,6409 |

√ $H_0$: o modelo de interesse é adequado

√ Conclusão:

– Modelo de regressão lognormal é adequado para ajustar os tempos até o desmame

– Variáveis prognósticas: V1, V3, V4, V6

Análise de Sobrevivência -- 2017 200

- Estimativas dos parâmetros do modelo de regressão lognormal

√ Coeficientes estimados expressos na escala logarítmica dos tempos

$$Y = \ln(T) = \boldsymbol{x}'\boldsymbol{\beta} + \sigma\nu$$

```
> desm.ln <- survreg(Surv(desmame$tempo, desmame$cens)~ V3 + V4 + V6 + V1,
+ dist = 'lognorm')
> summary(desm.ln)
Call:
survreg(formula = Surv(desmame$tempo, desmame$cens) ~ V3 + V4 +
    V6 + V1, dist = "lognorm")
             Value Std. Error     z        p
(Intercept)  3.293      0.304 10.84 2.29e-27
V3          -0.631      0.290 -2.18 2.93e-02
V4          -0.824      0.302 -2.73 6.33e-03
V6          -0.680      0.293 -2.33 2.00e-02
V1          -0.572      0.301 -1.90 5.72e-02
Log(scale)   0.364      0.090  4.04 5.36e-05
Scale= 1.44
Log Normal distribution
Loglik(model)= -215.3   Loglik(intercept only)= -228.9
        Chisq= 27.21 on 4 degrees of freedom, p= 1.8e-05
Number of Newton-Raphson Iterations: 4
```

$$\begin{aligned} \mathrm{ep}(\hat{\sigma}) &= \mathrm{ep}[\ln(\hat{\sigma})]\hat{\sigma} \\ &= (0,090)(1,44) \\ &= 0,1295 \end{aligned}$$

Análise de Sobrevivência -- 2017 201

- Estimativa dos parâmetros:

| Covariável | | Estimativa | Erro Padrão | p-valor |
|---|---|---|---|---|
| | Intercepto | 3,293 | 0,304 | 0,000 |
| V1 | Experiência anterior de amamentação | –0,572 | 0,301 | 0,0572 |
| V3 | Conceito sobre o tempo de amamentação | –0,631 | 0,290 | 0,0293 |
| V4 | Dificuldades de amamentação pós-parto | –0,824 | 0,302 | 0,0063 |
| V6 | Recebimento exclusivo de leite materno | –0,680 | 0,293 | 0,0020 |
| | Parâmetro de forma | 1,439 | 0,129 | |

Análise de Sobrevivência -- 2017 202

## Adequação do Modelo Ajustado

- Métodos gráficos:
  - √ Utilização dos resíduos para confirmar a adequação do modelo lognormal
- Tipos de resíduos:
  - √ Resíduos padronizados
  - √ Resíduos de Cox-Snell



- Resíduos padronizados
  - √ Baseado na representação dos modelos log-lineares

$$\hat{\nu}_i = \frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\hat{\sigma}}, \text{ com } y_i = \ln(t_i)$$

  - √ Modelo lognormal bem ajustado para os dados:
    - Resíduos com distribuição aproximadamente normal padrão
    - Exponencial dos resíduos com distribuição lognormal



√ Procedimento:
- Estimar sobrevivência dos resíduos com Kaplan-Meier
- Comparar estimação de Kaplan-Meier com sobrevivência da lognormal padrão



- Cálculo e transformação dos resíduos:

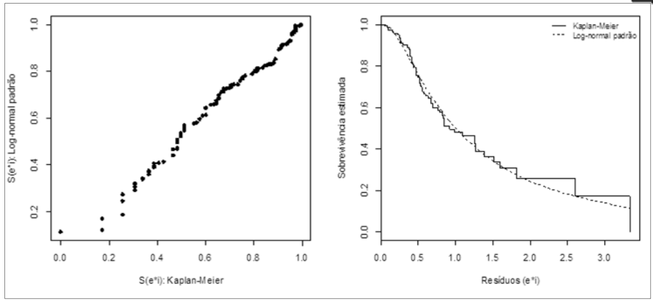
```
> # Determinacao Residuos e*i (log-normal padrão)
>
> residuos <- predict(desm.ln, type = "response")    # Escala original
> padr.log <-(log(tempo)-log(residuos))/desm.ln$scale # Padroniza na escala log
> padr.orig<-exp(padr.log)    # Escala original
> padr.km <- survfit(Surv(padr.orig, cens)~ 1)
> res <- padr.km$time
> sln <- pnorm(-log(res))
```

- Gráficos dos resíduos padronizados:

```
> # Gráficos Residuos e*i
>
> win.graph(width=12, height=6) # Default width = 7, height = 7
> par(mfrow = c(1,2))
> plot(padr.km$surv, sln, pch = 16, ylab = "S(e*i): Log-normal padrão",
+ xlab = "S(e*i): Kaplan-Meier")
> plot(padr.km, conf.int = F, mark.time = F, pch = 16,
+ xlab = "Resíduos (e*i)", ylab = "Sobrevivência estimada")
> lines(res, sln, lty = 2)
> legenda = c("Kaplan-Meier", "Log-normal padrão")
> legend(x = "topright", legend = legenda, lty = 1:2, cex = 0.8, bty = "n")
> par(mfrow = c(1,1))
```

√ Sobrevivência dos resíduos $e^*_i$:

– Estimativa Kaplan-Meier e Log-normal padrão




– Pode-se acreditar que o modelo de regressão lognormal se encontra bem ajustado aos dados sob análise



√ Resíduos de Cox-Snell

$$\boxed{\hat{e}_i = \hat{\Lambda}(t_i|x_1) = -\ln[\hat{S}(t_i|x_1)]}$$

√ Modelos:

– Exponencial: $\boxed{\hat{e}_i = \left[t_i \exp\{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\}\right]}$

– Weibull: $\boxed{\hat{e}_i = \left[t_i \exp\{\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\}\right]^{\hat{\gamma}}}$

– Lognormal: $\boxed{\hat{e}_i = -\ln\left[1 - \Phi\left(\frac{\ln(t_i) - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)\right]}$



- Se o modelo for adequado:
  - √ Resíduos seguem distribuição exponencial padrão
    - Curvas de sobrevivência dos resíduos, por Kaplan-Meier e exponencial padrão devem estar próximas
    - Gráfico dos pontos ($S_{KM}$(res), $S_{exp}$(res)) devem se aproximar de uma reta
    - Gráfico $e_i$ vs. $\Lambda(e_i)$ deve ser aproximadamente uma reta



- Cálculo dos resíduos de Cox-Snell:

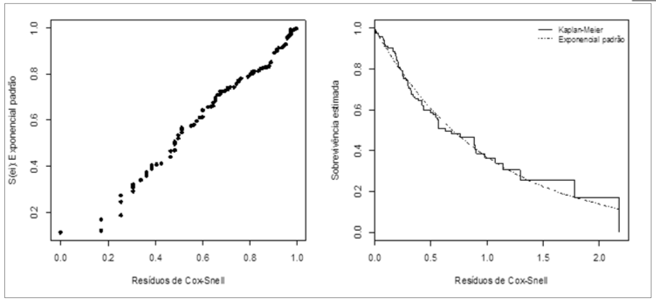
```
> # Determinação Resíduos Cox-Snell
>
> snell <- - log(1 - pnorm(padr.log))
> snell.km <- survfit(Surv(snell, cens)~ 1)
> t <- snell.km$time
> s.t <- snell.km$surv
> s.exp <- exp(-t)
```

- Gráficos dos resíduos padronizados:

```
> # Gráficos Resíduos Cox-Snell
>
> win.graph(width=12, height=6) # Default width = 7, height = 7
> par(mfrow = c(1,2))
> plot(s.t, s.exp, pch = 16, ylab = "S(ei): Exponencial padrão",
+ xlab = "Resíduos de Cox-Snell")
> plot(snell.km, conf.int = F, mark.time = F, pch = 16,
+ xlab = "Resíduos de Cox-Snell", ylab = "Sobrevivência estimada")
> lines(t, s.exp, lty = 4)
> legenda = c("Kaplan-Meier", "Exponencial padrão")
> legend(x = "topright", legend = legenda, lty = c(1, 4), cex = 0.8, bty = "n")
> par(mfrow = c(1,1))
```

√ Sobrevivência dos resíduos de Cox-Snell ($e_i$):

– Estimativa Kaplan-Meier e Exponencial padrão




– Evidências de que o modelo de regressão lognormal se encontra bem ajustado aos dados sob análise



## Interpretação dos Coeficientes

- Para variável binária

√ $\exp\{\beta_i\}$ é a razão dos tempos medianos:

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

√ Válido para os modelos Weibull e lognormal



- No exemplo

√ Efeito da experiência anterior de amamentação

0: sim
1: não

$$\boxed{e^{\hat{\beta}_1} = e^{-0,572} = 0,564}$$

– Tempo mediano até o desmame de mães que não tiveram experiência anterior de amamentação é aproximadamente a metade daquele das mães que já tiveram esta experiência

√ Efeito do conceito sobre o tempo de amamentação:

0: > 6 meses
1: ≤ 6 meses

$$\boxed{e^{\hat{\beta}_2} = e^{-0,631} = 0,532}$$

– Mães que acreditam que o tempo ideal de desmame é superior a 6 meses apresentam tempo mediano até o desmame de aproximadamente 2 vezes maior do que as mães que pensam ser este tempo igual ou inferior a 6 meses



√ Efeito da dificuldade de amamentação pós-parto

0: sim
1: não

$$\boxed{e^{\hat{\beta}_3} = e^{-0,824} = 0,439}$$

– Tempo mediano até o desmame de mães que não apresentaram dificuldades de amamentar nos primeiros dias até o parto é 2,3 vezes maior que o tempo das que sofreram este tipo de dificuldade.

√ Efeito do recebimento exclusivo de leite materno:

0: sim
1: não

$$\boxed{e^{\hat{\beta}_4} = e^{-0,680} = 0,507}$$

– Crianças que receberam exclusivamente leite materno na maternidade têm um tempo mediano de amamentação duas vezes maior do que o tempo daqueles que receberam outro tipo de alimentação juntamente com o leite materno

## Referências

## Bibliografia


- COLOSIMO, E.; GIOLO, S. *Análise de sobrevivência aplicada*. Edgard Blücher, 2006.
- CARVALHO, M. S. *et al*. *Análise de sobrevivência: teoria e aplicações em saúde*. Fiocruz, 2011.
- KLEIN, J. P.; MOESCHBERGER, M. L. *Survival analysis: techniques for censored and truncated data*. Springer, 2003.
- KLEINBAUM, D. G. KLEIN, M. *Survival analysis: a self-learning text*. Springer, 2011.




## Material de Apoio

- R:
  - √ www.r-project.org
- Tutorial online do R:
  - √ http://www.leg.ufpr.br/Rtutorial
  - √ http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa
- Conjuntos de dados e material Análise de Sobrevivência – Carvalho et al.
  - √ http://sobrevida.fiocruz.br