Análise de Sobrevivência

Camila Borelli Zeller
Lupércio França Bessegato
Departamento de Estatística/UFJF

Roteiro Geral

Análise de Sobrevivência - 2017

2

Modelos Probabilísticos

Roteiro do Módulo

3. Modelos probabilísticos:
   a) Introdução
   b) Modelos em análise de sobrevivência
   c) Estimação dos parâmetros dos modelos
   d) Intervalos de confiança e testes de hipóteses
   e) Escolha do modelo
   f) Exemplos

Análise de Sobrevivência - 2017

4

Introdução

Técnicas Não Paramétricas

- Técnicas não paramétricas:
  - √ Fáceis de entender
  - √ Não permitem inclusão de covariáveis na análise
- Estratificação:
  - √ Simplicidade dos cálculos e facilidade de entendimento
  - √ Análise com várias covariáveis gera um número muito grande de estratos que podem conter poucas observações

Modelos de Regressão

(Apropriados para dados de sobrevivência)

- Modelos Paramétricos:
  - √ São mais eficientes e menos flexíveis

$$Y = \log(T) = \beta_0 + \beta_1 + \cdots + \beta_p + \sigma W$$

    - W: erros com distribuição especificada
  - √ Classe denominada tempo de vida acelerado é muito utilizada
    - Tempo de vida obedece à seguinte relação:

$$Y = \log(T) = \mu + \sigma W$$

      - μ: parâmetro de locação
      - σ: parâmetro de escala

- Modelos Semi-paramétricos:
  - √ Modelo de riscos proporcionais de Cox
  - √ Formado por dois elementos: um paramétrico e outro não-paramétrico
  - √ Flexibilidade
  - √ Permite incorporação de covariáveis dependentes do tempo

Modelos em Análise de Sobrevivência

Distribuições

- Adotadas para se ajustar a distribuição do tempo de sobrevivência T:
  - √ Exponencial
  - √ Weibull
  - √ Log-normal
  - √ Log-logística
  - √ Etc.

Distribuição Exponencial

- Densidade: $f(t) = \alpha \exp\{-\alpha t\}$
  - √ $t \geq 0$ e $\alpha > 0$
- Função de sobrevivência: $S(t) = \exp\{-\alpha t\}$
- Parâmetro:
  - √ α: representa a queda da função de sobrevivência ao longo do tempo

- Função de risco:

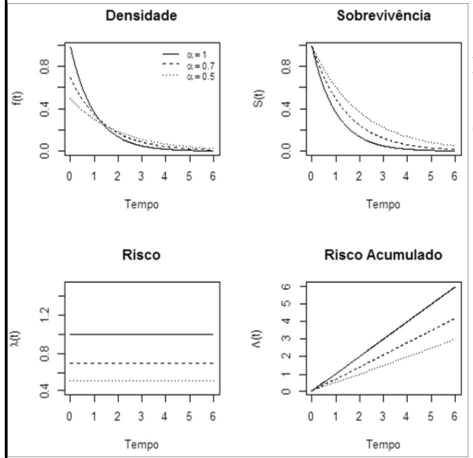
$$\lambda(t) = \frac{\alpha \exp\{-\alpha t\}}{\exp\{-\alpha t\}} = \alpha$$

  - √ Constante e igual a α
- Função de risco acumulado

$$\Lambda(t) = -\ln S(t) = \alpha t$$

  - √ É uma função linear no tempo

√ Exponencial – Forma típica das funções



Função de risco

- √ Paralela ao eixo do tempo

Quanto maior o risco α:

- √ Mais rapidamente decai a sobrevivência
- √ Maior a inclinação da reta do risco acumulado

- Exponencial – Esperança e Variância:

$$\mathrm{E}[T] = \frac{1}{\alpha}$$

$$\mathrm{Var}[T] = \frac{1}{\alpha^2}$$

- Função quantílica:

$$F(t_p) = p,\ 0 < p < 1$$

$$1 - \exp\{-\alpha t_p\} = p$$

$$t_p = -\frac{\ln(1-p)}{\alpha}$$

- Mediana:

$$t_{med} = -\frac{\ln(1-0,5)}{\alpha} = \frac{\ln(2)}{\alpha}$$

  - √ $F(t_{med}) = 0,5$

- Exponencial – Média e mediana:

$$\mathrm{E}[T] = \frac{1}{\alpha} \qquad t_{med} = \frac{\ln(2)}{\alpha}$$

  - √ Tempo de sobrevivência é assimétrico
    - – Tempo mediano é mais usual que tempo médio
  - √ $t_{med} < E[T]$, para qualquer valor de α
  - √ E[T] e t_{med} são inversamente proporcionais ao risco

Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Coorte completa entre 1986 e 2000
    - – Óbitos: 90
    - – Censuras: 103
  - √ Modelagem:
    - – Tempo entre diagnóstico de Aids e a morte segue uma distribuição exponencial
      - ▪ Estimativa: α = 0,000497

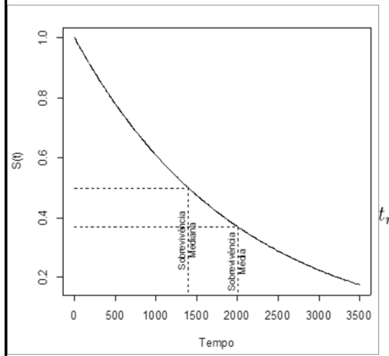
- Encerramento de acompanhamento: t = 3.500 dias
  - √ Espera-se que cerca de 20% dos pacientes ainda estejam vivos
  - √ Sobrevivência teórica adotada – Exponencial:

$$\hat{S}_e(3500) = \exp\{-(0,000497)(1500)\} = 0,17561$$

    - Sobrevivência teórica tende a zero
    - (Aconteceria se tempo de estudo se prolongasse até que todos os pacientes fossem acompanhados até o óbito)
- Função de risco é constante:

$$\lambda(t) = 0,000497$$

√ Coorte de Aids – Modelo paramétrico exponencial



Tempo médio de sobrevivência:

$$\mathrm{E}[T] = \frac{1}{\alpha} = \frac{1}{0,000497} = 2.012 \text{ dias}$$

Tempo mediano de sobrevivência:

$$t_{med} = \frac{\ln(2)}{\alpha} = \frac{\ln(2)}{0,000497} = 1.394 \text{ dias}$$

Sobrevivência ao tempo médio:

$$S(\mathrm{E}[T]) = \exp -1 = 0,3679$$

SEMPRE!!! (qualquer exponencial)

- Exponencial – Comentários:
  - √ Simplifica a modelagem
    - É expressa por um único parâmetro
  - √ Suposição de risco constante é pouco plausível na maioria dos fenômenos da saúde
    - Diz-se que a distribuição não tem memória
  - √ Pode ser uma aproximação válida quando o tempo de acompanhamento é curto o suficiente para se considerar constante o risco no período
    - Risco de acidentes domésticos em crianças de 4 a 6 anos
  - √ Utilizada em alguns estudos de tempos de vida e em remissão de doenças crônicas e infecciosas

Distribuição de Weibull

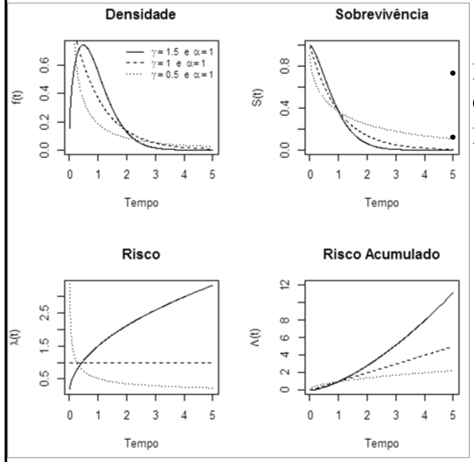
- Densidade: $f(t) = \gamma\alpha^{\gamma}t^{\gamma-1}\exp\{-(\alpha t)^{\gamma}\}$
  - √ $t \geq 0$, α e $\gamma > 0$
- Função de sobrevivência: $S(t) = \exp\{-(\alpha t)^{\gamma}\}$
- Parâmetros:
  - √ α: parâmetro de escala
    (mesma unidade de t)
  - √ γ: parâmetro de forma
    (adimensional)
- A exponencial é uma caso especial da Weibull (γ=1)

- Função de risco: $\lambda(t) = \gamma\, \alpha^{\gamma}\, t^{\gamma-1}$
  - √ Função potência
  - √ Risco não é constante ao longo do tempo
    - – A função de risco é monótona
  - √ Parâmetro de forma da função de risco
    - – Modela a variação do risco no tempo
    - $\gamma < 1$: função de risco decresce
    - $\gamma > 1$: função de risco cresce
    - $\gamma = 1$: função de risco constante
- Função de risco acumulado: $\Lambda(t) = -\ln S(t) = (\alpha t)^{\gamma}$

Análise de Sobrevivência - 2017

23

- Weibull – Forma típica das funções



- Flexibilidade de formas da distribuição
- Parâmetro de forma γ:
  - $\gamma < 1$: função de risco decresce
  - $\gamma > 1$: função de risco cresce
  - $\gamma = 1$: função de risco constante

Análise de Sobrevivência - 2017

24

- Exponencial – Esperança e Variância:

$$\mathrm{E}[T] = \frac{1}{\alpha}\Gamma\left(1+\frac{1}{\gamma}\right)$$

$$\mathrm{Var}[T] = \frac{1}{\alpha^2}\left[\Gamma\left(1+\frac{2}{\gamma}\right) - \Gamma^2\left(1+\frac{1}{\gamma}\right)\right]$$

$$\Gamma(k) = \int_0^{\infty} x^{k-1}\mathrm{e}^{-x}\mathrm{d}x$$

- Função quantílica: $F(t_p) = p,\ 0 < p < 1$

$$1 - \exp\{-(\alpha t_p)^{\gamma}\} = p$$

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

- Mediana: $t_{med} = \dfrac{[-\ln(1-0,5)]^{\frac{1}{\gamma}}}{\alpha} = \dfrac{[\ln(2)]^{\frac{1}{\gamma}}}{\alpha}$
  - √ $F(t_{med}) = 0,5$

Análise de Sobrevivência - 2017

25

Weibull – Comentários

- Para a maioria dos fenômenos de saúde
  - √ É plausível supor que o risco não é constante no tempo
- Função mais utilizada na modelagem paramétrica em saúde
- Muito flexível
  - √ O parâmetro γ modela a variação do risco no tempo

Análise de Sobrevivência - 2017

27

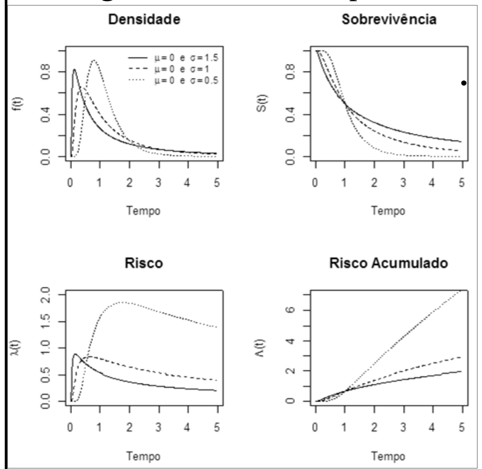
Distribuição Lognormal

- Densidade: $f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)\right\}$
  - √ $t > 0$, $-\infty < \mu < \infty$ e $\sigma > 0$
- Função de sobrevivência: $S(t) = \Phi\left(\frac{-\ln(t)+\mu}{\sigma}\right)$
  - √ Φ: função de distribuição acumulada da normal padrão
- Parâmetros:
  - √ O logaritmo dos dados de uma lognormal têm distribuição normal.

- Função de risco: $\lambda(t) = \frac{f(t)}{S(t)}$
  - √ A função de risco não é monótona como as da distribuição de Weibull
- Função de risco acumulado:

$$\Lambda(t) = -\ln S(t)$$
$$= -\ln\left[\Phi\left(\frac{-\ln(t)+\mu}{\sigma}\right)\right]$$

- Lognormal – Forma típica das funções



- Comportamento da função de risco:
  - Crescem, atingem um valor máximo e depois decrescem

- Lognormal – Esperança e Variância:

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

- Função quantílica:
  - √ z_p: 100p% percentil da normal padrão

$$F(t_p) = p,\ 0 < p < 1$$
$$\mathrm{P}\{T \le t_p\} = p$$
$$\mathrm{P}\{\ln(T) \le \ln(t_p)\} = \Phi\left(\frac{\ln(t_p) - \mu}{\sigma}\right)$$
$$t_p = \exp\{\mu + z_p\sigma\}$$

- Mediana: $t_{med} = \exp\{\mu\}$
  - √ $F(t_{med}) = 0{,}5$

Lognormal – Comentários

- Muito utilizada para caracterizar tempo de vida de indivíduos
  - √ Tempo de vida de pacientes com leucemia
- $\ln(T) \sim N(\mu, \sigma^2)$
  - √ μ: média do logaritmo do tempo de sobrevivência
  - √ σ: desvio padrão do logaritmo do tempo de sobrevivência

Análise de Sobrevivência - 2017

36

Estimação dos Parâmetros dos Modelos

Estimação de Parâmetros

- Parâmetros do modelo devem ser estimados a partir das observações amostrais
  - √ Determinação do modelo para responder as perguntas de interesse

Análise de Sobrevivência - 2017

40

Métodos de Estimação

- Método de Mínimos Quadrados:
  - √ Inadequado para estudo de tempo de sobrevida
  - √ Censura não é incorporada no processo de estimação

Análise de Sobrevivência - 2017

41

- Método de Máxima Verossimilhança:
  - √ Opção apropriada para dados censurados
  - √ Incorpora a censura
  - √ Relativamente simples de ser entendido
  - √ Possui boas propriedades para grandes amostras

Método da Máxima Verossimilhança

- Exemplo de Motivação:
  - √ Dados oriundos de população binomial com parâmetros 10 e p_0.

    p_0: constante e desconhecido

    Função de probabilidade de X:

$$f(x; p_0) = \mathrm{P}(X = x) = \binom{10}{x} p_0^x (1 - p_0)^{10-x}$$

    Observa-se X = 3
  - √ Objetivo:
    - Basear-se no dado disponível para estimar o valor verdadeiro do parâmetro

- Considerando p_0 = ½. (p_0 é desconhecido)
  - √ Probabilidade de gerar o dado que realmente observamos (X = 3) é:

$$\mathrm{P}(X = 3) = f(3; 0,5) = \binom{10}{3} (0,5)^3 (0,5)^7 \approx 0,117$$

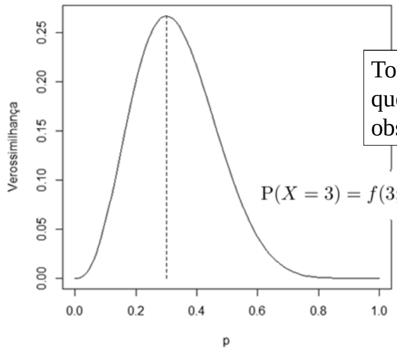
- Podemos calcular esta probabilidade sob a condição que $p_0 = p$, $0 \leq p \leq 1$

$$L(p; 3) = \binom{10}{3} (p)^3 (1 - p)^7,\ p \in [0, 1]$$

  - √ $L(.)$ é denominada função de verossimilhança.

Princípio da Máxima Verossimilhança

- Usamos como nossa estimativa de p_0 o valor de p que faz $L(p; 3)$ o maior possível



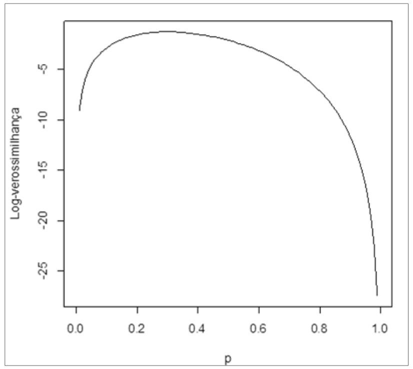
Toma-se o valor do parâmetro que torna mais provável o dado observado

$$\mathrm{P}(X = 3) = f(3; 0,3) = \binom{10}{3} (0,3)^3 (0,7)^7 \approx 0,267$$

Função de Log-Verossimilhança

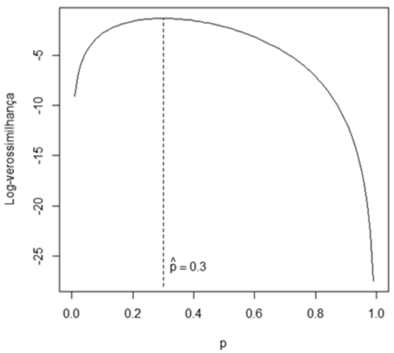
- A função log é crescente
  - √ Valor de p que maximiza $L(p; 3)$ é o mesmo que maximiza $\log L(p; 3)$
  - √ Em geral, é conveniente maximizar $\log L(p; 3)$ ao invés de $L(p; 3)$
  - √ No exemplo, a função de log-verossimilhança é definida como:

$$l(p;3) \stackrel{\text{def}}{=} \ln L(p;3)$$
$$= 3\ln p + 7\log(1-p) + \ln\binom{10}{3}$$



$$l(p;3) = 3\ln p + 7\ln(1-p) + \ln\binom{10}{3}$$

√ Pode-se calcular o ponto máximo da função
- ponto no domínio em que a derivada é zero



$$\hat{p} = \frac{3}{10}$$

- Estimativa é determinada pelo valor de X
  - √ Se tivéssemos observado X = k, teríamos a estimativa

$$\hat{p} = \frac{k}{10}$$

- No exemplo:
  - √ Estimador de máxima verossimilhança é:

$$\hat{p} = \frac{X}{10}$$

Máxima Verossimilhança – Procedimento de Estimação

- Baseado nos resultados observados, entre todas as distribuições definidas pelos possíveis valores de seus parâmetros, qual é aquela com maior probabilidade de ter gerado tal amostra?
  - √ A função de verossimilhança avalia o quanto os dados apoiam, concordam ou suportam cada valor possível do parâmetro a ser estimado

Análise de Sobrevivência - 2017 51

Função de Verossimilhança – Definição

- Suponha T uma variável aleatória com distribuição de probabilidades f(t; θ), em que θ é um único parâmetro desconhecido.
  - √ Sejam t_1, t_2, ..., t_n os valores observados de uma certa população de interesse.
  - √ A função de verossimilhança para θ é expressa por:

$$L(\theta; t_1, t_2, \ldots t_n) = f(t_1;\theta)f(t_2;\theta)\ldots f(t_n;\theta)$$
$$= \prod_{i=1}^{n} f(t_i, \theta)$$

Análise de Sobrevivência - 2017 52

Estimador de Máxima Verossimilhança

- O estimador de máxima verossimilhança (EMV) de θ é o valor $\hat{\theta}$ que maximiza a função de verossimilhança L(θ; t)

$$L(\hat{\theta}; \mathbf{t}) = \max_{\theta \in \Theta} L(\theta; \mathbf{t})$$

  - √ Valor de θ que maximiza a probabilidade da amostra observada ocorrer.
  - √ No caso discreto, o EMV é um estimador que maximiza a probabilidade de ocorrência dos valores da amostra

Análise de Sobrevivência - 2017 53

Exemplo – Distribuição de Bernoulli

- Seja X uma variável aleatória de Bernoulli
  - √ Função de probabilidade:

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x} & , x = 0, 1 \\ 0 & , \text{caso contrário} \end{cases}$$

p é o parâmetro desconhecido a ser estimado

Análise de Sobrevivência - 2017 54

- Função de verossimilhança da amostra

$$
\begin{aligned}
L(p;\mathbf{x}) &= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}\dots p^{x_n}(1-p)^{1-x_n} \\
&= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}
\end{aligned}
$$

- Função de log-verossimilhança

$$
l(p;\mathbf{x}) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)
$$

- Estimativa de máxima verossimilhança da amostra **x**:

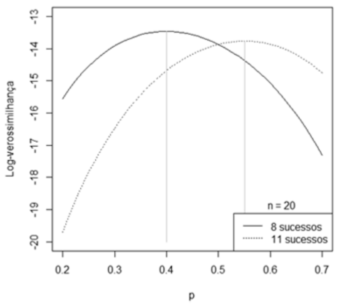
$$
\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}
$$

- Estimador de máxima verossimilhança para amostras de Bernoulli
  - √ Proporção de sucessos na amostra

$$
\hat{p} = \bar{X}
$$

Exemplo

- Bernoulli com parâmetro p desconhecido
  - √ Duas amostras de n= 20 com $\sum_{i=1}^{20} x_i = 8$ e $\sum_{i=1}^{20} x_i = 11$

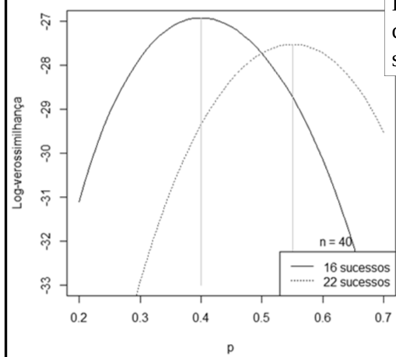


Pontos de máxima verossimilhança correspondentes à proporção de sucessos de cada amostra

$$
\hat{p}_1 = \frac{8}{20} = 0,40
$$

$$
\hat{p}_2 = \frac{11}{20} = 0,55
$$

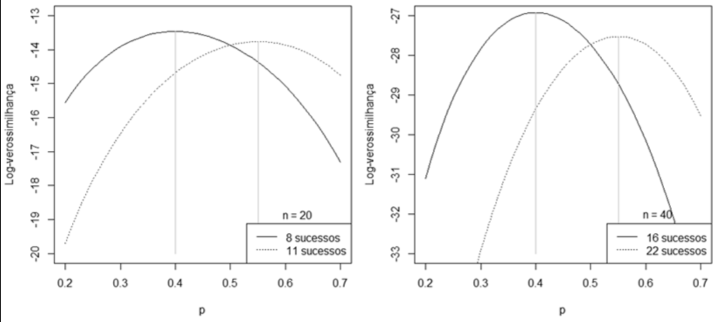
√ Duas amostras de n= 40 com $\sum_{i=1}^{20} x_i = 16$ e $\sum_{i=1}^{20} x_i = 22$



Pontos de máxima verossimilhança correspondentes à proporção de sucessos de cada amostra

$$
\hat{p}_1 = \frac{16}{40} = 0,40
$$

$$
\hat{p}_2 = \frac{22}{40} = 0,55
$$



√ As curvas crescem (decrescem) mais acentuadamente nas proximidades dos pontos de máximo na amostra maior (n = 40)

– Variância do EMV é menor para amostras maiores)

59

Análise de Sobrevivência - 2017

Exemplo

- Amostra para estimar prevalência de hipertensão
  - √ 10% dos indivíduos da amostra são hipertensos
  - √ É muita baixa a plausibilidade de a proporção de hipertensos na população ser 90%
  - √ Quanto mais próximo de 10% maior a verossimilhança
    - – Máxima verossimilhança

60

Análise de Sobrevivência - 2017

Função de Verossimilhança

- Dados não censurados:
  - √ Contribuição de cada observação:
    - – Função de densidade de probabilidade
  - √ Função de verossimilhança:

$$L(\theta; t_1, t_2, \ldots t_n) \propto \prod_{i \in O} f(t_i, \theta)$$

O: conjunto das observações de desfecho

61

Análise de Sobrevivência - 2017

- Dados censurados à direita:
  - √ Contribuição de cada observação:
    - – Informa que o tempo de sobrevida é maior que o tempo observado
    - – Contribuição para L(θ) é sua função de sobrevivência

      (não é sua função de densidade!)
  - √ Função de verossimilhança:

$$L(\theta) \propto \prod_{i \in O} f(t_i, \theta) \prod_{i \in D} S(t_i, \theta)$$

O: conjunto das observações de desfecho

D: conjunto das observações que sofreram censura à direita

62

Análise de Sobrevivência - 2017

- Dados censurados à esquerda:
  - √ Contribuição de cada observação:
    - – Sabemos somente que o tempo de sobrevivência é menor que o tempo observado
    - – Contribuição para L(θ) é sua função de distribuição acumulada:

      F(t; θ) = 1 – S(t; θ)
  - √ Função de verossimilhança:

$$L(\theta) \propto \prod_{i \in O} f(t_i, \theta) \prod_{i \in D} S(t_i, \theta) \prod_{i \in E} [1 - S(t_i; \theta)]$$

O: conjunto das observações de desfecho

E: conjunto observações com censura à esquerda

D: conjunto das observações com censura à direita

- Dados com censura intervalar:
  - √ Contribuição de cada observação:
    - – Sabemos que o tempo de sobrevivência está dentro de um intervalo entre t^- e t^+.
    - – Contribuição para L(θ) é a probabilidade de ocorrência de desfecho nesse intervalo de tempo:

      $S(t^-; \theta) - S(t^+; \theta)$
  - √ Função de verossimilhança:

$$L(\theta) \propto \prod_{i \in O} f(t_i, \theta) \prod_{i \in D} S(t_i, \theta) \prod_{i \in E} [1 - S(t_i; \theta)] \prod_{i \in I} [S(t_i^-; \theta) - (t_i^+; \theta)]$$

O: conjunto das observações de desfecho

D: conjunto das observações com censura à direita

E: conjunto observações com censura à esquerda

I: conjunto das observações com censura intervalar

Truncamentos

- São usadas probabilidades condicionais na construção da verossimilhança
  - √ Probabilidade de observação de um evento está condicionada à probabilidade do indivíduo ser incluído no estudo
    - – Probabilidade de inclusão no estudo de indivíduos com tempo de sobrevivência muito pequeno é menor do que a de indivíduos com tempo de sobrevivência grande
- Janela temporal do estudo:
  - √ (t_E, t_D)

- Truncamentos à esquerda:
  - √ Probabilidade de o evento ocorrer condicionado ao fato de o indivíduo ter sobrevivido ao estudo
  - √ Contribuição de indivíduo para a verossimilhança é ponderada pela sobrevivência no início do estudo

$$\frac{f(t)}{S(t_E)}$$

- Truncamentos à direita:
  - √ Contribuição indivídual para a verossimilhança é ponderada pela probabilidade acumulada de ocorrência do evento

$$\frac{f(t_i)}{1 - S(t_i)}$$

Estimadores de Máxima Verossimilhança – Dados Censurados

- Forma geral da verossimilhança
  - √ considerados os mecanismos de censura à direita: $L(\theta) = \prod_{i \in O} f(t_i, \theta) \prod_{i \in D} S(t_i, \theta)$
  - √ ou $$L(\theta) = \prod_{i=1}^{n} [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i}]$$
  - √ Estimadores de máxima verossimilhança são os valores de θ que maximizam L(θ) ou, equivalentemente, l(θ) = ln(L(θ)).

- √ Estimadores de máxima verossimilhança são encontrados resolvendo-se o sistema de equações

$$U(\theta) = \frac{\partial \ln(\theta)}{\partial \theta} = \frac{\partial l(\theta)}{\partial \theta} = 0$$

- √ A função de verossimilhança pode também ser expressa em termos da função de risco:

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

Distribuição Exponencial

- √ Suponha uma amostra de n itens de uma população exponencial, em que r ≤ n são desfechos e os demais (r – n) são censuras

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^{n} [f(t_i, \alpha)]^{\delta_i} [S(t_i, \alpha)]^{1-\delta_i}] \\ &= \prod_{i=1}^{n} [\alpha \exp\{-\alpha t_i\}]^{\delta_i} [\exp\{-\alpha t_i\}]^{1-\delta_i}] \\ &= \prod_{i=1}^{n} (\alpha)^{\delta_i} \exp\{-\alpha t_i\} \\ &= \alpha^{\sum_{i=1}^{n} \delta_i} \exp\{-\alpha \textstyle\sum_{i=1}^{n} t_i\} \end{aligned}$$

√ Função da log-verossimilhança:

$$l(\alpha) = \ln\left(\alpha^{\sum_{i=1}^n \delta_i} \exp\{-\alpha \textstyle\sum_{i=1}^n t_i\}\right)$$

$$= \sum_{i=1}^n \delta_i \ln(\alpha) - \alpha \sum_{i=1}^n t_i$$

√ Estimador de máxima verossimilhança:

$$\frac{\partial l(\alpha)}{\partial \alpha} = \frac{\sum_{i=1}^n \delta_i}{\alpha} - \sum_{i=1}^n t_i \qquad \frac{\partial l(\alpha)}{\partial \alpha} = 0 \Rightarrow \hat{\alpha} = \frac{r}{\sum_{i=1}^n t_i}$$

$$= \frac{r}{\alpha} - \sum_{i=1}^n t_i \qquad \hat{\alpha} = \frac{\text{total desfechos}}{\text{tempo total sob teste}}$$

Se não houver censuras $\hat{\alpha} = \frac{n}{\sum_{i=1}^n t_i} = \frac{1}{\bar{x}}$

Exemplo

- Sobrevivência ao diagnóstico de Aids
  - √ Pacientes com tempo de observação menor que 90 dias
    - Óbitos: 15
    - Censuras: 6
  - √ Suposição:
    - Tempo entre diagnóstico de Aids e a morte segue uma distribuição exponencial

- Estimação do parâmetro α da exponencial:

$$\hat{\alpha} = \frac{\text{total de desfechos}}{\sum_{i=1}^n t_i} = \frac{15}{1012} = 0,0148$$

  - √ Taxa de mortalidade por pessoa-tempo
    - 0,015 morte por pessoa-mês
    - 17,8 mortes por 100 pessoas/ano
- Tempo médio de sobrevivência:

$$\hat{\text{E}}[T] = \frac{1}{\hat{\alpha}} = 67,5 \text{ meses}$$

- Função de Sobrevivência estimada

$$\hat{S}_e(t) = \exp\{-\hat{\alpha}t\} = \exp\{0,0148t\}$$

- Estimativa da probabilidade de um paciente estar vivo depois de 16 dias do diagnóstico de Aids
  - √ Paramétrica – Exponencial:

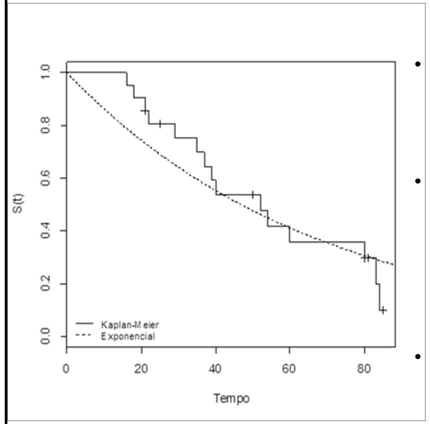
$$\hat{S}_e(16) = \exp\{-(0,0148)(16)\} = 0,7891$$

  - √ Não Paramétrica – Kaplan-Meier

$$\hat{S}_K M(16) = 0,9524$$

  - √ As estimativas são bastante diferentes
    - A suposição de distribuição exponencial é adequada aos dados?

- Estimativas paramétrica e não paramétrica



- Exponencial está afastada dos dados?
  - Ela é adequada à situação? Risco constante?
- As distribuições de Weibull ou Lognormal seriam mais próximas dos dados?
  - Como estimar seus parâmetros?
- O risco é constante?

77

Análise de Sobrevivência - 2017

Modelos Paramétricos de Sobrevivência

- Parametrização dos modelos no R
  - √ Exponencial:
    - Parâmetro α: $\alpha = \exp\{-\text{Intercept}\}$
  - √ Weibull:
    - Parâmetro α: $\alpha = \exp\{-\text{Intercept}\}$
    - Parâmetro γ: $\hat{\gamma} = \dfrac{1}{\text{scale}}$

78

Análise de Sobrevivência - 2017

Modelos Paramétricos de Regressão

- 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

79

Análise de Sobrevivência - 2017

√ Modelo Exponencial:

```
>           # Exponencial
>
> sobrev.e<-survreg(Surv(tempo, status)~1,data = ipec90, dist='exponential')
> sobrev.e
```

√ Saída:

```
> sobrev.e
Call:
survreg(formula = Surv(tempo, status) ~ 1, data = ipec90, dist = "exponential")
Coefficients:
(Intercept)
   4.211634
Scale fixed at 1
Loglik(model)= -78.2   Loglik(intercept only)= -78.2
n= 21
```

√ Estimação parâmetro:

```
> alfa.exp<-exp(-sobrev.e$coefficients[1]) #(na forma alpha t)
> alfa.exp
(Intercept)
 0.01482213
```

√ Sobrevivência estimada: $\hat{S}_e(t) = \exp\{-0,01482\,t\}$

80

Análise de Sobrevivência - 2017

Outras Distribuições

- Tempos de sobrevivência com distribuição de Weibull ou Lognormal:
  - √ Construção da verossimilhança é feita da mesma forma
    (substitui-se funções de densidade e sobrevivência)
  - √ Equações não podem ser resolvidas analiticamente
  - √ Necessário utilizar métodos iterativos de solução
    - Por exemplo, método de Newton-Raphson

√ Modelo Weibull:

```
        # Weibull

> sobrev.w<-survreg(Surv(tempo, status)~1,data = ipec90, dist='weibull')
> sobrev.w
```

√ Saída:

```
> sobrev.w<-survreg(Surv(tempo, status)~1,data = ipec90, dist='weibull')
> sobrev.w
Call:
survreg(formula = Surv(tempo, status) ~ 1, data = ipec90, dist = "weibull")
Coefficients:
(Intercept)
   4.157605
Scale= 0.5156146
Loglik(model)= -74.4   Loglik(intercept only)= -74.4
n= 21
```

√ Estimação parâmetro:

```
> alfa.weib<-exp(-sobrev.w$coefficients[1])
> gama.weib<-1/sobrev.w$scale
> cbind(gama.weib, alfa.weib)
            gama.weib  alfa.weib
(Intercept)  1.939433 0.01564499
```

√ Sobrevivência estimada: $\hat{S}_w(t) = \exp\left\{-(0,01564\,t)^{1,9394}\right\}$

√ Modelo Log-normal:

```
> # Lognormal
>
> sobrev.ln <- survreg(Surv(tempo, status)~1,data = ipec90, dist='lognorm')
> sobrev.ln
```

√ Saída:

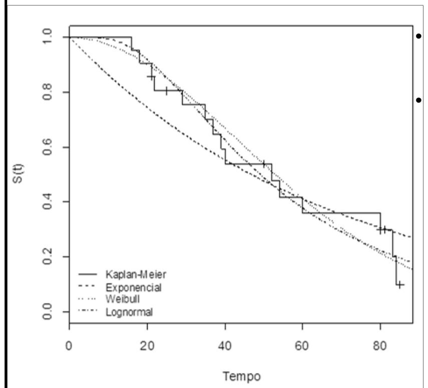
```
> sobrev.ln
Call:
survreg(formula = Surv(tempo, status) ~ 1, data = ipec90, dist = "lognorm")
Coefficients:
(Intercept)
   3.893372
Scale= 0.6425924
Loglik(model)= -73.8   Loglik(intercept only)= -73.8
n= 21
```

√ Estimação parâmetro:

```
> mi.ln <- sobrev.ln$coefficients[1]
> sigma.ln <- sobrev.ln$scale
> cbind(mi.ln, sigma.ln)
               mi.ln  sigma.ln
(Intercept) 3.893372 0.6425924
```

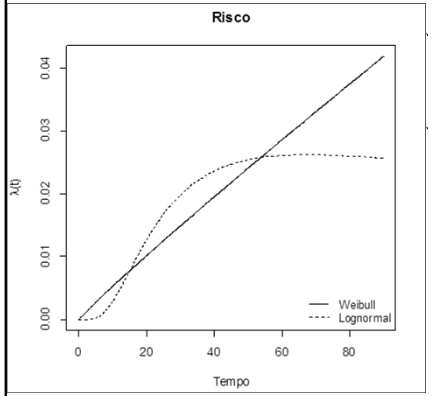
√ Sobrevivência estimada: $\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(t)-3,8934}{0,6426}\right]$

- Curva de Sobrevivência
  - √ Estimativas paramétricas e não paramétrica



- Exponencial está afastada dos dados
- Weibull e Lognormal estão mais próximas
  - Weibull se afasta para tempos de sobrevivência mais longos

- Função de Risco
  - √ Modelos Weibull e Lognormal



- Qual das funções de risco representam o problema mais adequadamente?
- Que critérios adotar para escolher o modelo?

87

Análise de Sobrevivência - 2017

Intervalos de Confiança e Testes de Hipóteses

Intervalos de Confiança

- Construção de intervalos de confiança para parâmetros e quantidades de interesse
  - √ Método de máxima verossimilhança.
- Distribuição assintótica do EMV $\hat{\theta}$:
  - √ Para grandes amostras:

$$\hat{\theta} \overset{as.}{\sim} N(\theta, \mathrm{Var}(\hat{\theta}))$$

  - √ $\widehat{\mathrm{Var}}(\theta)$: variância de $\hat{\theta}$:
    - Em geral, depende de θ.

89

Análise de Sobrevivência - 2017

√ Intervalo aproximado de (1 – a)100% de confiança para θ:

$$\hat{\theta} \pm z_{\alpha/2}\, \mathrm{ep}(\hat{\theta})$$

– $\widehat{\mathrm{ep}}(\theta)$: erro padrão de $\hat{\theta}$:

91

Análise de Sobrevivência - 2017

Testes de Hipóteses

- Seja um modelo com vetor de parâmetros

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$$

  - √ Deseja-se testar hipóteses relacionadas a este vetor ou a algum de seus subconjuntos
- Testes utilizados:
  - √ Teste da Razão de Verossimilhanças
  - √ Teste de Wald
  - √ Teste Escore

Teste da Razão de Verossimilhanças

- Compara os valores da log-verossimilhança maximizada sem restrição e sob H_0
  - √ Hipótese:
    - H_0: $\theta = \theta_0$.
  - √ Estatística de teste: $\mathrm{RV} = 2\ln\left[\frac{L(\hat{\theta}_0)}{L(\hat{\theta})}\right] = 2\left[\ln L(\hat{\theta}) - L(\hat{\theta}_0)\right]$

$$= 2[l(\hat{\theta}) - l(\hat{\theta}_0)]$$

  - √ Distribuição amostral: $\mathrm{RV} \overset{as.}{\sim} \chi^2_{gl}$
    - gl: diferença entre o número de parâmetros não especificados em H_0 e H_1

Teste de Wald

- Baseado na distribuição assintótica de θ:
  - √ Em geral, usado para testar hipóteses relativas a um único parâmetro θ_j.
  - √ Hipótese:
    - H_0: $\theta = \theta_0$.
  - √ Estatística de teste: $\mathrm{W} = \frac{(\hat{\theta}-\theta_0)^2}{\widehat{\mathrm{Var}}(\hat{\theta})} \overset{as.}{\sim} \chi^2_1$
    - ou, equivalentemente $\frac{\hat{\theta}-\theta_0}{\widehat{\mathrm{ep}}(\hat{\theta})} \overset{as.}{\sim} N(0,1)$

Escolha do Modelo

Escolha do Modelo Probabilístico

- Especificação do modelo
  - √ É de extrema importância na análise paramétrica de dados de tempo de sobrevivência
- Em muitas situações:
  - √ Não há evidências de testes anteriores de que um modelo se ajusta bem aos dados
  - √ Solução para essas situações é basicamente empírica

100

Análise de Sobrevivência - 2017

- Proposta empírica:
  - √ Ajustar modelos probabilísticos
    - – Comparação entre valores estimados e observados para decidir qual deles "melhor explica os dados amostrais"
  - √ Técnicas gráficas:
    - – Forma simples e eficiente para selecionar o "melhor" modelo para ser usado por um conjunto de dados
    - – Ferramenta exploratória para escolha de distribuições candidatas
  - √ Teste de hipóteses:
    - – Pode ser usado teste de hipóteses para modelos encaixados

101

Análise de Sobrevivência - 2017

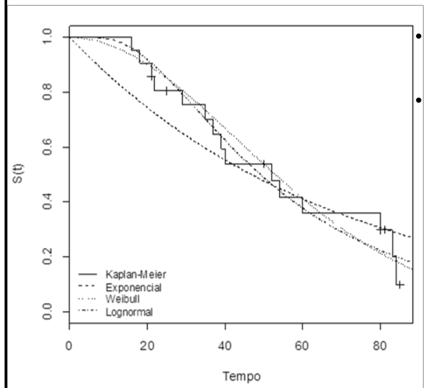
Métodos Gráficos

1. Comparação do modelo proposto com o estimador de Kaplan-Meier:
   - √ Ajuste do modelo proposto aos dados
   - √ Obtenção da estimativa de Kaplan-Meier
   - √ Comparação gráfica da função de sobrevivência do modelo paramétrico
   - √ Quanto mais próximo o modelo paramétrico estiver da curva do Kaplan-Meier, melhor o modelo

102

Análise de Sobrevivência - 2017

- Comparação das curvas de sobrevivência
  - √ Estimativas paramétricas e não paramétrica

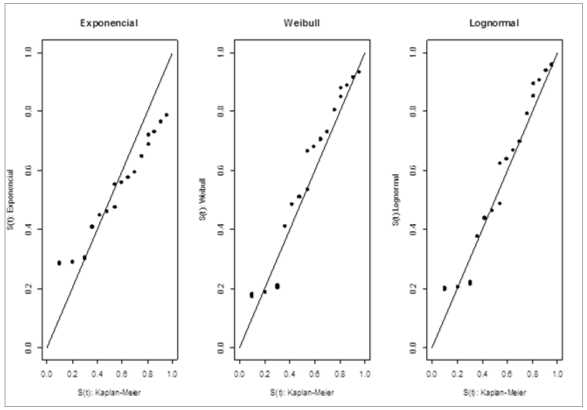


- Exponencial está afastada dos dados
- Weibull e Lognormal estão mais próximas
  - Weibull se afasta para tempos de sobrevivência mais longos

103

Análise de Sobrevivência - 2017

- Comparação entre os valores estimados
  - √ (S_{km}, $S_{paramétrico}$)



104

Análise de Sobrevivência - 2017

- Comentários:
  - √ Modelo exponencial parece não ser adequado aos dados:
    - (Curva encontra-se um tanto afastada da reta $y = x$)
  - √ Modelo de Weibull ou log-normal possivelmente é adequado aos dados
    - (acompanham mais de perto a reta $y = x$)

105

Análise de Sobrevivência - 2017

- Comparação do modelo proposto com o estimador de Nelson-Aalen:
  - √ Ajuste do modelo proposto aos dados
  - √ Obtenção da estimativa de Nelson-Aalen para a função de risco acumulado
  - √ Comparação gráfica das funções de sobrevivência ou de risco acumulado obtidas
  - √ Gráficos envolvendo função de sobrevivência ou a função de risco acumulado são úteis para discriminar modelos

106

Análise de Sobrevivência - 2017

Função de risco acumulado:

$$\Lambda(t) = -\ln[S(t)]$$

- √ Modelo exponencial:

$$\hat{\Lambda}(t) = \hat{\alpha} t$$

- √ Modelo de Weibull:

$$\hat{\Lambda}(t) = (\hat{\alpha} t)^{\hat{\gamma}}$$

- √ Modelo Lognormal:

$$\hat{\Lambda}(t) = -\ln\left[\Phi\left(\frac{-\ln(t) + \hat{\mu}}{\hat{\sigma}}\right)\right]$$

107

Análise de Sobrevivência - 2017

2. Linearização da função de sobrevivência:
   - √ Ideia básica:
     - Construção de gráficos que sejam aproximadamente lineares caso o modelo proposto seja apropriado
     - Violações da linearidade podem ser verificadas visualmente
   - √ Utiliza-se gráfico de transformação de linearização

- Linearização do modelo exponencial

$$S(t) = \exp\{-\alpha t\}$$
$$-\ln[S(t)] = \alpha t$$

√ Gráfico de $-\ln[\hat{S}_{KM}]$ vs. tempo deve ser aproximadamente linear se o modelo for adequado

- Linearização do modelo de Weibull

$$S(t) = \exp\{(\alpha t)^\gamma\}$$
$$-\ln S(t) = (\alpha t)^\gamma$$
$$\ln[-\ln[S(t)]] = \gamma \ln(\alpha) + \gamma \ln(t)$$

√ gráfico de $\ln[-\ln[\hat{S}_{KM}]]$ vs. tempo deve ser aproximadamente linear, se o modelo for adequado

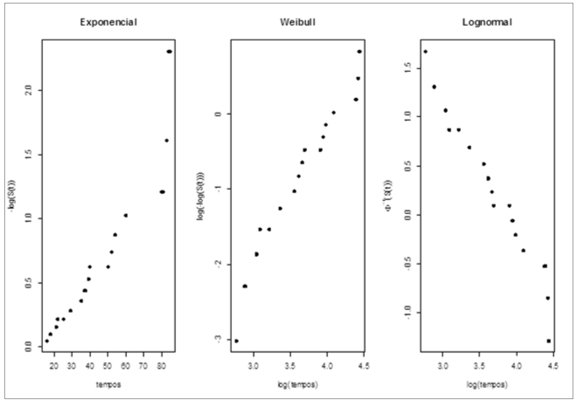
√ Indicação a favor do modelo exponencial
- Se além de linear o gráfico passar pela origem e tiver inclinação igual a 1.

- Linearização do modelo Lognormal

$$S(t) = \Phi\left(\frac{-\ln(t) + \mu}{\sigma}\right)$$
$$\Phi^{-1}[S(t)] = \frac{-\ln(t) + \mu}{\sigma}$$

√ gráfico de $\Phi^{-1}[\hat{S}_{KM}]$ vs. ln(tempo) deve ser aproximadamente linear, se o modelo for adequado
- Intercepto: μ/σ
- Inclinação: $1/\sigma$

- Gráficos linearizados da sobrevivência
  - √ Modelos exponencial, Weibull e lognormal



- Comentários:
  - √ Modelos de Weibull e Log-normal não mostram afastamentos marcantes de uma reta
  - √ Observa-se desvio para modelo Exponencial
  - √ Não há discriminação entre os modelos de Weibull e Lognormal
    - Principal razão pode ser tamanho amostral pequeno

Análise Gráfica – Comentários

- Avaliação da qualidade de ajuste do modelo
  - √ Análises gráficas são mais indicadas que testes formais de qualidade global de ajuste
  - √ Testes formais
    - Para amostras pequenas: tendem a ter baixo poder
    - Para amostras grandes: tendem a rejeitar o modelo
  - √ Se nenhum desses modelos for adequado
    - Podem ser necessários modelos mais flexíveis envolvendo mais de dois parâmetros
      - Ex.: gama generalizada

- Usar métodos gráficos para descartar modelos claramente inadequados
  - √ Gráficos não discriminam modelos, mas indicam que eles são igualmente bons
    - Tamanho amostral pequeno ou pequena quantidade de desfechos
  - √ Ajustes razoáveis de diferentes modelos para o mesmo conjunto de dados
    - Em especial, nos intervalos de tempo que concentram maior número de observações
    - Apresentam resultados semelhantes das quantidade de interesse

√ Presença de componente subjetivo na interpretação de técnicas gráficas

– Conclusões podem diferir para diferentes analistas

- Análise de resíduos é mais adequada para demonstrar que um modelo paramétrico é melhor

Comparação de Modelos – Teste de Hipóteses

Outra maneira de discriminar modelos

√ Hipóteses

– H_0: modelo de interesse é adequado

– H_1: modelo de interesse não é adequado

- Comparação de modelos encaixados:

√ Modelo com maior número de parâmetros contém todos os parâmetros do modelo menor

√ Identificar um modelo generalizado tal que modelo de interesse seja seu caso particular

- Ajustes para realizar o teste:

√ Obtenção da log-verossimilhança do modelo generalizado (modelo maior):

$$l(\theta_G) = \ln[L(\theta_G)]$$

√ Obtenção da log-verossimilhança do modelo de interesse (modelo menor):

$$l(\theta_M) = \ln[L(\theta_M)]$$

- Estatística de teste:

√ Razão de verossimilhanças: $\mathrm{RV} = 2\left[l(\hat{\theta}_G) - l(\hat{\theta}_M)\right]$

√ Distribuição amostral:

$$\mathrm{RV} \sim \chi^2_{gl}$$

– gl: diferença entre quantidade de parâmetros dos dois modelos

- Comparação modelos

√ Weibull (maior) vs. Exponencial(menor)

√ Logaritmos da verossimilhança:

```
> sobrev.e$loglik[2]
[1] -78.1745
> sobrev.w$loglik[2]
[1] -74.39242
```

√ Razão de Verossimilhanças:

$$\mathrm{RV} = 2\left[l(\hat{\theta}_G) - l(\hat{\theta}_M)\right]) = 2[78,1745 - 74,3942] = 7,5642$$

√ p-valor: $\mathrm{P}\left\{\chi^2_1 > 7,5642\right\} = 0,0059$

```
> 1 - pchisq(7.5642, 1)
[1] 0.005953908
```

– Rejeita-se H_0: modelo exponencial é adequado

- Condução do teste – comando **anova**:

```
> anova(sobrev.e, sobrev.w)
  Terms Resid. Df    -2*LL Test Df Deviance     Pr(>Chi)
1     1        20 156.3490      NA       NA           NA
2     1        19 148.7848    =  1 7.564177 0.005953983
```

√ Resultado indica inadequação do modelo exponencial

- No contexto de Análise de Sobrevivência, teste da razão de verossimilhanças é em geral utilizado com a distribuição gama generalizada.
  - √ Apresenta os modelos exponencial, de Weibull, lognormal e gama como modelos encaixados
- Para ajuste do modelo gama generalizado:
  - > **`library(flexsurv)`**
  - > **`flexsurvreg(formula, data, dist='gengamma')`**

121

Análise de Sobrevivência - 2017

- No exemplo:
  - √ Ajuste do modelo gama generalizado (3 parâmetros)

```
> library(flexsurv)
> sobrev.gama<-flexsurvreg(Surv(tempo, status)~1,data = ipec90,dist='gengamma')
> sobrev.gama
Call:
flexsurvreg(formula = Surv(tempo, status) ~ 1, data = ipec90,     dist =
"gengamma")
Estimates:
       est     L95%    U95%    se
mu      3.802   3.055   4.548   0.381
sigma   0.656   0.453   0.950   0.124
Q      -0.318  -2.642   2.006   1.186
N = 21,  Events: 15,  Censored: 6
Total time at risk: 1012
Log-likelihood = -73.8024, df = 3
AIC = 153.6048
```

122

Análise de Sobrevivência - 2017

- Logaritmos das verossimilhanças:

```
> sobrev.gama$loglik
[1] -73.8024
> sobrev.e$loglik[2]
[1] -78.1745
> sobrev.w$loglik[2]
[1] -74.39242
> sobrev.ln$loglik[2]
[1] -73.83833
```

- Teste da Razão das Verossimilhanças

	l(θ)	**Razão**	**GL**	**p-valor**
Gama generalizada	–73,80			
Exponencial	–78,17	2(78,17–73,80)=8,74	2	0,013
Weibull	–74,39	2(74,39–73,80)=1,18	1	0,277
Log-normal	–73,84	2(73,84–73,80)=0,08	1	0,777

```
> 1-pchisq(8.74,2)
[1] 0.01265124
```

  - √ Resultados indicam adequação dos modelos de Weibull e Log-normal

123

Análise de Sobrevivência - 2017

- Importante:
  - √ No R, a função **`anova`** não funciona com objeto **`flexsurvreg`**

124

Análise de Sobrevivência - 2017

AIC – Akaike Information Criterion

- Medida de qualidade relativa de um modelo estatístico para um conjunto de dados
  - √ Estimativa da informação perdida ao se usar um modelo para representar o processo de geração dos dados
  - √ Considera o ‘trade-off’ entre a bondade do ajuste e a complexidade do modelo
  - √ Não informa nada no sentido absoluto
    - O AIC não informa nada se todos os modelos candidatos não se ajustarem bem aos dados

- Critério de Informação de Akaike:

$$\text{AIC} = 2\left[p - l(\hat{\theta}_M)\right]$$

  - √ Correção para amostras finitas

$$\text{AIC}_c = \text{AIC} + \frac{2p(p+1)}{n-p-1}$$

  - √ Decisão:
    - Prefere-se o modelo com o menor valor de AIC
  - √ No exemplo

Modelo	l(θ)	AIC
Gama generalizada	–73,80	2[3 – (–73,80)] = 153,6
Exponencial	–78,17	2[1 – (–78,17)] = 158,3
Weibull	–74,39	2[2 – (–74,39)] = 152,8
Log-normal	–73,84	2[2 – (–73,84)] = 151,7

- Comentários:
  - √ Recomenda-se usar o AIC para selecionar modelos quando o número de observações (n) é maior que pelo menos 40 vezes o número de parâmetros (p)
    - (Burham e Anderson, 2002)
  - √ O AICc é recomendado para pequenas amostras e respostas com distribuição normal
    - Aumenta consideravelmente a probabilidade de escolha do modelo adequado. (Davison, 2001)
    - Deve-se considerar arriscado usar para distribuições não normais um critério que foi desenvolvido para distribuição normal

Exemplo de Aplicação

Modelos Paramétricos de Sobrevivência

- Parametrização dos modelos no R
  - √ Exponencial:
    - Parâmetro $\alpha = \exp\{-\text{Intercept}\}$
  - √ Weibull:
    - Parâmetro $\alpha = \exp\{-\text{Intercept}\}$
    - Parâmetro $\gamma = \dfrac{1}{\text{scale}}$

Análise de Sobrevivência - 2017 129

Exemplo

- Pacientes com câncer de bexiga
  - √ Tempo de reincidência de pacientes com câncer de bexiga submetidos a procedimento cirúrgico com laser
  - √ 20 pacientes
  - √ Modelos analisados:
    - Exponencial
    - Weibull
    - Log-normal

Análise de Sobrevivência - 2017 130

- Tempo (em meses):
  - √ 3, 5, 6, 7, 8, 9, 10, 10+, 12, 15, 15+, 18, 19, 20, 22, 25, 28, 30, 40, 45+

Análise de Sobrevivência - 2017 131

- Carregamento dos dados:

```
> library(survival)
> library(flexsurv)
>
> # Exemplo 3.1 - Câncer de Bexiga
>
> tempos <- c(3, 5, 6, 7, 8, 9, 10, 10, 12, 15, 15, 18, 19, 20, 22, 25, 28, 30,
+   40, 45)
> cens <- c(1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0)
> sobrevida.cancer <- Surv(tempos, cens)
> class(sobrevida.cancer)
>
[1] "Surv"
> str(sobrevida.cancer)
 Surv [1:20, 1:2]  3   5   6   7   8   9  10  10+ 12  15  ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:2] "time" "status"
 - attr(*, "type")= chr "right"
```

Análise de Sobrevivência - 2017 132

- Estimativas de máxima verossimilhança:
  - √ Modelo exponencial:

```
> (ajuste.e <- survreg(sobrevida.cancer ~ 1, dist = 'exponential'))
Call:
survreg(formula = sobrevida.cancer ~ 1, dist = "exponential")

Coefficients:
(Intercept)
   3.016111

Scale fixed at 1

Loglik(model)= -68.3   Loglik(intercept only)= -68.3
n= 20
```

  - √ Estimativa parâmetro:

```
> (alfa.exp <- exp(-ajuste.e$coefficients[1]))
(Intercept)
 0.04899135
```

$$\hat{\alpha} = \exp\{-\hat{\beta}_0\} = \exp\{-3,016111\} = 0,04899135$$

  - √ Sobrevivência estimada: $\hat{S}_e(t) = \exp\{-0,04899t\}$

133

Análise de Sobrevivência - 2017

- Estimativas de máxima verossimilhança:
  - √ Modelo Weibull:

```
> (ajuste.w <- survreg(sobrevida.cancer ~ 1, dist = 'weibull'))
Call:
survreg(formula = sobrevida.cancer ~ 1, dist = "weibull")

Coefficients:
(Intercept)
   3.060529

Scale= 0.647922

Loglik(model)= -66.1   Loglik(intercept only)= -66.1
n= 20
```

  - √ Estimativa parâmetro:

```
>   # estimativa dos parâmetros
> alfa.weib <- exp(-ajuste.w$coefficients[1])
> gama.weib <- 1/ajuste.w$scale
> cbind(alfa.weib, gama.weib)
             alfa.weib gama.weib
(Intercept) 0.04686288  1.543396
```

$$\hat{\alpha} = \exp\{-\hat{\beta}_0\} = \exp\{-3,060529\} = 0,04686288$$

$$\hat{\gamma} = 1/\hat{\sigma} = 1/0,647922 = 1,543396$$

134

Análise de Sobrevivência - 2017

- √ Sobrevivência estimada:

$$\hat{S}_w(t) = \exp\left\{-(0,04686t)^{1,5433}\right\}$$

135

Análise de Sobrevivência - 2017

- Estimativas de máxima verossimilhança:
  - √ Modelo Lognormal:

```
> (ajuste.ln<-survreg(sobrevida.cancer~1,dist='lognorm'))
Call:
survreg(formula = sobrevida.cancer ~ 1, dist = "lognorm")

Coefficients:
(Intercept)
   2.717176

Scale= 0.7648167

Loglik(model)= -65.7   Loglik(intercept only)= -65.7
n= 20
```

  - √ Estimativa parâmetro:

```
>   # estimativa dos parâmetros
> mi.ln<- ajuste.ln$coefficients[1]
> sigma.ln <- ajuste.ln$scale
> cbind(mi.ln, sigma.ln)
               mi.ln  sigma.ln
(Intercept) 2.717176 0.7648167
```

$$\hat{\mu} = 2,717176$$
$$\hat{\sigma} = 0,7648167$$

136

Análise de Sobrevivência - 2017

√ Sobrevivência estimada:

$$\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(t) - 2{,}7172}{0{,}7648}\right]$$

- Estimativas da função de sobrevivência:
  - √ Modelo exponencial: $\hat{S}_e(t) = \exp\{-0,04899t\}$
  - √ Modelo de Weibull: $\hat{S}_w(t) = \exp\left\{-(0,04686t)^{1,5433}\right\}$
  - √ Modelo Log-normal: $\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(t)-2,7172}{0,7648}\right)$
- Valores estimados para t = 10:
  - √ Exponencial: $\hat{S}_e(t) = \exp\{-(0,04899)(10)\} = 0,612$
  - √ Weibull: $\hat{S}_w(t) = \exp\left\{-(0,04686 \times 10)^{1,5433}\right\} = 0,732$
  - √ Log-normal: $\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(10)-2,7172}{0,7648}\right) = 0,708$
  - √ São bem próximas as estimativas obtidas pelos modelos de Weibull e Log-normal

- Estimativas da função de sobrevivência:
  - √ Modelo exponencial: $\hat{S}_e(t) = \exp\{-0,04899t\}$
  - √ Modelo de Weibull: $\hat{S}_w(t) = \exp\left\{-(0,04686t)^{1,5433}\right\}$
  - √ Modelo Log-normal: $\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(t)-2,7172}{0,7648}\right)$
- Valores estimados para t = 10:
  - √ Exponencial: $\hat{S}_e(t) = \exp\{-(0,04899)(10)\} = 0,612$
  - √ Weibull: $\hat{S}_w(t) = \exp\left\{-(0,04686 \times 10)^{1,5433}\right\} = 0,732$
  - √ Log-normal: $\hat{S}_{ln}(t) = \Phi\left[\frac{-\ln(10)-2,7172}{0,7648}\right) = 0,708$
  - √ São bem próximas as estimativas obtidas pelos modelos de Weibull e Log-normal

- Comparação das estimativas da sobrevivência para os tempos de reincidência:
  - √ Estimador Kaplan-Meier e função de sobrevivência dos modelos exponencial, Weibull e log-normal

- Comparação das sobrevivências:

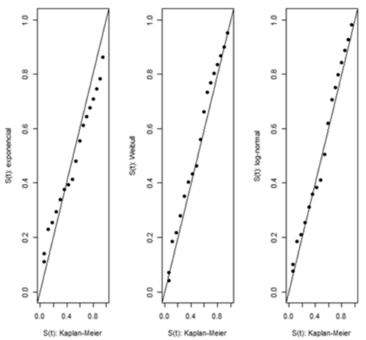
```
>   # estimativa de Kaplan-Meier
> e.km <- survfit(sobrevida.cancer~1)
> tempo.km <- e.km$time
> s.km <- e.km$surv
>   # esimativas modelos paramétricos
>  s.e <- exp(-alfa.exp*tempo.km)
>  s.w <- exp(-(alfa.weib*tempo.km)^gama.weib)
>  s.ln <- pnorm((-log(tempo.km)+ mi.ln)/sigma.ln)
>   # comparação das estimativas
>  cbind(tempo.km,s.km,s.e,s.w,s.ln)
 tempo.km       s.km       s.e        s.w       s.ln
 [1,]       3 0.95000000 0.8633164 0.95274148 0.98283934
 [2,]       5 0.90000000 0.7827384 0.89897484 0.92624322
 [3,]       6 0.85000000 0.7453152 0.86839357 0.88685752
 [4,]       7 0.80000000 0.7096812 0.83609525 0.84337638
 [5,]       8 0.75000000 0.6757509 0.80253272 0.79781416
 [6,]       9 0.70000000 0.6434428 0.76809812 0.75169629
 [7,]      10 0.65000000 0.6126794 0.73313414 0.70611769
 [8,]      12 0.59583333 0.5554947 0.66278292 0.61931883
 [9,]      15 0.54166667 0.4795676 0.55966698 0.50475984
[10,]      18 0.48148148 0.4140186 0.46346069 0.41042396
[11,]      19 0.42129630 0.3942241 0.43345774 0.3831769
```

√ Estimativas Weibull e lognormal bem próximas

- Escolha do modelo – Método Gráfico:
  - √ Gráfico das estimativas da sobrevivência obtidas pelo estimador de Kaplan-Meier vs. estimativas das sobrevivências a partir dos modelos paramétricos

```
> par(mfrow=c(1,3))
>   # Exponencial
> plot(s.km, s.e, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
+ xlab = "S(t): Kaplan-Meier", ylab="S(t): exponencial")
> abline(0, 1, lty = 1)
>   # Weibull
> plot(s.km, s.w, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
+ xlab = "S(t): Kaplan-Meier", ylab="S(t): Weibull")
> abline(0, 1, lty = 1)
> # Log-normal
> plot(s.km, s.ln, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
+ xlab = "S(t): Kaplan-Meier", ylab = "S(t): log-normal")
> abline(0, 1, lty = 1)
```


- Comparação gráfica: KM vs. modelos

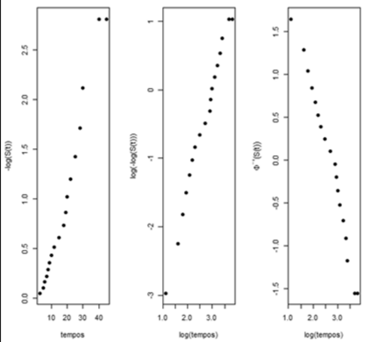


- Modelo exponencial parece não ser adequado aos dados
  - √ Curva afastada da reta
- Modelos Weibull e lognormal mais próximos da reta
  - √ acompanham mais de perto a reta $y = x$
  - √ Indicação de que um desses modelos possa ser adequado aos dados

- Gráficos linearizados dos modelos
  - √ Comandos

```
> par(mfrow=c(1,3))
> inv.km <- qnorm(s.km)
> plot(tempo.km, -log(s.km),pch=16,xlab="tempos",ylab="-log(S(t))")
> plot(log(tempo.km),log(-log(s.km)),pch=16,xlab="log(tempos)",ylab="log(-
log(S(t)))")
> plot(log(tempo.km),inv.km, pch=16,xlab="log(tempos)", ylab=expression(Phi^-1 *
(S(t))))
> par(mfrow = c(1, 1))
```

- Gráficos linearizados dos modelos



- Modelos exponencial e Weibull não se afastam da reta de forma marcante
  - √ Exponencial apresenta desvio
- Resultados anteriores são confirmados
- Tamanho amostral pequeno pode ser principal razão de não haver discriminação entre os modelos de Weibull e lognormal

145

Análise de Sobrevivência - 2017

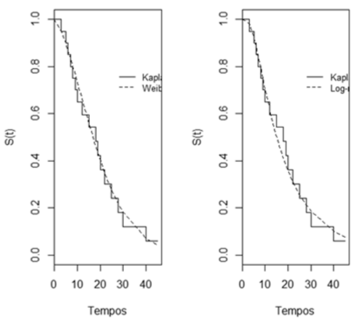
- Curvas estimados pelos modelos vs. KM
  - √ Comandos

```
>  par(mfrow=c(1,2))
>  plot(e.km, conf.int=F, xlab="Tempos", ylab="S(t)")
>  lines(c(0,tempo.km),c(1,s.w), lty=2)
>  legend(25,0.8,lty=c(1,2),c("Kaplan-Meier", "Weibull"),bty="n",cex=0.8)
>  plot(e.km, conf.int=F, xlab="Tempos", ylab="S(t)")
>  lines(c(0,tempo.km),c(1,s.ln), lty=2)
>  legend(25,0.8,lty=c(1,2),c("Kaplan-Meier", "Log-normal"),bty="n",cex=0.8)
```

146

Análise de Sobrevivência - 2017

- Curvas estimadas dos modelos vs. KM



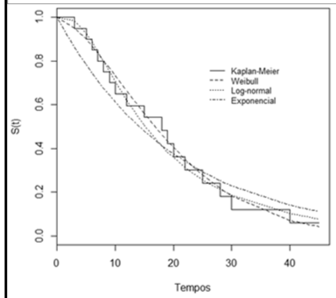
- Ambos modelos apresentam ajustes satisfatórios

147

Análise de Sobrevivência - 2017

- Curvas estimadas dos modelos:

```
> par(mfrow=c(1,1))
> plot(e.km, conf.int=F, xlab="Tempos", ylab="S(t)")
> lines(c(0,tempo.km),c(1,s.w), lty=2)
> lines(c(0,tempo.km),c(1,s.ln), lty=3)
> lines(c(0,tempo.km),c(1,s.e), lty=4)
> legend(25,0.8,lty=c(1,2,3,4),c("Kaplan-Meier", "Weibull", "Log-normal",
"Exponencial"), bty="n",cex=0.8)
```



- Curvas Weibull e Lognormal se afastam para maiores tempos

148

Análise de Sobrevivência - 2017

- Teste da razão de verossimilhança:
  - √ H_0: modelo de interesse é adequado
  - √ H_1: modelo gama generalizada (3 parâmetros)
    - – modelo de interesse é caso particular
- Ajuste da gama generalizada:

```
> # Gama-generalizada
> ajuste.gama <- flexsurvreg(sobrevida.cancer ~ 1, dist = 'gengamma')
> ajuste.gama$loglik
 [1] -65.69074
```


- Teste da razão de verossimilhanças:
- Modelos de interesse:
  - √ Exponencial, Weibull e log-normal
- Cálculo manual

```
> # TRV - manual
> modelos <- c("Exponencial", "Weibull", "Log-normal")
> loglik <- c(ajuste.gama$loglik, ajuste.e$loglik[2],
ajuste.w$loglik[2],
+   ajuste.ln$loglik[2])
> LR <- -2*(loglik[-1] - loglik[1])
> g <- 3 - c(1, 2, 2)
> p.valor <- 1 - pchisq(LR, g)
> data.frame(modelos, loglik[-1], LR, g, p.valor)
  modelos loglik..1.         LR g    p.valor
1 Exponencial  -68.27389 5.16630281 2 0.07553559
2     Weibull  -66.13336 0.88523500 1 0.34677185
3  Log-normal  -65.73990 0.09831756 1 0.75385805
```

$\boxed{H_0 : \text{modelo de interesse é adquado } (\theta_M)}$

$\boxed{RV = 2[l(\hat{\theta}_G) - l(\hat{\theta}_M)] \sim \chi_g}$

- – Resultados indicam a adequação dos modelos de Weibull e lognormal aos dados

- Logaritmos da verossimilhança:

```
> ajuste.gama$loglik
[1] -65.69074
>  ajuste.e$loglik[2]
[1] -68.27389
>  ajuste.w$loglik[2]
[1] -66.13336
>  ajuste.ln$loglik[2]
[1] -65.7399
```

- Teste da Razão de Verossimilhanças

	ln(θ)	Razão	GL	p-valor
Gama generalizada	–65,69			
Exponencial	–68,27	2(68,27–65,69)=5,16	2	0,075
Weibull	–66,13	2(66,13–65,69)=0,88	1	0,348
Log-normal	–65,74	2(65,74–65,69)=0,10	1	0,752

```
> 1-pchisq(5.16,2)
[1] 0.075774
```

- √ Resultados indicam adequação dos modelos de Weibull e Log-normal

- Teste da razão de verossimilhanças:
  - √ Weibull e exponencial

```
> anova(ajuste.e, ajuste.w)
  Terms Resid. Df    -2*LL Test Df Deviance    Pr(>Chi)
1     1        19 136.5478      NA       NA          NA
2     1        18 132.2667    =  1 4.281068 0.03853913
```

RV

- √ Modelo exponencial é rejeitado em relação ao Weibull (geral)

- Critério de Informação de Akaike:

$$\mathrm{AIC} = 2\left[p - l(\hat{\theta}_M)\right]$$

√ Correção para amostras finitas

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2p(p+1)}{n-p-1}$$

√ Prefere-se o modelo com o menor valor de AIC

√ No exemplo

Modelo	l(θ)	AIC
Gama generalizada	–65,69	2[3 – (–65,69)] = 137,38
Exponencial	–68,27	2[1 – (–68,27)] = 138,54
Weibull	–66,13	2[2 – (–66,13)] = 136,26
Log-normal	–65,74	2[2 – (–65,74)] = 135,48

√ Confirmam o descarte do modelo exponencial

- Critério de informação de Akaike

√ Cálculo manual

```
> modelos <- c("Gama", "Exponencial", "Weibull", "Log-normal")
> loglik <- c(ajuste.gama$loglik, ajuste.e$loglik[2], ajuste.w$loglik[2],
+   ajuste.ln$loglik[2])
> gl <- c(3, 1, 2, 2)
> aic <- 2*(gl - loglik)
> data.frame(modelos, aic)
      modelos      aic
1        Gama 137.3815
2 Exponencial 138.5478
3     Weibull 136.2667
4  Log-normal 135.4798
```

√ Comandos

```
> # comando
> AIC <- c(ajuste.gama$AIC, extractAIC(ajuste.e)[2], extractAIC(ajuste.w)[2],
+ extractAIC(ajuste.ln)[2])
> data.frame(modelos, AIC)
      modelos      AIC
1        Gama 137.3815
2 Exponencial 138.5478
3     Weibull 136.2667
4  Log-normal 135.4798
```


- Comentários:

√ Recomenda-se usar o AIC para selecionar modelos quando o número de observações (n) é maior que pelo menos 40 vezes o número de parâmetros (p)

– (Burham e Anderson, 2002)

√ O AICc é recomendado para pequenas amostras e respostas com distribuição normal

– Aumenta consideravelmente a probabilidade de escolha do modelo adequado. (Davison, 2001)

– Deve-se considerar arriscado usar para distribuições não normais um critério que foi desenvolvido para distribuição normal

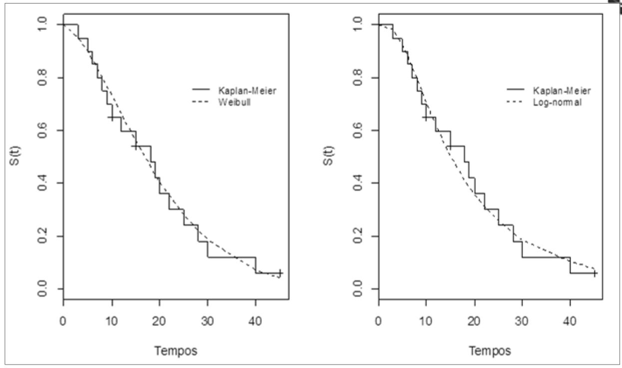
- Curvas de sobrevivência estimadas:
  – Modelos Weibull e log-normal
  – Estimador de Kaplan-Meier

√ Comandos:

```
        # Curvas de Sobrevivência Estimadas

par(mfrow=c(1,2))
plot(e.km, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,tempo.km),c(1,s.w), lty=2)
legend(25,0.8,lty=c(1,2),c("Kaplan-Meier", "Weibull"),bty="n",cex=0.8)
plot(e.km, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,tempo.km),c(1,s.ln), lty=2)
legend(25,0.8,lty=c(1,2),c("Kaplan-Meier", "Log-normal"),bty="n",cex=0.8)
```

- Curvas de sobrevivência estimadas:

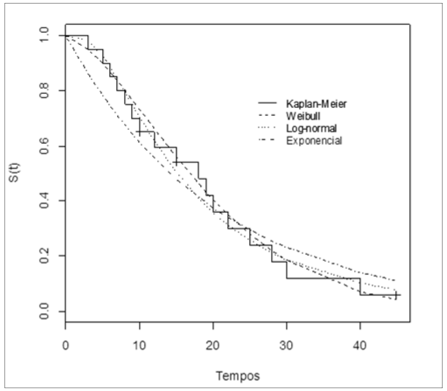


√ Ambos os modelos apresentam ajustes satisfatórios

158

Análise de Sobrevivência - 2017

- Comparação das curvas estimadas de sobrevivência



159

Análise de Sobrevivência - 2017

- Estimativa para o tempo médio:

√ Modelo Log-normal:

$$\begin{aligned}\hat{E}(T) &= \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \\ &= \exp\left\{2,72 + \frac{0,76^2}{2}\right\} \\ &= 20,263 \text{ meses}\end{aligned}$$

√ Modelo de Weibull

$$\begin{aligned}\hat{E}(T) &= \frac{1}{\alpha}\Gamma\left(1 + \frac{1}{\gamma}\right) \\ &= \frac{1}{0,0468}\Gamma\left(1 + \frac{1}{1,54}\right) \\ &= 19,206 \text{ meses}\end{aligned}$$

160

Análise de Sobrevivência - 2017

√ Estimativa variância EMV's – Lognormal

```
> ajuste.ln$var
            (Intercept)  Log(scale)
(Intercept) 0.031061677 0.002706896
Log(scale)  0.002706896 0.030119031
```

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\mu}) &= 0,031061677 \\ \widehat{\mathrm{Var}}(\log\hat{\sigma}) &= 0,030119031 \\ \widehat{\mathrm{Cov}}(\hat{\mu}, \log\hat{\sigma}) &= 0,002706896\end{aligned}$$

√ Estimativa variância $\hat{\sigma}$:

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\mu}) &= 0,0311 \\ \widehat{\mathrm{Var}}(\hat{\sigma}) &= \widehat{\mathrm{Var}}(\ln\hat{\sigma})(\hat{\sigma})^2 \\ &= (0,0311)(0,7648)^2 \\ &= 0,0176 \\ \widehat{\mathrm{Cov}}(\hat{\mu}, \hat{\sigma}) &= 0,0027\end{aligned}$$

161

Análise de Sobrevivência - 2017

- Estimativa de Var[$\hat{E}$(T)]:
  - √ Modelo Log-normal:

```
> ajuste.ln$var
            (Intercept)  Log(scale)
(Intercept) 0.031061677 0.002706896
Log(scale)  0.002706896 0.030119031
```

$$\widehat{\mathrm{Var}}(\hat{\mu}) = 0,0311$$
$$\widehat{\mathrm{Var}}(\hat{\sigma}) = \widehat{\mathrm{Var}}(\ln\hat{\sigma})(\hat{\sigma})^2 = (0,0311)(0,7648)^2 = 0,0176$$
$$\widehat{\mathrm{Cov}}(\hat{\mu},\hat{\sigma}) = 0,0027$$

  - √ Estimativa :

$$\widehat{\mathrm{Var}}[\hat{E}(T)] = \widehat{\mathrm{Var}}(\hat{\mu})\left[\exp\left\{\hat{\mu}+\frac{\hat{\sigma}^2}{2}\right\}\right]^2 + \widehat{\mathrm{Var}}(\hat{\sigma})\left[\hat{\sigma}\exp\left\{\hat{\mu}+\frac{\hat{\sigma}^2}{2}\right\}\right]^2 + 2\widehat{\mathrm{Cov}}(\hat{\mu},\hat{\sigma})\left[\exp\left\{\hat{\mu}+\frac{\hat{\sigma}^2}{2}\right\}\right]\left[\hat{\sigma}\exp\left\{\hat{\mu}+\frac{\hat{\sigma}^2}{2}\right\}\right]$$
$$= (0,031)(20,263)^2 + (0,0176)[(0,76)(20,263)]^2 + 2(0,00207)(0,76)(20,263)^2 = 18,2$$

162

Análise de Sobrevivência - 2017

- Intervalo de 95% de confiança para E[T]:

$$\hat{E}[T] \pm z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}[\hat{E}(T)]} = 20,263 \pm 1,96\sqrt{18,2}$$
$$[11,901; 28,625]$$

163

Análise de Sobrevivência - 2017

- Estimativa de Var[$\hat{E}$(T)]:
  - √ Modelo de Weibull:
    - Nesse caso, a estimativa é mais complicada pois aparece a derivada da função gama
    - Envolve a função digama

164

Análise de Sobrevivência - 2017

- Estimativa para o tempo mediano:
  - √ Modelo Log-normal: $\hat{t}_{0,5} = \exp\{\hat{\mu}\} = \exp\{2,72\} = 15,18$ meses
  - √ Variância do estimador do tempo mediano

$$\mathrm{Var}(\hat{t}_{0,5}) = \mathrm{Var}(\mu)\,[\exp\{\hat{\mu}\}]^2$$
$$\widehat{\mathrm{ep}}(\hat{t}_{0,5}) = \sqrt{0,031}\exp\{2,72\} = 2,675 \text{ meses}$$

  - √ Intervalo com 95% de confiança para o tempo mediano

$$\hat{t}_{0,5} \pm z_{\alpha/2}\widehat{\mathrm{ep}}(\hat{t}_{0,5}) = 15,18 \pm (1,96)(2,675)$$
$$[9,94; 20,42]$$

165

Análise de Sobrevivência - 2017

- Estimativa para o tempo mediano:
  - √ Modelo de Weibull: $$\hat{t}_{0,5} = \frac{[\ln(2)]^{\frac{1}{\gamma}}}{\alpha} = \frac{[\ln(2)]^{\frac{1}{1,54}}}{0,0468} = 16,84 \text{ meses}$$
  - √ Interpolação linear da função de sobrevivência

	tempo.km	s.km
[1,]	3	0.95000000
[2,]	5	0.90000000
[3,]	6	0.85000000
[4,]	7	0.80000000
[5,]	8	0.75000000
[6,]	9	0.70000000
[7,]	10	0.65000000
[8,]	12	0.59583333
[9,]	15	0.54166667
[10,]	18	0.48148148
[11,]	19	0.42129630

$$\hat{t}_{0,5} = 17,05 \text{ meses}$$

166

Análise de Sobrevivência - 2017

- Estimativa para S(20):
  - √ Modelo Log-normal: $$\hat{S}_{ln}(20) = \Phi\left(\frac{-\ln(t) + \hat{\mu}}{\hat{\sigma}}\right) = \Phi\left(\frac{-\ln(20) + 2,72}{0,76}\right) = 0,3583$$
  - √ Estimador Kaplan-Meier: $\hat{S}_{KM}(20) = 0,361$
  - √ Paciente tem uma probabilidade de cerca de 36% de estar livre de reincidência 20 meses após da realização do procedimento cirúrgico

171

Análise de Sobrevivência - 2017

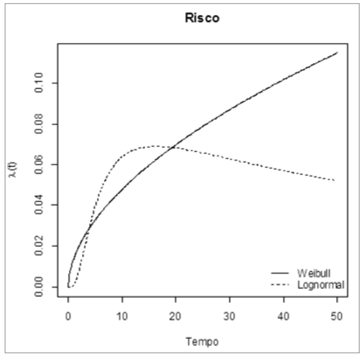
Comentário

- Utilizado o modelo lognormal para ilustrar o cálculo das estimativas intervalares de E(T) e t_{med}.

174

Análise de Sobrevivência - 2017

- Funções de risco estimadas dos modelos



  - √ Qual função de risco expressa melhor o problema?

175

Análise de Sobrevivência - 2017

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.