Aplicações de Análise de Sobrevivência na Área da Saúde

Lupércio França Bessegato Dep. Estatística/UFJF

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Material de Apoio

- R:
 - √ www.r-project.org
- Tutorial online do R:
 - √ http://www.leg.ufpr.br/Rtutorial
 - $\sqrt{\, \underline{http://www.leg.ufpr.br/\sim}paulojus/embrapa/Rembrapa}$
- Conjuntos de dados e material Análise de Sobrevivência Carvalho et al. (Fiocruz)
 - √ http://sobrevida.fiocruz.br

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Prof. Lupércio França Bessegato

Departamento de Estatística – UFJF

E-mail: lupercio.bessegato@ufjf.edu.br Site: www.ufjf.br/lupercio_bessegato

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Roteiro

- 1. Introdução
- 2. Características de Dados de Sobrevivência
- 3. Funções Básicas de Sobrevivência
- 4. Estimação Não Paramétrica
- 5. Modelos de Regressão de Cox
- Referências

Introdução

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Modelos Estatísticos

- Estruturas de causalidade fundamentas em raciocínio epidemiológico e evidências empíricas
- Associação de efeitos principais e de interação a cada variável
- Possibilidade de afirmações probabilísticas para descrição e representação do conhecimento
- Variabilidade associada a predições produzidas pelo modelo (IC's)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Modelo de Sobrevivência

- Classe de modelos estocásticos usados para analisar características e fatores associados ao tempo até ocorrência de evento de interesse
- Evento de interesse:
 - √ Falha ou desfecho
 - Óbito
 - Nascimento
 - Cura
 - Diagnóstico de uma doença
 - Aparecimento de tumor
 - Aparecimento de efeito colateral de medicamento

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Análise de Sobrevivência

- Tempo é o objeto de interesse:
 - √ Tempo até a ocorrência de evento de interesse
 - √ Risco de ocorrência por unidade de tempo
- Exemplos;
 - √ Qual o efeito de anticancerígeno sobre o tempo de sobrevivência?
 - $\sqrt{\mbox{Quais}}$ os fatores associados ao tempo de duração da amamentação?
 - √ Quais os fatores preditivos para reinternação hospitalar, considerando o tempo entre internações?

Modelo Estatístico de Sobrevivência

- Descrição da força de incidência como função de variáveis explicativas
 - √ Procura-se conhecer o efeito de cada uma delas e suas interações, livres da presença de confusão
 - à um modelo de regressão, com os seguintes elementos:
 - Variável resposta
 - Covariáveis explicativas
 - Função de ligação
 - Estrutura de erro

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Caracterização de Dados de Sobrevivência

59ª Reunião Anual da Região Brasileira da

Variável Resposta

- Probabilidade de sobrevivência:
 - $\sqrt{\text{Probabilidade}}$ de não ocorrer o evento de interesse no período [t_0 , t]
 - $\sqrt{P\{\text{sobreviver por mais que t unidades de tempo}\}}$
- Taxa de incidência (ou de *hazard*)
 - $\sqrt{\text{Taxa}}$ instantânea de ocorrência do evento de interesse no tempo t, condicionada à sua não ocorrência anterior à i.
- Taxa de incidência acumulada

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Exemplo

- Hemodiálise em pacientes com insuficiência renal
 - √ Registro mensal de sessões de hemodiálise
 - √ Período: Janeiro/1988 a Outubro/2001
 - $\sqrt{\text{Evento de interesse:}}$
 - Óbito
 - √ Variável tempo:
 - Intervalo entre o início tratamento e o óbito
 - √ Dados: *dialise.csv*

- Informações parciais
 - √ Causas de perda da informação temporal completa
 - Óbitos por causas não relacionadas ao estudo
 - Término do estudo
 - Perda de contato
 - Recusa em continuar participando
 - Mudança de procedimento
 - Abandono devido a efeitos adversos de tratamento
 - Desconhecimento do início do tempo da modelagem

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Tipos de Censura

- · Censura à direita
- Censura à esquerda
- Censura intervalar

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Censura

- Perda de informação decorrente de não se ter observado a data de ocorrência do desfecho
 - √ Mesmos censurados, todos os resultados de um estudo de sobrevivência devem ser usados na análise
 - Mesmo incompletas, as observações censuradas fornecem informações sobre o tempo de vida de pacientes
 - Omissão das censuras pode acarretar conclusões viciadas

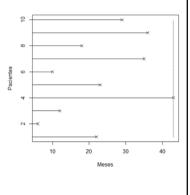
59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Censura à Direita

- Tempo de ocorrência do desfecho está à direita do tempo registrado
 - √ Aproveita-se a informação do tempo durante o qual a pessoa esteve sob observação sem que ocorresse o evento

• Tempos de observação (em meses) de 10 pacientes em diálise

Paciente (i)	Tempo (T _i)	Status (δ_i)
1	22	1
2	6	0
3	12	1
4	43	0
5	23	1
6	10	1
7	35	1
8	18	0
9	36	1
10	29	1



59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Exemplo

- Estudo para investigar os fatores associados à infecção por leptospirose em comunidade de baixa renda, recém-criada.
 - $\sqrt{\text{Inclus}}$ ão de participantes quando de mudança para a área
 - √ Anualmente coleta-se sangue dos participantes para verificar se houve soroconversão

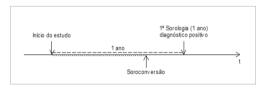
59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Censura à Esquerda

- Não se conhece o momento da ocorrência do desfecho, mas sabemos que ele ocorreu antes do tempo observado
 - √ Tempo observado é maior que o tempo de falha

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Censura à Esquerda



- √ Alguns participantes já estavam soropositivos logo na 1ª coleta
 - Tempo para soroconversão (tempo de sobrevivência) é menor que o tempo para o primeiro exame

Censura Intervalar

- Tipo mais geral de censura
 - √ Ex.: estudo em que os pacientes são acompanhados em visitas periódicas e conhece-se apenas que o evento de interesse ocorreu em um certo intervalo de tempo
- Representação: $(t^- < T < t^+)$

 \sqrt{t} e t⁺: tempos observados

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Truncamento

- É caracterizado por uma condição que exclui certos indivíduos do estudo
 - √ Indivíduos são excluídos por motivo relacionado à ocorrência do estudo
- Os pacientes não são acompanhados a partir do tempo inicial, mas somente após experimentarem um certo evento
 - $\sqrt{\rm Estudo}$ inclui somente indivíduos em que o evento ocorreu dentro de uma janela temporal préestabelecida ($T_{\rm E},\,T_{\rm D}$)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Exemplo

• Estudo de leptospirose



- √ Participante em que a soroconversão foi observada no segundo exame
 - Tempo exato de sobrevivência é maior que o tempo até o primeiro exame e menor do que o tempo até o segundo exame
 - $-(t^- < T < t^+)$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Truncamento à Esquerda

- Inclui somente observações em que o desfecho ocorreu após o limite inferior da janela temporal de observação
 - $\sqrt{T} \ge T_E$
 - √ Só ocorre quando a perda de informação está relacionada a indivíduos que foram excluídos porque já tinham experimentado o evento antes do início do evento e não podiam ser observados

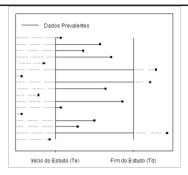
Exemplo

- Hemodiálise em pacientes com insuficência renal
 - √ Registro mensal de sessões de hemodiálise
 - √ Período: Janeiro/1988 a Outubro/2001
 - $\sqrt{\text{Ingresso no estudo:}}$
 - Pacientes em tratamento em janeiro de 1988 ou que iniciaram o tratamento numa data posterior
 - √ Dados truncados:
 - Pacientes diagnosticados e mortos antes de Janeiro/1988
 - √Truncamento à esquerda devido a uso de dados prevalentes

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Viés de seleção:
 - √ Indivíduos com sobrevivência mais longa têm maior probabilidade de entrar no estudo (subestima risco de ocorrência do evento)
- · Correção do viés:
 - √ Pacientes que iniciaram a diálise antes do início do estudo só terão considerados os tempos após a data de início do estudo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



- Parte do tempo é observado prospectivamente e parte retrospectivamente
 - √ Informações dos indivíduos que já sofreram o desfecho antes o início do estudo não podem ser incluídos no estudo, embora se esteja considerando os sobreviventes

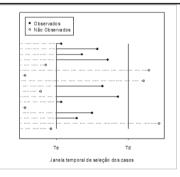
59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Interpretação adequada da modelagem com dados prevalentes:
 - √ Resultados encontrados estão condicionados ao fato de o paciente estar vivo em Janeiro/1988

Truncamento à Direita

- Ocorre quando o critério de seleção inclui somente os indivíduos que sofreram o evento
 - $\sqrt{\mbox{Data}}$ de ocorrência do evento é sempre menor que o limite superior da janela temporal (T \leq T_D)
 - √ Comum em estudos de sobrevivência que partem do dado do óbito para selecionar as observações

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



- √ Amostra definida pelos eventos na janela temporal
- √ Trajetórias em pontilhado não são observadas (mesmo que cruzem a janela evento não ocorreu)
- √ Nunca há censura à direita

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Exemplo

- Análise para estimar fatores associados à sobrevivência de pacientes com Aids que partem do registro de óbito
 - $\sqrt{\text{Indiv}(\text{duos que não sofrem o evento não são}}$ incluídos no estudo, mesmo tendo todos os fatores de risco presentes.

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Viés de seleção:
 - √ Risco de ocorrência do evento é superestimado
 - √Em doenças de curta duração e muito bem caracterizadas este viés de seleção será pouco importante

Coortes Abertas

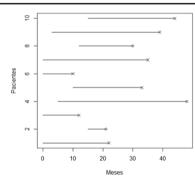
- Representação:

 $\sqrt{(I_i, F_i, \delta_i)}$: informação sobrevivência indivíduo i

- Ii: tempo calendário de início de observação
- Fi: tempo calendário de término de observação

$$\sqrt{(I_5, F_5, \delta_5)} = (10, 33, 1) \rightarrow (T_i, \delta_i) = (23, 1)$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



- Eixo das abcissas:
 - √ Tempo calendário
- Ponto inicial:

√ Data do 1º registro ou data de início de estudo prospectivo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Exemplo: Pacientes em hemodiálise

Paciente (i)	Tempo inicial	Tempo final	Tempo (T _i)	$\begin{array}{c} Status \\ (\delta_i) \end{array}$
1	0	22	22	1
2	15	21	6	0
3	0	12	12	1
4	5	48	43	0
5	10	33	23	1
6	0	10	10	1
7	0	35	35	1
8	12	30	18	0
9	3	39	36	1
10	15	44	29	1

Dados em formato de processo de contagem

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Tempo de Sobrevivência no R

library(survival)
 Surv(tempo, status)
 Surv(inicio, fim, status)

- Comando Surv()
 - √ Combina, em uma única variável as informações sobre o tempo de sobrevivência e o status de cada indivíduo
 - Status = 1, se ocorreu o evento de interesse
 - Status = 0, se o tempo foi censurado

Formato clássico

 $\sqrt{ }$ + ao lado do tempo de observação indica censura

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Funções Básicas de Sobrevivência

59ª Reunião Anual da Região Brasileira da

• Formato de processo de contagem

√ + ao lado da data de saída da coorte indica censura √ Tempo início da observação não faz parte do intervalo √ Último tempo é parte do intervalo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Introdução

- Exemplo: Sobrevivência ao diagnóstico de Aids
 - √ Coorte de 50 pacientes acompanhados por 4 anos (1460 dias)
 - √ Observados 32 óbitos
 - √ Objetivo: descrição do processo
 - Taxa de mortalidade média: 32/50 = 64%
 - Taxa de mortalidade anual: (0,64)(100)/4 = 16 óbitos por 100 pessoas-ano
 - √ Taxa não é homogênea no tempo
 - A probabilidade de um paciente vir a morrer de Aids é função que depende do tempo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Questões em Análise de Sobrevivência

- √ Qual o **risco** de um paciente com Aids vir a falecer em até **três anos** após o diagnóstico?
- √ Qual a **probabilidade** de um paciente sobreviver por mais de 2 anos após o diagnóstico?
- √ Qual o **número esperado** de óbitos em uma coorte de pacientes acompanhados por 5 anos?
- √ Qual o **tempo mediano** de sobrevivência?

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Função de Sobrevivência

- Qual a probabilidade de um paciente sobreviver por mais de 365 dias?
 - $\sqrt{\text{Oual \'e o valor de P}\{T > 365}\}$?
- *S*(*t*): função de sobrevivência

$$S(t) = P\{T > t\}$$

= 1 - P\{T \le t\} = 1 - F(t)

 $\sqrt{F(t)}$: função de distribuição acumulado de T

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Função de Densidade de Probabilidade

- T: tempo de sobrevivência (tempo até desfecho)
 √ Variável aleatória contínua e positiva
- f: função de densidade de probabilidade de T

$$f(t) = \lim_{\epsilon \to 0^+} \frac{P\{t \le T \le t + \epsilon\}}{\epsilon}$$

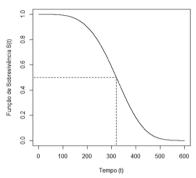
√ Pode ser interpretada como a probabilidade de um indivíduo sofrer um evento em um intervalo instantâneo de tempo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Propriedades

- S(0) = 1
- $S(+\infty)=0$
- S(t) é não crescente
- A partir da função de sobrevivência S(t) é possível calcular os percentis de T.

• Curva teórica da função de sobrevivência:



√ Tempo de sobrevivência mediano: 319,4

$$\sqrt{S(319.4)} = 0.5$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Função de Risco

- Qual o risco de um paciente com Aids vir a óbito após sobreviver 365 dias?
 - $\sqrt{\text{Esse risco}}$ de morrer aumenta ou diminui com o tempo?
- $\lambda(t)$: risco instantâneo de um indivíduo sofrer o evento entre o tempo t e $t+\epsilon$, dado que ele sobreviveu até o tempo t
 - $\sqrt{\varepsilon}$ é infinitamente pequeno \Rightarrow $\lambda(t)$ expressa risco instantâneo de ocorrência de uma evento, dado que até então o evento não tenha ocorrido

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimação Função Sobrevivência – Sem Censura

• Estimação de S(t) na ausência de censura:

$$\hat{S}_x(t) = \frac{R_x(t)}{\text{n}^o \text{ total de pacientes}}$$

 $\sqrt{R_x(t)}$: número de pessoas sob risco no início do intervalo de tempo x.

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

$$\lambda(t) = \lim_{\epsilon \to 0^+} \frac{P\{t \le T \le t + \epsilon | T \ge t\}}{\epsilon}$$

 $\sqrt{\lambda(t)}$ também é denominada:

- função ou taxa de incidência,
- força de infecção,
- taxa de falha,
- força de mortalidade,
- força de mortalidade condicional.

 $\sqrt{\lambda(t)}$ é uma taxa (tempo⁻¹).

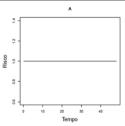
√ Pode assumir qualquer valor positivo

– (não é probabilidade)

$$\lambda(t) = \frac{f(t)}{S(t)}$$
 $\lambda(t) = -\frac{\mathrm{d}\ln(S(t))}{\mathrm{d}t}$

- $\lambda(t)$ e S(t) são inversamente proporcionais
 - √ Quando o risco aumenta, a probabilidade de sobrevivência diminui e vice-versa

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



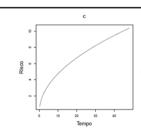
- · Curva de risco constante
 - √ Ex: Ocorrência de fratura em população escolar em 12 meses
 - (caso fatores de exposição a risco forem os mesmos)
- A ocorrência do evento é sem memória
- · Curva decrescente de risco
 - √ Ex.: Óbito pós-cirúrgico
 - (risco é maior no primeiros dias, caindo rapidamente a zero com o passar do tempo)

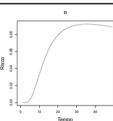
598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Comportamento da Função de Risco

- Permite avaliar o risco de um indivíduo sofrer um evento em um determinado tempo *t*, dado que ele já sobreviveu até aquele momento
- Qual é o risco de um paciente falecer após uma cirurgia?
 - √ O risco de vir a óbito no 1º dia pós-cirúrgico é o mesmo de vir a óbito no 30º dia?

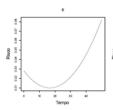
598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014





- Curva de risco crescente
 - √ Ex: Estudo sobre o tempo até solidificação de fratura óssea
 - Desfecho: solidificação de fratura
 - Risco nos primeiros dias é nulo até que praticamente todas as fraturas estejam solidificadas
- Exemplo de possível descrição de metástase
 - √ Período inical: risco é baixo e não varia (quimioterapia)
 - √ Aumenta em seguida
 - √ Após um tempo fica aproximadamente constante até decair novamente

Análise de Sobrevivência - Minicurso



- Comportamento do risco:
 - √ Risco começa alto
 - √ Decai durante certo tempo
 - √ Volta a subir

- · Comportamento do risco?
 - √ Sobe rapidamente no início
 - √ Cai rapidamente depois
 - √ Exemplo: comportamento possível da tuberculose (após introdução dos tuberculostáticos, risco de óbito decai rapidamente)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Função de Risco Acumulado

 Qual é o risco de um paciente com Aids vir a óbito no 1º ano após diagnóstico? E nos primeiros dois anos?

$$\Lambda(t) = \int_0^t \lambda(u) \, \mathrm{d}u$$

 $\sqrt{\Lambda(t)}$: função de risco acumulado

- mede o risco de ocorrência do evento até o tempo t
- é uma taxa e não está restrito ao intervalo [0,1]

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Relação entre Funções de Sobrevivência

- As funções f(t), S(t), $\lambda(t)$, e $\Lambda(t)$ são equivalentes $\sqrt{}$ Diferentes formas de representar o mesmo fenômeno
- Respondem a três perguntas basicas:
 - $\sqrt{\text{Qual a probabilidade de sobreviver por mais de } t}$ unidades de tempo?
 - $\sqrt{\text{Qual o risco}}$ de sofrer o evento no tempo t se sabemos que o paciente sobreviveu até aquele momento?
 - $\sqrt{\text{Qual o risco de sofrer o evento até um determinado tempo } t$?

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

 Relações entre as funções básicas: de sobrevivência:

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

$$S(t) = \exp(-\Lambda(t))$$

$$\lambda(t) = -\frac{d \ln(S(T))}{dt}$$

$$\lambda(t) = \frac{f(t)}{S(t)}$$

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

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

Estimação Não Paramétrica

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimador de Kaplan-Meier

- Desdobra a condição de sobreviver até o tempo t
 - $\sqrt{\text{Sequência}}$ de elementos independentes, caracterizando a sobrevivência em cada intervalo de tempo anterior a t
 - √ A probabilidade de sobreviver ao tempo t é o produto das probabilidades de chegar até cada um dos tempos anteriores

$$S(5) = P\{T {>} 5\} = P\{T {>} 1;\, T {>} 5\} = P\{T {>} 1\}\; P\{T {>} 5\mid T {>} 1\}$$

- Sobrevivência por 5 semanas: necessário sobreviver à 1^a semana e depois, à 5^a semana, sabendo-se que sobreviveu à 1^a
- √ Estimador produto (ou estimador limite produto)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimadores Não Paramétricos

- Abordagem não paramétrica:
 - √ Não adota qualquer suposição sobre a distribuição de probabilidades do tempo de sobrevivência (T)
- Estimadores:
 - √ Kaplan-Meier:
 - Estima a função de sobrevivência S(t)
 - √ Nelson-Aalen:
 - Estima a função de risco acumulado $\Lambda(t)$
 - √ Ambos utilizam dados censurados

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimador Produto

• Amostra com *n* indivíduos e *m* diferentes tempos de ocorrência dos eventos:

$$\sqrt{t_1 < t_2 < \dots < t_m}$$

- $\sqrt{R(t_i)}$: total de pessoas sob risco no tempo tj
- $\sqrt{\Delta N(t_j)}$: número de eventos ocorridos precisamente em t_i
- $\sqrt{R(t_i)} \Delta N(t_i)$: número de sobreviventes no tempo t_i

P {sobrevivência ao tempo
$$t_j \} = \frac{R(t_j) - \Delta N(t_j)}{R(t_j)}$$

√ Como os eventos são independentes S(t) é o produto da probabilidades de sobrevivência a cada tempo $t_i \le t$

$$\hat{S}_{KM}(t) = \left(\frac{R(t_1) - \Delta N(t_1)}{R(t_1)}\right) \times \left(\frac{R(t_2) - \Delta N(t_2)}{R(t_2)}\right) \times \dots$$

$$\left(\frac{R(t_m) - \Delta N(t_m)}{R(t_m)}\right)$$

$$\hat{S}_{KM}(t) = \prod_{t_j \le t} \left(\frac{R(t_m) - \Delta N(t_m)}{R(t_m)} \right)$$

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

√Função de sobrevivência estimada pelo método Kaplan-Meier

```
> sobrev.km<-survfit( formula =
> summary(sobrev.km)
Call: survfit(formula = Surv(te
time n.risk n.event survival s
 16 21 1 0.9524
 18
      20
             1 0.9048
      19
 21
             1 0.8571
 22
     17
            1 0.8067
 29
     15
           1 0.7529
 35
     14
           1 0.6992
             1 0.6454
 39
      12
                0.5916
 40
      11
             1 0.5378
             1 0.4781
 54
             1 0.4183
 60
             1 0.3585
                 0.2988
 83
                 0.1992
                0.0996
```

```
\hat{S}_{KM}(18) = 0,9524 \times \left(\frac{20-1}{20}\right) = 0,9048
```

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

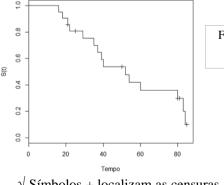
Exemplo

- Sobrevivência ao diagnóstico de Aids
 - √ Pacientes com tempo de observação menor que 90 dias
 - Óbitos: 15
 - Censuras: 6
 - $\sqrt{\text{Banco de dados: ipec.csv}}$

```
> ipec90 <- ipec[ipec$tempo < 90,]</pre>
> Surv(ipec90$tempo, ipec90$status)
[1] 60 84 25+ 54 80+ 37 18 29 50+ 83 80 81+ 35 52 21 40 22 85+ 39
[20] 16 21+
```

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Função de sobrevivência dos pacientes com Aids



Função escada: Salta em cada tempo de ocorrência de evento

√ Símbolos + localizam as censuras

Estimador de Kaplan-Meier para $\Lambda(t)$

• Função de risco acumulado:

$$\hat{\Lambda}_{KM}(t) = -\ln \hat{S}_{KM}(t)$$

√ Pode-se estimar qualquer das funções

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimador de Nelson-Aalen

$$\hat{\Lambda}_{NA}(t) = \sum_{t_j \le t} \frac{\Delta N(t_j)}{R(t_j)}$$

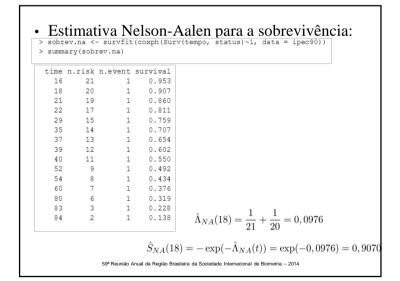
√ Pode-se estimar a sobrevivência por:

$$\hat{S}_{NA}(t) = \exp\left(-\hat{\Lambda}_{NA}(t)\right)$$

- √ Indicado para amostras muito pequenas
- √ Para amostras suficientemente grandes as estimativas de K-M e N-A para a função de sobrevivência são bem próximas

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Gráfico da função de risco acumulado $\sqrt{\text{Estimativa por Kaplan-Meier}} > \text{plot}(\text{sobrev.km, conf.int} = \text{F, xlab} = \text{"Tempo", ylab} = \text{"S(t)", fun = "cumhaz"})}$ $\hat{\Lambda}_{KM}(21) = -\ln(\hat{S}_{KM}(21)) = -\ln(0,8571) = 0,1542$ S9º Reunião Anual da Região Brasileira da Sociedade Internacional de Blometria - 2014



Intervalos de Confiança

Variância do estimador Kaplan-Meier

 √ Estimador de Greenwood:

$$\widehat{\mathrm{Var}}(\hat{S}_{KM}(t)) = (\hat{S}_{KM}(t))^2 \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)(R(t_j) - \Delta N(t_j))}$$

 $\sqrt{\acute{E}}$ uma expressão assintótica

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Alternativa para contornar incompatibilidade em limites:
 - $\sqrt{\text{Construir intervalo simétrico para ln}(\Lambda(t))}$
 - $\sqrt{\mbox{Construir}}$ intervalo para S(t) através da transformação $\ln(\Lambda(t)) = \ln(-\ln(S(t))$
 - $\sqrt{\text{Intervalo}}$ será assimétrico, porém positivo e menor ou igual a 1
- Estimador da variância assintótica de $ln(\Lambda(t))$

$$\widehat{\mathrm{Var}}[\ln(\hat{\Lambda}_{KM}(t))] = \frac{1}{[\ln(\hat{S}_{KM}(t))]^2} \sum_{t_j \leq t} \frac{\Delta N(t_j)}{R(t_j)(R(t_j) - \Delta N(t_j))}$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Intervalo de confiança para S(t):
 - √O estimador KM, para t fixo, tem distribuição assintótica normal
 - $\sqrt{\text{Intervalo aproximado de }100\;(1-\alpha)\%}$ de confiança para S(t)

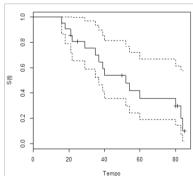
$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{S}_{KM}(t))}$$

√ Este intervalo levar a limites incompatíveis com a definição de sobrevivência:

(permite valores negativos e maiores do que 1)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

 $\sqrt{\text{Curva}}$ de sobrevivência estimada pelo Kaplan-Meier com intervalo de confiança de 95%



√ Estimativas do intervalo de confiança exponencial de Greenwood

Tempo de Sobrevivência Mediano

- · Tempo mediano:
 - √ Tempo após o qual 50% dos indivíduos estão vivos
 - √ Devido a sua robustez, é a medida-resumo mais comum em Análise de Sobrevivência
- Estimação do tempo mediano:
 - √ A sobrevivência é uma função em degraus (O tempo não é observado de forma contínua)

$$\hat{t}_{med} = \min\{t_j | \hat{S}(t_j) \le 0, 5\}$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Exemplo √ Coorte completa de Aids: - Óbitos: 90 e censuras: 103 - Óbitos: 90 e censuras: 103 √ Esta é uma boa estimativa da curva de sobrevivência tanto para homens quanto para mulheres? - S9º Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria - 2014

Estratificação

- Peculiaridades dos indivíduos contribuem para diferentes curvas de sobrevivências dos grupos
 - √ Pessoas não adoecem ou morrem da mesma forma
 - √ Descrição da sobrevivência segundo característica:
 - Sexo, faixa etária, etc.
 - √ Divide-se o conjunto total de observações em estratos de acordo com covariável de interesse
 - √ Utilização de Kaplan-Meier para estimação da função de sobrevivência para cada estrato

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Comandos no R:

√ Carregamento dados:

> library(survival)
> ipec <- read.table("dados/ipec.csv", header = T, sep = ";")</pre>

√ Gráfico coorte completa:

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Estratificação por sexo:

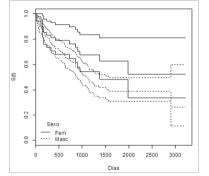
> survaids <- survfit(Surv(tempo, status)~sexo, data = ipec)
> survaids
Call: survfit(formula = Surv(tempo, status) ~ sexo, data = ipec)

records n.max n.start events median 0.95LCL 0.95UCL
sexo=F 49 49 49 16 NA 1371 NA
sexo=M 144 144 144 74 1116 887 1563

- √ Há predominância do sexo masculino (75%) e 82% das mortes se concentram neste grupo
- √ Não foi possível estimar o tempo de sobrevivência mediano para mulheres (67% estão vivas)
- √ Nesse caso é comum utilizar outro percentil:
 - Percentil 70: 905 (F) e 516 (M)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

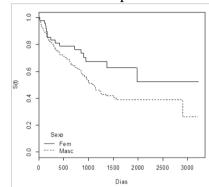
• Curvas de sobrevivência com IC de 95%



- $\sqrt{\dot{E}}$ visível a sobreposição das bandas de confiança
 - Sugere que a diferença observada não é significativa

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Curvas de sobrevivência por sexo



√ Curva de sobrevivência dos homens sempre abaixo

– A diferença é significativa?

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Comparação das Curvas

- Testes de hipóteses:
 - $\sqrt{\text{Log-Rank ou Mantel-Haenszel}}$
 - √ Peto
- Hipótese nula: não há diferença entre estratos

$$\sqrt{\mathrm{H}_0}$$
: $\lambda_1(t) = \lambda_2(t) = \dots = \lambda_k(t)$

Teste Log-rank

Distribuição esperada de eventos
 √ Sob H₀ é igual em todos os estratos:

$$E_k(t) = N(t) \frac{R_k(t)}{R(t)}$$

• Estatística de teste para 2 estratos (k = 2)

$$T_{LR} = \frac{(O_1 - E_1)^2}{\text{Var}(O_1 - E_1)}$$

 $\sqrt{O_1}$: total de eventos observados no estrato 1

 $\sqrt{E_1}$: total de eventos esperados no estrato 1

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Teste de Peto

 Maior peso às diferenças (ou semelhanças) no início da curva

 $\sqrt{\text{Local de concentração da maior parte dos dados}}$

• Estatística de teste: $T_P = \frac{(O_1 - E_1)^2}{\text{Var}(O_1 - E_1)}$

com
$$Var(O_1 - E_1) = \sum_{t_j} S(t_j)(O_1(t_j - E_1(t_j))$$

 $\sqrt{\text{Usa um ponderador S(t) no estimador}}$

• Distribuição amostral

$$\sqrt{\text{Sob H}_0}$$
: $T_P \stackrel{as.}{\sim} \chi_{k-1}^2$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Fator de padronização da estatística (k = 2)

$$Var(O_1 - E_1) = \sum_{t} \frac{R_1(t)R_2(t)\Delta N(t)[R(t) - \Delta N(t)]}{R^2(t)[R(t) - 1]}$$

Distribuição amostral da estatística (caso geral)
 √ Sob H₀

 $T_{LK} \stackrel{as.}{\sim} \chi^2_{k-1}$

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 20

Exemplo

 $\sqrt{\text{Use rho}} = 0$ (default)

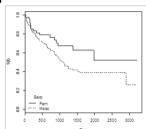
Modelo de Regressão de Cox

59ª Reunião Anual da Região Brasileira da

Comentários

Resultados

Teste	Estatística	p-valor
Log-rank	4,03	0,0447
Peto	3,54	0,0598



 $\sqrt{\text{Teste Log-rank rejeita H}_0}$ (ainda que a diferença seja pequena)

- √ Inspeção gráfica:
 - Diferença nas curvas é maior mais tardiamente
 - Teste Peto não detecta essa diferença (mais peso para os tempos menores)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Introdução

- Interesse:
 - √ Modelar o efeito de covariáveis sobre o tempo de sobrevivência (risco)
- Modelo de riscos proporcionais de Cox
 - $\sqrt{\text{Modelo}}$ de regressão mais utilizado em Sobrevivência
 - √ Efeito multiplicativo das covariáveis
- Modelo de Cox estendido
 - √ Modelagem de situações complexas com base em processo de contagem

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Modelo de Riscos Proporcionais

- Ajusta a função de risco λ(t)
 - $\sqrt{\text{Considera um risco basal } \lambda_0(t)}$
 - $\sqrt{\text{Inclui o vetor de covariáveis } x}$

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) = \lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})$$

 Razão entre o risco de ocorrência do evento para dois indivíduos k e l, com covariáveis x_k e x_l

$$\frac{\lambda_k(t|\mathbf{x}_k)}{\lambda_l(t|\mathbf{x}_l)} = \frac{\exp(\mathbf{x}_k'\boldsymbol{\beta})}{\exp(\mathbf{x}_l'\boldsymbol{\beta})}$$

√ Razão de riscos NÃO varia ao longo do tempo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Outras formas do modelo de riscos proporcionais:

$$\Lambda(t|\mathbf{x}) = \Lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})
S(t|\mathbf{x}) = [S_0(t)]^{\exp(\mathbf{x}'\boldsymbol{\beta})}$$

$$\sqrt{\text{Risco acumulado basal}} \quad \hat{\Lambda}_0(t) = \sum\nolimits_{i:t_i \leq t} \frac{\Delta N_i(t)}{\sum_{j \in R(t_i)} \exp(\mathbf{x}'\hat{\boldsymbol{\beta}})}$$

$$\sqrt{\mathrm{Sobreviv\hat{e}ncia}}$$
 basal $\hat{S}_0(t) = \exp\left[\hat{\Lambda}_0(t) \right]$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Riscos proporcionais:
 - $\sqrt{\rm O}$ risco de desenvolvimento do câncer ao longo do tempo para quem fuma é sempre o mesmo ao longo do tempo
 - √ Algumas pessoas vão desenvolver rapidamente e outras mais tarde, mas sempre na mesma proporção estimada

(Exponencial do coeficiente [exp $\{\beta_{Fumo}\}\]$)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Modelo de Cox é dito semi-paramétrico:
 - √ Supõe que as covariáveis agem multiplicativamente (parte paramétrica do modelo)
 - $\sqrt{\text{N}}$ ão assume distribuição para o tempo de sobrevivência na estimação dos efeitos das covariáveis e na função de risco basal $\lambda_0(t)$ (parte não paramétrica do modelo)

Modelo de Cox - Suposições

- Covariáveis agem multiplicativamente sobre o risco
 - √ Parte paramétrica do modelo
- Razão de risco é constante ao longo do tempo √ Riscos proporcionais
- Os tempos de ocorrência do evento são independentes
- Não há empates na ocorrência do evento √ Tempo é contínuo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Censura à direita:
 - √ Sabemos apenas que o tempo de sobrevivência exato é maior que o observado
 - (contribuição da observação censurada à direita é dada pela sobrevivência)

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

- O: conjunto das observações que sofreram o evento
- D: conjunto das observações censuradas à direita

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Estimação dos Coeficientes

- Modelos de regressão paramétricos:
 - √ Determinação da função de verossimilhança deve considerar a situação das observações
 - Sem censura
 - Censura à direita
 - Censura à esquerda
 - Censura intervalar
- Sem censura: $L \propto \prod_i f(t_i)$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Censura à esquerda:
 - √ Sabemos apenas que o tempo de sobrevivência exato é menor que o observado

(contribuição da observação censurada é dada pela função de distribuição acumulada)

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

- O: conjunto das observações que sofreram o evento
- D: conjunto das observações censuradas à direita
- E: conjunto das observações censuradas à esquerda

• Censura intervalar:

√ Sabemos a probabilidade de ocorrência do evento dentro de um intervalo de tempo t⁻ e t⁺

(contribuição da observação censurada é dada pela função de distribuição acumulada)

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

- O: conjunto das observações que sofreram o evento
- D: conjunto das observações censuradas à direita
- E: conjunto das observações censuradas à esquerda
- I: conjunto das observações com censura intervalar

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• 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

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

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

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:

$$\sqrt{(t_{\rm E}, t_{\rm D})}$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

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

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

Estimação dos Coeficientes

- Modelo de regressão de Cox:
 - √ Determinação da função de verossimilhança parcial:
 - Elimina-se a função de risco basal
 - Considera a informação dos indivíduos sob risco em cada tempo t
 - Formulação semelhante aos modelos não paramétricos (Kaplan-Meier), mas que permite a estimação dos efeitos das covariáveis

(fatores de risco no tempo de sobrevivência)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Definição da verossimilhança parcial em notação de processo de contagem:

$$L_i = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\sum_{t\geq 0} Y_j(t) \exp(\mathbf{x}_j'\boldsymbol{\beta})}$$

- $\sqrt{Y_j(t)}$: indicadora de risco do indivíduo j em t (1 se o indivíduo j estiver em risco no tempo t)
- √ A função de verossimilhança não depende do risco basal

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Verossimilhança Parcial

 $\sqrt{\text{Sejam }} t_1 < t_2 < ... < t_m$, m tempos diferentes ordenados (sem empate)

 $\sqrt{L_i}$: verossimilhança individual (contribuição de cada indivíduo para o tempo de sobrevivência t_i)

$$L_i = \frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j'\boldsymbol{\beta})}$$

 Razão entre o risco do indivíduo i apresentar o evento em ti e a soma dos riscos até a ocorrência do evento e, todos os indivíduos em risco

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Verossimilhança parcial $L(\beta)$ = produto das L_i

$$L(\boldsymbol{\beta} = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{j} Y_j(t) \exp(\mathbf{x}_j' \boldsymbol{\beta})} \right\}^{\Delta N_i(t)}$$

- $\sqrt{\Delta N_i(t)}$: diferença entre a contagem de eventos até o instante t e a contagem no momento imediatamente anterior a t
- √ Numerador: indivíduos que experimentaram o evento
- √ Denominador: todos os indivíduos que não experimentaram o evento (inclusive as censuras)

Exemplo

- Transplante de medula óssea
 - √ Objetivo: avaliação dos fatores prognósticos associados a transplante de medula óssea
 - √ Pacientes: 96

 Óbitos: 49

 Censuras: 47
 - $\sqrt{\text{Banco de dados: } tmoclas.dat}$

```
> library(survival)
> tmo <- read.table("dados/tmoclas.dat", header = T, sep = " ")
> tmo.tot = survfit(formula = Surv(os, status) ~1, data = tmo)
> tmo.tot

Call: survfit(formula = Surv(os, status) ~ 1, data = tmo)

records n.max n.start events median 0.95LCL 0.95UCL
96 96 96 49 453 370 NA
```

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Análise Exploratória

- Curvas de sobrevivência estimadas pelo método de Kaplan-Meier:
 - √ Objetivo:
 - Verificar o pressuposto de proporcionalidade dos riscos das variáveis a serem incluídas no modelo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Resposta:

√ os: tempo de sobrevivência desde o transplante até o óbito

√ status: 0: censura; 1: óbito

• Covariáveis:

 $\sqrt{\text{sexo: 1: masculino; 2: feminino}}$

 $\sqrt{\text{idade: idade na data do transplante (anos)}}$

 $\sqrt{}$ fase: fase da doença no momento do transplante

(1: crônica; 2: aguda; 3: crise blástica)

√ deag: doença enxerto contra hospedeiro aguda

(0: não; 1: sim)

√ decr: doença enxerto contra hospedeiro crônica

(0: não; 1: sim)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Covariáveis com pequeno número de categorias:
 - $\sqrt{\text{Gráfico}}$ da sobrevivência estratificado poderá dar indicação de proporcionalidade
 - Curvas que são razoavelmente paralelas ao longo de todo o tempo

(podem indicar proporcionalidade)

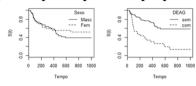
 Curvas com entrelaçamento ou variação na distância entre as curvas das categorias

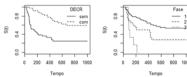
(podem indicar ausência de proporcionalidade)

- √ Outra forma de visualização
 - Gráfico do logaritmo da função de risco acumulado em cada categoria

(duas retas paralelas podem indicar proporcionalidade)

• Avaliação do pressuposto de proporcionalidade





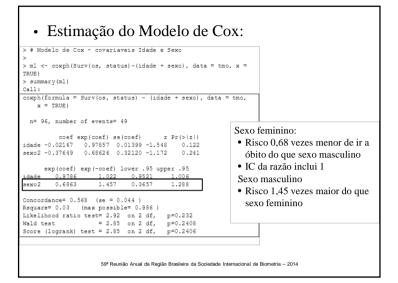
- √ Parece razoável admitir proporcionalidade
 - Variável sexo apresenta superposição, afastando-se no final

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Interpretação dos Coeficientes

- · Coeficiente:
 - √ Valores positivos:
 - Variáveis que contribuem para o aumento do risco
 - √ Valores negativos:
 - Variáveis que contribuem para a redução do risco
- Exponencial do coeficiente
 - √ Risco relativo ou razão de riscos (*hazard ratio*)
 - Maiores que 1: sobrerisco
 - Entre 0 e 1: proteção

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



```
exp(coef) exp(-coef) lower .95 upper .95 idade 0.9786 1.022 0.9521 1.006 sexo2 0.6863 1.457 0.3657 1.288
```

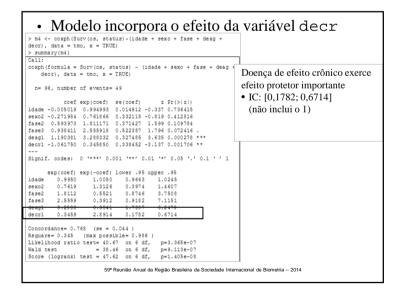
- No exemplo:
 - $\sqrt{\text{Paciente do sexo feminino tem risco 0,68 vezes (ou 32%)}$ menor de ir a óbito do que do sexo masculino
 - IC da razão de riscos: [0,365; 1,288]
 - √ Paciente do sexo masculino tem risco 1,45 vezes (ou 45%) maior de ir a óbito a cada unidade de tempo
 - IC da razão de riscos: [1/1,29; 1/0,37] = [0,78; 2,74]
 - √ Estimativa de risco atribuído ao sexo e idade não é significativo
 - Intervalo de confiança engloba a unidade

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Modelo de Cox Estratificado

- $\lambda_0(t)$ não é o mesmo para todos os indivíduos $\sqrt{\lambda_{0A}(t)} \neq \lambda_{0B}(t) \neq \lambda_{0C}(t)$, definindo diferentes estratos
- Usado quando alguma covariável não atende à proporcionalidade
- Não será estimado o efeito da variável para a qual se estratifica

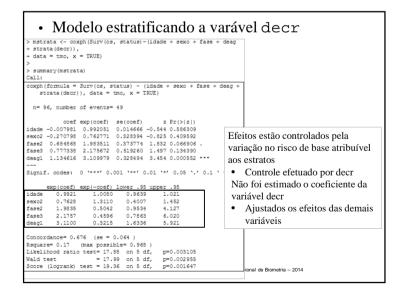
598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



Exemplo

- Modelagem da sobrevivência dos pacientes
 - √ Consideração:
 - aparecimento de doença de enxerto crônica divide em estratos com linhas de base diferentes
 - √ Raciocínio:
 - Considerar que para que ocorra doença crônica, o paciente necessariamente sobreviveu algum tempo
 - (efeito da variável varia com o tempo)
 - √ Não é possível estimar o efeito para a variável que estratifica o risco

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



Comentários

- √ Coeficientes estimados são semelhantes para os modelos m4 e mstrata
 - Em geral, os ICs em mstrata são ligeiramente maiores
- √ Efeito é esperado pois o número de quantidades a serem estimados no modelo mstrata é maior
 - s riscos de base $\lambda_{0s}(t)$ a serem estimados
 - É maior a incerteza dos coeficientes dos modelos

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Teste da Razão de Verossimilhança

- H₀: não há diferença entre os modelos
 √ (modelo maior = modelo menor)
- Estatística de teste: $RV = 2(l_{\text{maior}} l_{\text{menor}})$
 - $\sqrt{l_{\text{mod}}} = \ln(L_{\text{mod}})$
- Distribuição amostral: $RV \stackrel{H_0}{\sim} \chi_d^2$
 - \sqrt{d} : diferença no número de covariáveis nos modelos em questão

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Seleção de Modelos

· Teste de Wald

$$\sqrt{\mathbf{H}_0}$$
: $\beta_i = 0$

- Efeito da covariável é não significativo

√ Estatística de teste:

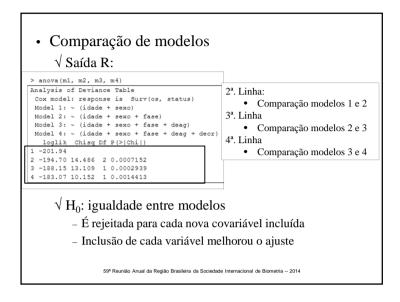
$$Z_j = \frac{\hat{\beta}_j}{\operatorname{ep}(\hat{\beta}_j)}$$

√ Distribuição amostral

$$Z_j \stackrel{H_0}{\sim} \mathrm{N}(0,1)$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 20

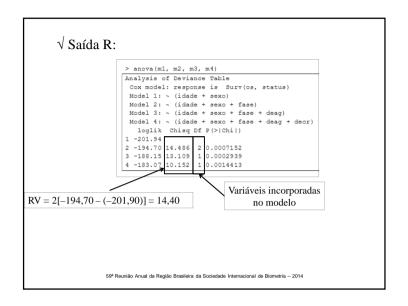
- Teste da razão de verossimilhança:
 - √ Compara modelos aninhados
 - Modelo com maior número de parâmetros tem de conter todos os parâmetros do modelo menor
 - √ Avalia se a inclusão de covariáveis no modelo aumenta significativamente a verossimilhança
 - √ Assintoticamente semelhante ao teste de Wald
 - É mais robusto quando o número de observações é pequeno
 - $\sqrt{\,\text{Modelos}\,}$ aninhados perdem comparabilidade caso existam valores ausentes
 - √ Não se pode comparar modelos estratificados com não estratificados



Qualidade de Ajuste

- O modelo se ajusta bem aos dados?
- Qual o poder explicativo do modelo
- Existem poucas estatísticas de ajuste:
 - √ Função desvio (deviance)
 - $\sqrt{R^2}$
 - √ Gráfico de Índice Prognóstico

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014



Variabilidade Total Explicada

- Proporção explicada pelas covariáveis
 - √ Modelo nulo: menor poder de explicação
 - Modelo sem covariáveis
 - Verossimilhança é apenas a linha de base
 - √ Modelo saturado: maior poder de explicação
 - Cada indivíduo está totalmente explicado
 - (um coeficiente para cada indivíduo)
 - Modelo sem resíduos (saturado)
 - Não explica nem prediz coisa alguma

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Modelo nulo:

```
mod.nulo <- coxph(Surv(os, status)~1, data = tmo)
mod.nulo
Call: coxph(formula = Surv(os, status) ~ 1, data = tmo)

Null model
log likelihood= -203.404
n= 96
```

Modelo saturado:

```
> mod.saturado <- coxph(Surv(os, status)~as.factor(id), data=tmo)
> summary(mod.saturado)
Call:
coxph(formula = Surv(os, status) ~ as.factor(id), data = tmo)

n= 96, number of events= 49
[...]

Concordance= 1 (se = 0.044)
Rsquare= 0.985 (max possible= 0.986)
Likelihood ratio test= 404 on 95 df, p=0

Wald test = 1826 on 95 df, p=0

Score (logrank) test = 394.1 on 95 df, p=0
```

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Exemplo:

√ Modelo m1:

$$R_{LR}^{2} = 1 - \exp\left(\frac{2\{l(0) - l(\hat{\beta})\}}{n}\right)$$

$$= 1 - \exp\left(\frac{2\{-201, 94 - (-203, 40)\}}{96}\right)$$

$$= 1 - \exp(-0, 003042) = 0,0300$$

√ Porcentagem de explicação:

$$\% \ \text{explicação} \ = \ \frac{R_{modelo}^2}{R_{saturado}^2} \times 100$$

$$= \ \frac{0,0300}{0,986} \times 100 = 2,958\%$$

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Medida Global de Ajuste – R²

• *R*²: poder explicativo das covariáveis no tempo de ocorrência do evento em estudo

$$R_{LR}^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\beta})} \right\}^{\frac{2}{n}} = 1 - \exp\left(\frac{2\{l(0) - l(\hat{\beta})\}}{n} \right)$$

- L(0): verossimilhança do modelo nulo
- $L(\hat{\beta})$: verossimilhança do modelo ajustado
- $\sqrt{\text{Valor mínimo possível } L(0)} = L(\hat{\beta})$
- √ Valor máximo não é 1 (100%)
 - É a razão entre as verossimilhanças dos modelos saturados e nulo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Verossimilhança dos modelos:

Modelo	$l_{ m modelo}$	\mathbb{R}^2	Variabilidade Explicada (%)
Nulo	-203,40	0,000	0,0
Saturado	-1,39	0,986	100,0
m1:idade + sexo	-201,94	0,030	3,0
m2:m1 + fase	-194,70	0,166	16,8
m3: m2 + deag	-188,15	0,272	27,6
m4: m3 + decr	-183,07	0,345	35,0

% explicação =
$$\frac{R_{modelo}^2}{R_{saturado}^2} \times 100$$

√ m4 é o melhor e explica 35% da variabilidade

Comentários

- Em modelos de sobrevivência é raro encontrar poder explicativo muito maior
 - √ Tempos de sobrevivência têm grande variabilidade individual
 - √ Diferenças causadas por fatores incorporados ao modelo trazem informação importante em
 - Avaliação terapêutica
 - Fatores prognósticos
 - Fatores de risco ou proteção

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Probabilidade de concordância estimada pelo modelo de Cox:
 - $\sqrt{\text{Entre 0,3 e 0,4:}}$
 - Modelo possui baixo poder preditivo (ou discriminatório)
 - $\sqrt{0,5}$:
 - A concordância pode ter ocorrido por acaso
 - $\sqrt{\text{Entre } 0,6 \text{ e } 0,7}$:
 - Resultado comum em Análise de Sobrevivência
 - $\sqrt{\text{Entre } 0.7 \text{ e } 0.8}$:
 - Resultado discriminatório muito bom
 - $\sqrt{\text{Entre }0.8 \text{ e }0.9}$
 - Resultado discriminatório excelente

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Probabilidade de Concordância

- Medida global de ajuste útil quando o objetivo do estudo é obter modelo preditivo
 - √ Usada para avaliar o poder discriminatório e a acurácia preditiva do modelo de Cox
- Concordância:
 - √ Ao selecionar aleatoriamente duas observações, a que possui menor tempo de sobrevivência é também aquela que possui o maior risco estimado (predito) pelo modelo de Cox

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

```
    Modelo m4

> m4 <- coxph(Surv(os, status)~(idade + sexo + fase + deag -
decr), data = tmo, x = TRUE)
> summary(m4)
coxph(formula = Surv(os, status) ~ (idade + sexo + fase + deag
    decr), data = tmo, x = TRUE)
  n= 96, number of events= 49
            coef exp(coef) se(coef) z Pr(>|z|)
idade -0.005019 0.994993 0.014912 -0.337 0.736415
sexo2 -0.271984 0.761866 0.332115 -0.819 0.412816
fase2 0.593973 1.811171 0.371427 1.599 0.109784
fase3 0.938411 2.555918 0.522357 1.796 0.072416 .
deag1 1.190381 3.288332 0.327485 3.635 0.000278 ***
decr1 -1.061750 0.345850 0.338452 -3.137 0.001706 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
      exp(coef) exp(-coef) lower .95 upper .95
idade 0.9950 1.0050 0.9663 1.0245
sexo2 0.7619 1.3126 0.3974 1.4607
        1.8112
                     0.5521
                                0.8746
fase3 2.5559
                     0.3912 0.9182
                                          7.1151
decr1 0.3459
                     2.8914 0.1782 0.6714
                                                                        Comando no R:
Concordance= 0.768 (se = 0.044)
                                                                         survConcordance()
Rsquare= 0.345 (max possible= 0.986)
Likelihood ratio test= 40.67 on 6 df, p=3.365e-07
                      = 38.46 on 6 df, p=9.113e-07
Score (logrank) test = 47.62 on 6 df, p=1.405e-08
                        598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014
```

• Exemplo:

Modelo	$l_{ m modelo}$	\mathbb{R}^2	Variabilidade Explicada (%)	
Nulo	-203,40	0,000	0,0	
Saturado	-1,39	0,986	100,0	1
m1:idade + sexo	-201,94	0,030	3,0	0,568
m2:m1 + fase	-194,70	0,166	16,8	0,650
m3: m2 + deag	-188,15	0,272	27,6	0,722
m4: m3 + decr	-183,07	0,345	35,0	0,768

√ Poder preditivo (ou discriminatório) do modelo m4 é muito bom

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Construção do gráfico:
 - √ Indivíduos são estratificados em grupos de tamanhos aproximadamente iguais
 - Grupos de alto, médio e baixo IP
 - √ Determinação dos valores médios de cada uma das covariáveis dentro de cada grupo
 - √ Obtenção das curvas de sobrevivência sob o modelo ajustado utilizando os valores médios
 - $\sqrt{\text{Curvas}}$ são comparadas às curvas de sobrevivência estimadas por K-M para cada grupo
 - $\sqrt{\text{Se o modelo for razoável:}}$
 - Curvas ajustadas sejam próximas das curvas K-M, por estrato

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Índice Prognóstico - IP

- Gráfico de sobrevivência estratificado por IP
 - $\sqrt{\text{Procedimento exploratório para avaliar qualidade de ajuste do modelo}}$
 - $\sqrt{\text{IP: preditor linear do modelo de Cox}}(x, \hat{\beta})$
 - √ Cálculo do IP para cada indivíduo

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Cálculo de índice prognóstico

√ Modelo m4

Covariável	O	Indivíduo A		Indivíduo B	
Covariavei	β	$x_{\rm A}$	βx_{A}	$x_{\rm B}$	$\beta x_{\rm B}$
idade	-0,005019	56	-0,281064	20	-0,100380
sexo	-0,271984	0	0	1	-0,271984
fase2	0,593973	1	0,593973	0	0
fase3	0,938411	0	0	1	0,938411
deag	1,190381	1	1,190381	1	1,190381
decr	-1,061750	0	0	0	0
Índice de P	Índice de Prognóstico		1,50329		1,756428

- Indivíduo A: sexo masculino, fase intermediária
- Indivíduo B: sexo feminino, fase avançada

√ Estratificação por índice de prognóstico – Comandos > source("funcoes/Rfun.r") Funcoes do livro Analise de Sobrevivencia: Teoria e Aplicações em Saude Ultima atualizacao: Julho de 2011 Funcoes disponiveis: plot.pi(), plot.frail() Função para estimar quantis do tempo de sobrevivencia. desenvolvida pelo Prof. John Fox plot.pi(m1, main = "M1", xlab = "Tempo") The solid line is the fitted model, the dashed one is the K-M > plot.pi(m2, main = "M2", xlab = "Tempo") The solid line is the fitted model, the dashed one is the K-M plot.pi(m3, main = "M3", xlab = "Tempo") The solid line is the fitted model, the dashed one is the K-M plot.pi(m4, main = "M4", xlab = "Tempo") The solid line is the fitted model, the dashed one is the K-M par(mfrow = c(1,1)) 59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

- Modelos m2, m3 e m4 são mais capazes de discriminar os grupos
 - √ Ajustam-se razoavelmente bem
 - √ Melhoram a cada nova variável incluída
 - $\sqrt{\ }$ % variabilidade explicada

16,8% **→** 27,6% **→** 35,0%

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

√ Gráficos estratificados por índice de prognóstico

M1

M2

Tempo

M3

M4

Apenas m1 apresenta-se mal ajustado

59º Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Função Desvio

- $\sqrt{H_0}$: modelo se ajusta aos dados
- $\sqrt{\text{Estatística de teste:}}$ $D = 2(l_{\text{saturado}} l_{\text{modelo}})$
 - $-l_{\text{saturado}}$: logverossimilhança do modelo saturado
 - ${\scriptstyle -}\ l_{\rm modelo}$: log
verossimilhança do modelo
- √ Modelo saturado:
 - Ajustados n parâmetros para n observações
- $\sqrt{\text{Distribuição amostral}}$ $D \stackrel{H_0}{\sim} \chi^2_{n-p-1}$
 - Quanto melhor a qualidade, menor a estatística e maior o p-valor

• Função desvio dos modelos:

Modelo	$l_{ m modelo}$	D	GL	P-valor
Nulo	-203,40	404,02		
Saturado	-1,39	0		
m1:idade + sexo	-201,94	401,10	93	0
m2: m1 + fase	-194,70	386,62	91	0
m3: m2 + deag	-188,15	373,52	90	0
m4: m3 + decr	-183,07	363,36	89	0

- √ Rejeita-se a hipótese de que o modelo se ajusta aos dados
 - Teste assintótico

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Análise de Resíduos

59ª Reunião Anual da Região Brasileira da

Tempos Empatados

- Modelo considera tempo uma variável contínua √ Na prática, tempo é medido de maneira discreta
- Empate na censura não é problema
 √ Estimação só é feita quando ocorre evento
- Se censura e evento estão empatados √ Considera-se que o evento aconteceu primeiro
- Empate de eventos
 √ Estimação por Efron, Breslow, exata

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Análise de Resíduos

- Premissas e ajuste de modelo quanto à
 - $\sqrt{\text{Proporcionalidade do risco}}$
 - √ Observações mal ajustadas pelo modelo
 - Pontos aberrantes e influentes
- Tipos de resíduos
 - √ Schoenfeld
 - √ Martingale
 - √ Deviance
 - √ Escore

Introdução

- Proporcionalidade:
 - √ A relação entre a variável resposta e tempo é sempre a mesma, independente do momento de ocorrência do evento
- Linearidade (log-linearidade) $\lambda(t) = \lambda_0(t)e^{\mathbf{x}'\boldsymbol{\beta}}$
 - √ A razão de riscos entre um indivíduo de 45 anos e um de 50 anos é idêntica àquela entre um indivíduo de 80 anos e um de 85 anos
- O modelo estima efeito médio de covariáveis:
 - $\sqrt{\text{Pontos influentes podem afetar estimativa fortemente}}$

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

menos a esperada não pode ser usado para os dados de sobrevivência:

• O resíduo obtido como resposta observada

√ Há censura

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Riscos Proporcionais - Schoenfeld

- · O efeito de uma variável é sempre o mesmo durante todo o tempo observado?
- Resíduo de Schoenfeld:
 - √ Diferença entre os valores observados de covariáveis de um indivíduo com tempo de ocorrência do evento t_i e os valores esperados em t_i dado o grupo em risco $R(t_i)$.
 - √ Haverá tantos vetores de resíduos quanto covariáveis ajustadas no modelo
 - √Estes são definidos somente nos tempos de ocorrência do evento.

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

• Para cada covariável x_i no tempo do evento t_i :

$$a_{ik} = \delta_i(x_{ik} - a_{ik})$$

$$a_{ik} = \frac{\sum_{j \in R(t_i)} x_{jk} \exp\left(\mathbf{x}'_j \hat{\boldsymbol{\beta}}\right)}{\sum_{j \in R(t_i)} \exp\left(\mathbf{x}'_j \hat{\boldsymbol{\beta}}\right)}$$

 $\sqrt{R(t_i)}$: conjunto dos indivíduos em risco no tempo t_i

 $\sqrt{x_{ik}}$: valor da covariável k do indivíduo j pertencente ao grupo de risco

- Supor que coeficiente β_k varia com o tempo. β_k pode ser dividido em duas partes:
 - $\sqrt{\text{M\'edia}}$ constante no tempo $E[r_i(\beta_k) \mid R(t_i)],$ com variância $V(\beta_k)$
 - √ Função U(t)
 - \sqrt{O} resíduo padronizado de Schoenfeld em t_i pode ser obtido por: $r_i^*(\beta_k) = \frac{r_i(\beta_k)}{V(\beta_k)}$
 - $\sqrt{\mbox{Se}}$ a premissa de proporcionalidade não é violada esperamos que o gráfico de r*k(t_j) vs. (t_j) apresente uma reta com inclinação zero

598 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

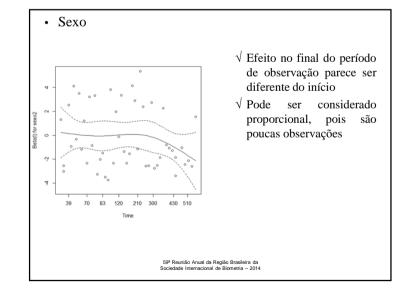
• Idade e Fase intermediária da doença: Idade Fase 2 Idade Fase 2 V Proporcionais ao longo do tempo

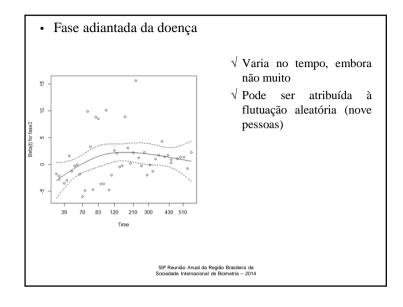
59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

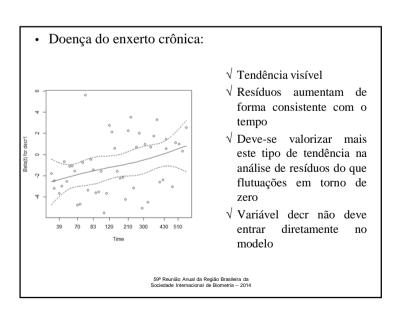
Schoenfeld no R

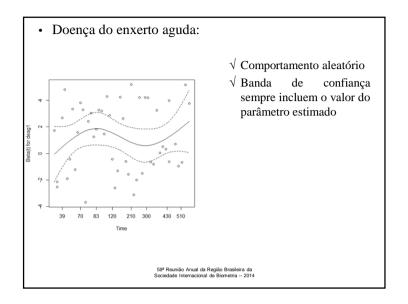
- residuo <- cox.zph(modelo)
- > plot(residuo[1])
- Observar a escala do tempo:
 - √ Kaplan-Meier nos tempos de falha
 - √ Calendário
 - bom quando ajuste utiliza processo de contagem
 - pode ficar pouco visível se concentra grande quantidade de eventos em um mesmo momento
 - $\sqrt{\text{Rank (ordem dos eventos):}}$
 - util quando os tempos são muito dispersos
- Suavização por lowess

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014









Comentários

- Em caso de problema de proporcionalidade:
 - √ Estratificar pela covariável tempo-dependente
 - √ Avaliar magnitude e pontos influentes
 - √ Particionar o eixo do tempo
 - $\sqrt{}$ Outro tipo de modelo
 - Modelo de tempo de vida acelerado

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Referências

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

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

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014

Bibliografia

- Carvalho, M. S. et al. *Análise de Sobrevivência: Teoria e Aplicações em Saúde*. (Fiocruz)
- Colosimo, E. A. e Giolo, S. R. *Análise de Sobrevivência Aplicada*. (Edgard Blucher)
- Klein, J. P. e Moeschberger, M. L. Survival Analysis: Techniques for Censored and Truncated Data. (Springer)

59ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria -- 2014