

Análise de Sobrevida

Valeska Andreozzi¹

valeska.andreozzi@ic.ufrj.br

&

Marília Sá Carvalho²

carvalho@fiocruz.br

¹Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

²Escola Nacional de Saúde Pública e Programa de Computação Científica da Fundação Oswaldo Cruz, Brasil

Julho, 2008

Agradecimentos

- À Fiocruz, que viabilizou o projeto.
- Às instituições e seus pesquisadores que cederam, mais do que seus dados, seus problemas, idéias, perguntas:
 - Departamento de Informação e Informática do SUS - Datasus;
 - Escola Nacional de Saúde Pública – Fundação Oswaldo Cruz;
 - Hospital Geral de Betim;
 - Hospital Universitário Clementino Fraga Filho – Universidade Federal do Rio de Janeiro;
 - Hospital Universitário Gaffrée e Guinle – Universidade Federal do Estado do Rio de Janeiro;
 - Instituto de Pesquisa Clínica Evandro Chagas – Fundação Oswaldo Cruz;
 - Instituto de Saúde Coletiva – Universidade Federal da Bahia;
 - Instituto Nacional do Câncer.

Objetivos do curso

- Apresentar, de forma simples e acessível, as várias formas de modelar o tempo de sobrevivência na área da Saúde.
- Descrever os modelos de Cox e as suas extensões para tratar covariáveis dependentes do tempo, eventos múltiplos e fragilidade (efeitos aleatórios). Não serão contemplados os modelos paramétricos.
- Apresentar o software R e o pacote `survival` para a implementação dos modelos estatísticos apresentados utilizando vários exemplos reais.
- Proporcionar aos participantes a aplicação dos conceitos teóricos e práticos na sua própria investigação.

Material do curso

- Notas de aula
- Web page do livro-texto
Análise de Sobrevida: Teoria e Aplicação em Saúde, Carvalho, MS, Andreozzi V, Codeço, CT, Barbosa, MTS, Shimakura, SE, Editora Fiocruz 2005. (<http://dengue.procc.fiocruz.br/~sobrevida/>)
- R software (www.r-project.org)
- Tutorias online do R
 - <http://www.leg.ufpr.br/Rtutorial/>
 - <http://www.leg.ufpr.br/~paulojust/embrapa/Rembrapa/>
- Outros pacotes para análise de sobrevida
 - Pesquise a palavra *survival* em <http://neacm.fe.up.pt/CRAN/>

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Sobrevida ou sobrevivência

- Em que tipo de desenho de estudo se aplica a *Análise de Sobrevida*?
- Estudo de coorte, observacional ou de intervenção (ensaios clínicos): pressupõe o acompanhamento dos indivíduos ao longo do tempo.

Sobrevida ou sobrevivência

- A **análise de sobrevida**, também chamada de **análise de sobrevivência**, será utilizada quando o tempo for o objeto de interesse, seja este interpretado como **o tempo até a ocorrência de um evento** ou o **risco de ocorrência de um evento por unidade de tempo**.
- As perguntas passíveis de resposta neste tipo de abordagem são:
 - Qual o efeito de um determinado anticancerígeno sobre o tempo de sobrevida?
 - Quais os fatores associados ao tempo de duração da amamentação?
 - Quais os fatores preditivos para reinternação hospitalar, considerando o tempo entre internações?
 - Qual o efeito da unidade assistencial na sobrevida após um infarto agudo do miocárdio?

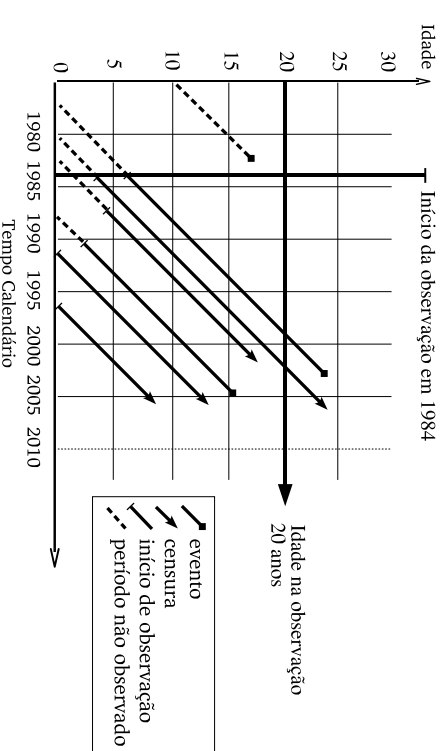
O Tempo

- Três efeitos do tempo surgem com frequência num contexto epidemiológico:
 - idade - tempo desde o nascimento
 - período - momento da medida
 - coorte - grupo de indivíduos nascidos no mesmo ano
- Frequentemente é difícil distinguir da idade o efeito de período e de coorte.

Duração

- Existe vasta literatura sobre efeito de coorte, de período e de idade.
- No contexto dos estudos de sobrevivida o interesse central é a **duração** de uma dada condição.
 - Problemas?
 - Exemplos?

Diagrama de lexis: idade, tempo, coorte



Programa

- 1 Introdução
- 2 **O Tempo**
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

O Tempo

Tempo até...

- óbito
- transplante
- doença
- cura

Medir o tempo

Tabela: Tempo de sobrevivida (em meses) de 10 pacientes em diálise.

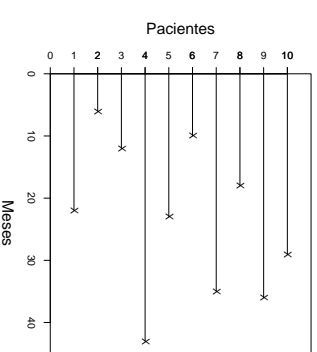
Paciente (i)	Tempo (T_i)
1	22
2	6
3	12
4	43
5	23
6	10
7	35
8	18
9	36
10	29

Informação incompleta

- óbito por outras causas – morte do paciente por causas externas;
- término do estudo;
- perda de contato – mudança de residência;
- recusa em continuar participando do estudo;
- mudança de procedimento;
- abandono devido a efeitos adversos de tratamento (iii);
- desconhecimento da data de início – em pacientes HIV+ com data de infecção desconhecida;
- dados truncados – prevalentes.

Censura e truncamento

Representar o tempo



Cada linha representa a trajetória de um paciente e o símbolo **X** indica a ocorrência do evento ou falha.

Mecanismos de censura

Censura à direita

- É a mais comum.
- Sabe-se que o tempo entre o início do estudo e o evento é maior do que o tempo observado.
- Nesse caso aproveita-se a informação do tempo durante o qual a pessoa esteve sob observação sem que ocorresse o evento.
- Desprezar essa informação faria com que o risco fosse superestimado, pois o tempo até a evento é desconhecido, mas o paciente estava em risco de sofrer o evento pelo menos até o último momento observado.

Dados com censura à direita

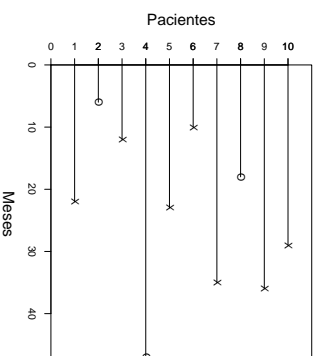
Exemplo

- Visando estudar o tempo entre o diagnóstico de Aids e o óbito, 193 pacientes foram acompanhados em um ambulatório especializado de 1986 a 2000. Durante esse período, foram observados 92 óbitos. Sabemos que até a data de término do estudo, em dezembro de 2000, 101 permaneciam vivos. Não há informações mais recentes. Dizemos, então, que ocorreram 92 eventos e 101 censuras (à direita).

<http://dengue.procc.fiocruz.br/~sobrevida/dados/aids.html>

Dados com censura à direita

Graficamente



X indica ocorrência do evento e O corresponde à presença de censura.

Dados com censura à direita

Dados de 10 pacientes

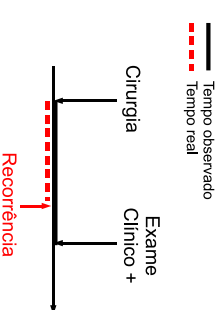
Paciente (i)	Tempo (T_i)	Status (δ_i)
1	22	1
2	6	0
3	12	1
4	43	0
5	23	1
6	10	1
7	35	1
8	18	0
9	36	1
10	29	1

Status=1 (evento); Status=0 (censura)

Mecanismos de censura

Censura à esquerda

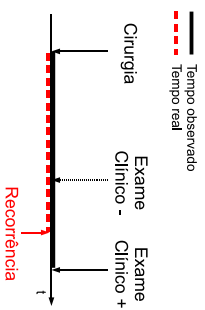
- Acontece quando não conhecemos o momento da ocorrência do evento, mas sabemos que a duração do evento é menor do que a observada.
- Considere um estudo para investigar o tempo de recorrência de um câncer após cirurgia para remoção. Após três meses da operação, pacientes são examinados e alguns apresentam recorrência. Para tais pacientes, sabemos que o tempo entre a cirurgia e a recorrência é menor que três meses, como indica o quadro abaixo



Mecanismos de censura

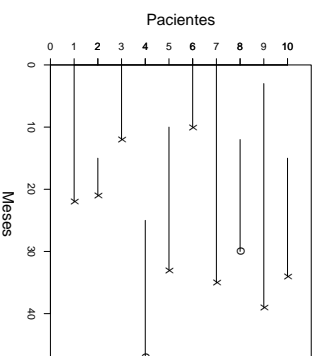
Censura intervalar

- Ocorrência do evento entre tempo conhecidos
- Aqui o paciente não apresenta recorrência na consulta após três meses da cirurgia, mas sim na consulta seguinte após 2 meses da consulta anterior.
- Neste caso sabemos o tempo até a recorrência é maior que 3 meses e menor que 5 meses.



Coorte aberta

Momento de entrada dos pacientes na coorte varia



Trajelórias individuais de pacientes com censura e com diferentes tempos de entrada em observação.

Informativa???

A censura ainda pode ser classificada em:

- Informativa
 - perda do indivíduo em decorrência de causa associada ao evento estudado; abandono do tratamento devido a piora do paciente
- NÃO Informativa
 - quando não há razão para suspeitar que o motivo da perda de informação esteja relacionado ao desfecho

Registro do tempo

Tempo de observação de pacientes de uma coorte aberta.

Paciente	Tempo* inicial (I)	Tempo* final (F)	Tempo* T (final - inicial)	Status (δ)
1	0	22	22	1
2	15	21	6	1
3	0	12	12	1
4	25	47	22	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	34	19	1

*Tempo calendário em meses

Truncamento

- À esquerda ocorre quando a perda da informação está relacionada a indivíduos que foram excluídos do estudo porque já tinham experimentado o evento antes do início do estudo e não podiam ser observados (**dados prevalentes**).
- À direita ocorre quando o critério de seleção inclui somente indivíduos que sofreram o evento. Não apresenta problema em doenças com curta duração

Processo de contagem

O par (T_i, δ_i) é substituído por $(N_i(t), Y_i(t))$, onde:

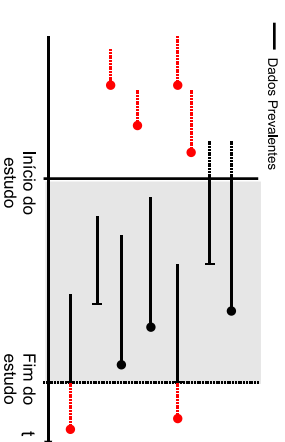
$N_i(t)$ é o número de eventos observados em $[0, t]$

$Y_i(t) = 1$, se o indivíduo i está sob observação e sujeito ao risco do evento no instante t

$Y_i(t) = 0$, se o indivíduo i não está em risco.

Truncamento

Representação gráfica



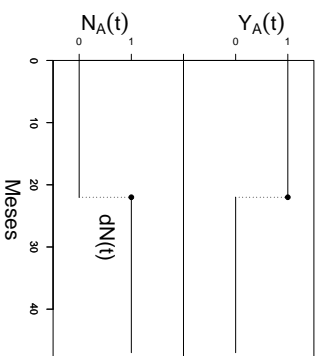
Pacientes em vermelho e trajetórias em pontilhado não serão incluídos no estudo

Processo de contagem

Formalmente:

- um processo de contagem é um processo estocástico $N(t)$ com $t > 0$, de tal forma que $N(0) = 0$ e $N(t) < \infty$;
- a trajetória de $N(t)$ é contínua à direita a partir de uma função escada com saltos de tamanho igual a um;
- a análise de sobrevida pode ser pensada como um processo de contagem onde $N(t)$ é o número de eventos observados até o tempo t e $dN_i(t)$ é a diferença entre a contagem de eventos até o instante t e a contagem no momento imediatamente anterior a t .

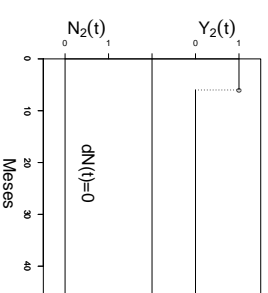
Graficamente



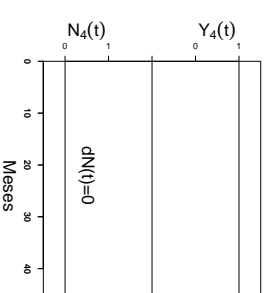
Paciente A: Diagnosticado no mês zero, acompanhado até o mês 22. A ocorrência do evento é assinalada pelo sinal •

Graficamente

Trajatória de dois pacientes censurados



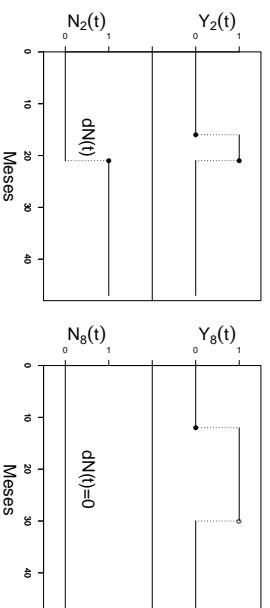
censura aos 6 meses



censura ao término do estudo

Graficamente

Trajetória de dois pacientes censurados que entraram na coorte ao longo do estudo



Qual o ganho?

O que se ganha com o processo de contagem?

Possibilidade de analisar:

- Mudança no valor de covariável
- Evento múltiplos
- Dados prevalentes

Organização dos dados

id	tempo (T)	status (δ)	sexo	idade
1	30	0	F	54
2	14	1	F	34
3	23	1	M	65
4	11	1	F	45
5	12	0	M	44

Organização dos dados

id	inicio (I)	fim (F)	status (δ)	sexo	idade
1	0	30	0	F	54
2	5	19	1	F	34
3	3	26	1	M	65
4	0	11	1	F	45
5	4	16	0	M	44

Tempo de Sobrevida no R

- O R aceita os dois formatos de registro do tempo de sobrevida.
- O comando `Surv()` tem como função combinar, em uma única variável, a informação referente ao tempo de sobrevida de cada indivíduo e a informação a respeito do status do paciente.
 - Status = 1 (um), se ocorreu o evento
 - Status = 0 (zero) se o tempo foi censurado
- `require(survival)`
 - `Surv(tempo, status)`
 - `Surv(inicio, fim, status)`

O objeto sobrevida

```
> require(survival)
> ipec<-read.table("ipec.csv", header=T, sep=";")
> ipec[1:9,c("id", "tempo", "status")]
  id tempo status
1 1 852 1
2 2 123 1
3 3 1145 1
4 4 2755 0
5 5 2117 0
6 6 329 0
7 7 60 1
8 8 151 1
9 9 1563 1

> Surv(ipec$tempo, ipec$status)
 [1] 852 123 1145 2755+ 2117+ 329+ 60 151 1563
```

Programa

- 1 Introdução
- 2 O Tempo
- 3 **Funções de Sobrevida**
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Funções de sobrevida

- Densidade de Probabilidade
- Sobrevida
- Risco (instantâneo)
- Risco Acumulado

Função – densidade de probabilidade

- T – tempo de sobrevida (até a ocorrência de um evento);
- T é uma variável aleatória contínua e positiva;
- $f(t)$ é a sua função de densidade de probabilidade;
- a função $f(t)$ pode ser interpretada como a probabilidade de um indivíduo sofrer um evento em um intervalo instantâneo de tempo.

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T \leq t + \Delta t)}{\Delta t}$$

Estimativa de probabilidade sem censura

Se não houver censura, isto é, se **todos** os pacientes apresentarem o evento antes do fim do estudo, a função $f(t)$ pode ser estimada a partir da tabela de frequência.

Nesta tabela, os valores observados de T são distribuídos em classes e para cada classe x , calcula-se $\hat{f}_x(t)$:

$$\hat{f}_x(t) = \frac{n^o \text{ de ocorrências na classe } x}{n^o \text{ total de ocorrências} \times (\text{amplitude de } x)} \quad (1)$$

Função de sobrevida

Qual é a probabilidade de um paciente com aids sobreviver 365 dias ou mais? Isto é, qual a probabilidade de T ser maior do que um determinado valor $t = 365$? Ou, mais formalmente, qual é $Pr(T \geq 365)$?

A função de sobrevida, $S(t)$, é a probabilidade de um indivíduo sobreviver por mais do que um determinado tempo t .

$$S(t) = Pr(T \geq t)$$

Estimando a sobrevida – sem censura

$$\hat{S}_x(t_{inf}) = \frac{n^o \text{ pacientes com } T > t_{inf}}{n^o \text{ total de pacientes}}$$

em que t_{inf} é o limite inferior do intervalo de tempo considerado x .

Função de sobrevida

Relembrando: a função de distribuição acumulada, $F(t)$, de uma variável aleatória é definida como a probabilidade de um evento ocorrer até o tempo t .

$$F(t) = Pr(T < t)$$

Logo, $S(t)$ é o complemento da função de distribuição acumulada $F(t)$:

$$S(t) = Pr(T \geq t) = 1 - Pr(T < t) = 1 - F(t)$$

Função de Risco

- Qual é o risco de um paciente com aids vir a óbito após sobreviver 365 dias?
 - Esse risco de morrer aumenta ou diminui com o tempo?
- $\lambda(t) \Rightarrow$ probabilidade instantânea de um indivíduo sofrer o evento em um intervalo de tempo t e $(t + \epsilon)$ dado que ele sobreviveu até o tempo t .

Sendo ϵ infinitamente pequeno, $\lambda(t)$ expressa o risco instantâneo de ocorrência de um evento, dado que até então o evento não tenha ocorrido.

Função de Risco

$$\lambda(t) = \lim_{\epsilon \rightarrow +\infty} \frac{\text{Pr}((t < T < t + \epsilon) | T > t)}{\epsilon}$$

- $\lambda(t)$ também é denominada:
 - função ou taxa de incidência,
 - força de infecção,
 - taxa de falha,
 - força de mortalidade,
 - força de mortalidade condicional.
- Apesar do nome risco, $\lambda(t)$ é uma taxa (tempo^{-1}).
- Pode assumir qualquer valor positivo (**não** é probabilidade).

Estimando risco sem censura

$$\hat{\lambda}_x(t) = \frac{n^o \text{ ocorrências na classe } x}{R_x(t) \times (\text{amplitude de } x)}$$

Número de eventos observados no intervalo de classe x divididos pelo número de pacientes em risco no início do intervalo x e pela amplitude de x .

Uma maneira alternativa de estimar $\lambda(t)$ é utilizar as relações entre $S(t)$, $f(t)$ e $\lambda(t)$.

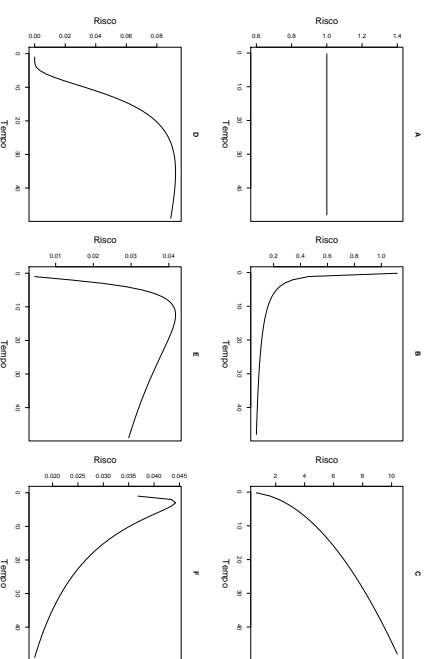
Função de Risco e de Sobrevida

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

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

Sobrevida e risco são inversamente proporcionais: quando o risco aumenta, a probabilidade de sobrevida diminui e vice-versa.

Comportamento do Risco



Função de risco com diversos formatos.

Função de risco acumulado

- Qual o risco de um paciente com aids vir a óbito no primeiro ano após o diagnóstico?
- Qual é o risco dele vir a óbito nos primeiros 2 anos?

$\Lambda(t)$ \rightarrow função de risco acumulado.

Mede o risco de ocorrência do evento até o tempo t .

É a soma (integral) de todos os riscos em todos os tempos até o tempo t .

$$\Lambda(t) = \int_0^t \lambda(u) du$$

Também é uma taxa, logo não está restrita ao intervalo $[0; 1]$.

Estimando risco acumulado sem censura

$$\hat{\Lambda}_x(t) = \sum_{k=2}^{k=x-1} \hat{\lambda}_k(t) \times \text{amplitude de } k$$

- O risco acumulado até o tempo t é igual a:
 - o risco acumulado até o tempo $t - 1$ mais
 - o risco instantâneo do período anterior vezes o intervalo de tempo até t .

Planilha tempo.xls

Relação entre as funções básicas de sobrevida

$$\begin{aligned} S(t) &= 1 - F(t) \\ \lambda(t) &= -\frac{d \ln(S(t))}{dt} \\ \lambda(t) &= \frac{f(t)}{S(t)} \\ \lambda(t) &= \frac{-f(t)}{1-F(t)} \\ \Lambda(t) &= -\ln(S(t)) \end{aligned}$$

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 **Estimação Não-Paramétrica**
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Estimação Não-Paramétrica

- Estimadores de sobrevivência e risco
- Kaplan-Meier e Nelson Aalen
- Intervalos de confiança
- Kaplan-Meier estratificado
- Testes de Log-Rank e Peto

Incorporando a censura

Sem suposições sobre a distribuição do tempo

Kaplan-Meier

- A probabilidade de sobrevivência até o tempo t é estimada considerando que a sobrevivência até cada tempo é independente da sobrevivência até outros tempos.
- A probabilidade de chegar até o tempo t é o produto da probabilidade de chegar até cada um dos tempos anteriores.

Kaplan-Meier

- Seja $t_1 < t_2 < \dots < t_m$ os tempos onde ocorreram os eventos;
- $Y_i(t) = 1$ se a pessoa i está em risco no tempo t e 0 caso contrário.
- $R(t_i)$ é o total de pessoas a risco no tempo t_i .
- A cada tempo t_i em que houver um evento, a probabilidade de sobrevivência será o número dos que sobreviveram até aquele tempo ($R(t_i) - N(t_i)$) sobre os que estavam em risco naquele tempo ($R(t_i)$).
- O estimador da distribuição $S(t)$ é o produto das probabilidades de sobrevivência a cada tempo $t_i \leq t$.

Kaplan-Meier

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

ou na forma de produto:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{R(t_i) - N(t_i)}{R(t_i)}$$

Da sobrevida ao risco

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

Logo... pode-se estimar qualquer das funções.

Intervalos de confiança

Variância do estimador Kaplan-Meier para a sobrevida
Estimador de Greenwood

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

Estimador de Nelson-Aalen

$$\hat{\Lambda}_{NA}(t) = \sum_{t_i \leq t} \frac{N(t_i)}{R(t_i)}$$

Melhor para amostras muito pequenas

planilha kaplanmeier.xls

Intervalos de confiança

Assumindo erro α , o intervalo fica assim:

$$\left[\hat{S}_{KM}(t) - z_{\alpha/2} \sqrt{Var(\hat{S}_{KM}(t))}; \hat{S}_{KM}(t) + z_{\alpha/2} \sqrt{Var(\hat{S}_{KM}(t))} \right]$$

Entretanto, este intervalo permite valores negativos e maiores do que 1, o que é incompatível com a definição de sobrevida.

Intervalos de confiança

Construindo intervalo simétrico para o risco $\ln \Lambda(t) = \ln(-\ln S(t))$, pode-se obter um intervalo assimétrico para $S(t)$, porém sempre positivo e menor do que 1 e igual a

$$[\exp(-\exp(t_s)); \exp(-\exp(t_t))]$$

onde

$$[t_s; t_t] = \left[\ln(\hat{\Lambda}_{KM}(t)) - z_{\alpha/2} dp; \ln(\hat{\Lambda}_{KM}(t)) + z_{\alpha/2} dp \right]$$

e o desvio padrão dp é:

$$dp = \sqrt{\frac{\frac{\sum_{t_i \leq t} R(t_i)(R(t_i) - N(t_i))}{N(t_i)}}{\left\{ \sum_{t_i \leq t} \ln \left[\frac{R(t_i) - N(t_i)}{N(t_i)} \right] \right\}^2}}$$

Saídas do R – summary(KM)

```
time n.risk n.event survival std.err lower95%CI upper95%CI
16 21 1 0.9524 0.0465 0.8655 1.000
18 20 1 0.9048 0.0641 0.7875 1.000
21 19 1 0.8571 0.0764 0.7198 1.000
22 17 1 0.8067 0.0869 0.6531 0.996
29 15 1 0.7529 0.0963 0.5859 0.968
35 14 1 0.6992 0.1034 0.5232 0.934
37 13 1 0.6454 0.1085 0.4642 0.897
39 12 1 0.5916 0.1120 0.4082 0.857
40 11 1 0.5378 0.1140 0.3550 0.815
52 9 1 0.4781 0.1160 0.2972 0.769
54 8 1 0.4183 0.1158 0.2431 0.720
60 6 1 0.3585 0.1137 0.1926 0.667
80 6 1 0.2988 0.1093 0.1459 0.612
83 3 1 0.1992 0.1092 0.0680 0.583
84 2 1 0.0996 0.0891 0.0172 0.575
```

no R

- Criando o objeto sobrevida (tempo, censura):

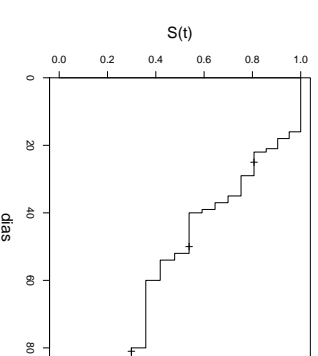

```
> Surv(tempo,status)
# variável status=1 indica evento, 0 censura
16 18 21+ 21 22 25+ 29 35 37 39 40 50+ 52 54 60 80+ 80 81+ 83 84 85+
```
- Kaplan-Meier


```
> KM <- survfit(Surv(tempo,status), data = ipec90)
> summary(KM)
> plot(KM)
```
- Nelson-Aalen


```
> sob.NA <- survfit(coxph("y~1", data = ipec90))
> sob.NA
> summary(sob.NA)
```

Saídas do R – plot(KM)

Função de sobrevida dos pacientes com aids, utilizando o estimador produto Kaplan-Meier.
Os símbolos + localizam as censuras.



Kaplan-Meier estratificado

- A sobrevivência é estimada separadamente para cada estrato, utilizando Kaplan-Meier.

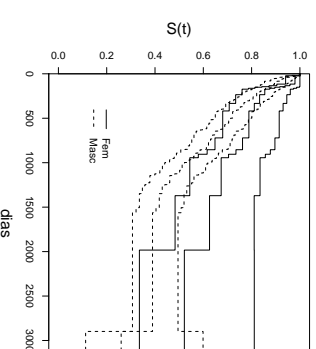
- no R

```
> ippec <- read.table("ippec.csv", header=T, sep=";", "
> survivals <- survfit(Surv(tempo,status) ~ sexo, data = ippec)
> survivals
```

```
Call: survfit(formula = resp ~ sexo, data = ippec)

      n  events  mean se(rmean) median 0.95LCL 0.95UCL
sexo=F   49     16  2096     229      Inf   1371      Inf
sexo=M  144     74  1581     122     1116    887   1563
```

Gráfico sobrevida estratificada



Curvas de sobrevida de pacientes com aids, estratificado por sexo. Estimação por Kaplan-Meier, com intervalo de confiança de 95%.

Testes

Hipótese nula: não há diferença entre estratos

$$H_0 : \lambda_1(t) = \lambda_2(t) = \dots = \lambda_k(t)$$

Log-rank (ou Mantel-Haenszel)

Distribuição esperada de eventos igual em todos os estratos:

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

Estatística de teste log-rank para dois estratos ($k = 2$):

$$\text{Log-rank} = \frac{(N_1 - E_1)^2}{\text{Var}(N_1 - E_1)}$$

com N_1 = ao total de eventos **observados** no estrato 1 e E_1 = ao total de eventos **esperados** no estrato 1.

Teste log-rank

A variância, que entra no cálculo como um fator de padronização, tem a fórmula (para $k = 2$):

$$\text{Var}(N_1 - E_1) = v_i$$

em que

$$v_i = \sum_{t_i} \frac{R_1(t_i)[R(t_i) - R_1(t_i)]N(t_i)[R(t_i) - N(t_i)]}{R(t_i)^2[R(t_i) - 1]}.$$

A estatística log-rank, sob a hipótese nula, segue uma distribuição χ^2 , com $k - 1$ graus de liberdade.

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

69 / 264

no R

Estimação Não-Paramétrica

```
> survdiff(Surv(tempo,status)~sexo, data=ipec, rho=0)
Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec, rho = 0)

      N Observed Expected (0-E)^2/E (0-E)^2/V
sexo=F  49      16      24.5      2.93      4.03
sexo=M 144      74      65.5      1.09      4.03

Chisq= 4 on 1 degrees of freedom, p= 0.0447
```

O argumento `rho` determina o tipo de teste a ser realizado. Para log-rank, use `rho = 0 (default)`. Para o teste Peto, use `rho = 1`.

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

71 / 264

Teste de Peto

Dá maior peso às diferenças (ou semelhanças), no início da curva, onde se concentra a maior parte dos dados e por isso é mais informativa. Usa um ponderador $S(t_i)$ no estimador.

$$\text{Peto} = \frac{(N_1 - E_1)^2}{\text{Var}(N_1 - E_1)}$$

sendo que

$$N_1 - E_1 = \frac{\sum S(t_i)(N_1(t_i) - E_1(t_i))}{\sum S(t_i)}$$

$$\text{Var}(N_1 - E_1) = \frac{(\sum S(t_i)(N_1(t_i) - E_1(t_i)))^2}{\sum (S(t_i))^2 v_i}$$

Também a estatística Peto segue aproximadamente uma distribuição χ^2 com $k - 1$ graus de liberdade.

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

70 / 264

no R

Estimação Não-Paramétrica

```
> survdiff(Surv(tempo,status)~sexo, data=ipec, rho=1)
Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec, rho = 1)

      N Observed Expected (0-E)^2/E (0-E)^2/V
sexo=F  49      12.1      18.2      2.011      3.54
sexo=M 144      55.1      49.0      0.746      3.54

Chisq= 3.5 on 1 degrees of freedom, p= 0.0598
```

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

72 / 264

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 **Modelagem Paramétrica**
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Distribuições

- distribuições estatísticas para modelar as funções de sobrevida:
 - Exponencial
 - Weibull
 - Lognormal

Paramétrica e Não-paramétrica

Os estimadores de Kaplan-Meier e Nelson-Aalen para as funções $S(t)$ e $\lambda(t)$ são obtidos a partir dos dados, supondo que a cada momento do tempo existe um processo diferente gerando as observações.

Como cada intervalo de tempo é estimado de forma independente, a estimação não-paramétrica possui tantos parâmetros quantos intervalos de tempo.

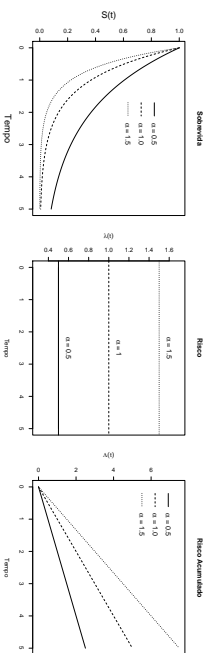
Distribuição Exponencial

Se a variável T possui uma distribuição exponencial,

- Densidade de probabilidade: $f(t) = \alpha \exp(-\alpha t)$, $\alpha > 0$
- Função de sobrevida: $S(t) = \exp(-\alpha t)$
- A função risco é constante para todo o tempo de observação t , ou seja: $\lambda(t) = \frac{f(t)}{S(t)} = \alpha = \text{constante}$
- A função de risco acumulado é uma função linear no tempo e é dada por: $\Lambda(t) = -\ln S(t) = \alpha t$

Algumas exponenciais

Função de sobrevivida, de risco e de risco acumulado para a distribuição exponencial considerando diferentes valores de α



A distribuição exponencial é conhecida como distribuição exponencial padrão quando $\alpha = 1$.

Distribuição Weibull

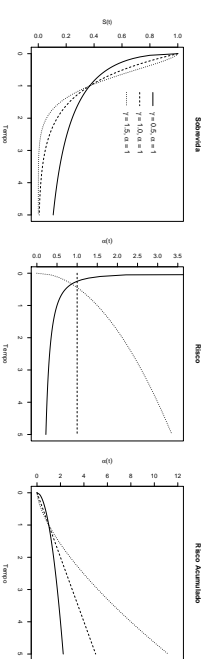
- permite variação do risco no tempo
- é uma generalização da distribuição exponencial, sendo a densidade $f(t) = \gamma \alpha^\gamma t^{\gamma-1} \exp(-(\alpha t)^\gamma)$ e a sobrevivida $S(t) = \exp(-(\alpha t)^\gamma)$ ($\alpha > 0$ e $\gamma > 0$)
- o parâmetro γ determina a forma da função de risco sendo chamado de parâmetro de forma:
 - $\gamma < 1$ função de risco decrescente
 - $\gamma > 1$ função de risco crescente
 - $\gamma = 1$ função de risco constante (equivalente ao modelo exponencial)
- o parâmetro α determina a escala da distribuição
- a função de risco é: $\lambda(t) = -\ln S(t) = (\alpha t)^\gamma - 1$

Interpretando risco exponencial

- média e a variância: $\bar{T} = \frac{1}{\alpha}$ e $var(T) = \frac{1}{\alpha^2}$
- quanto maior o risco, menor o tempo médio de sobrevivida e menor a variabilidade deste em torno da média
- como a distribuição do tempo de sobrevivida T é assimétrica, usa-se mais o tempo mediano
- o modelo exponencial é matematicamente simples, mas a suposição de risco constante no tempo é pouco plausível
- é aplicável quando o tempo de acompanhamento é curto o suficiente para que o risco naquele período possa ser considerado constante (por ex., o risco de óbito de crianças entre dois e cinco anos, independente da causa, pode ser considerado constante neste intervalo)

Algumas Weibull

Função de sobrevivida, de risco e de risco acumulado com parâmetro escala $\alpha = 1$ e diferentes valores do parâmetro de forma γ



Regressão Paramétrica

- Nos modelos paramétricos, a inclusão de covariáveis segue a forma utilizada em modelos lineares generalizados, podendo ser tanto contínuas – pressão sanguínea, idade, dosagens bioquímicas – como categóricas – gênero, tratamento, comportamentos.
- O objetivo de um modelo de regressão é o de estimar o efeito de covariáveis (ou variáveis independentes ou preditores), x_1, x_2, \dots, x_p , sobre uma variável resposta (ou variável dependente), Y .
- Supondo uma distribuição da família exponencial para a variável resposta teremos um modelo linear generalizado.
- Ainda que a distribuição exponencial e a Weibull sejam parte desta família, os modelos de regressão paramétricos para tempo de sobrevivida não são parte dos GLM por causa de dados censurados.

Revisando

- Estimação paramétrica
- Regressão paramétrica

Exemplo

Assumindo que o risco de morrer é constante ao longo do tempo, pode-se estimar o efeito da *idade* na sobrevivida e no risco de 6.805 pacientes em diálise acompanhados durante um ano (1.603 morreram) através do modelo exponencial:

$$\lambda(t|idade) = \exp(\beta_0 + idade\beta_1)$$

Os parâmetros estimados são: $\beta_0 = 6, 14$ e $\beta_1 = -0, 04$, ou seja, para cada ano a mais de vida o risco aumenta de $\exp(0, 04) = 0, 96$.

Pode-se comparar o risco constante de morte no tempo, entre dois indivíduos submetidos à diálise, um com 30 anos e outro com 70, substituindo as estimativas dos parâmetros β :

$$\frac{\lambda(t|x_1 = 30)}{\lambda(t|x_1 = 70)} = \frac{\exp(\beta_0 + 30\beta_1)}{\exp(\beta_0 + 70\beta_1)} = \frac{0, 006}{0, 029} = 4, 39$$

Distribuição Exponencial

$$T \sim \text{Exp}(\alpha)$$

$$f(t) = \alpha \exp(-\alpha t), \quad \alpha > 0$$

$$E(t) = \frac{1}{\alpha}$$

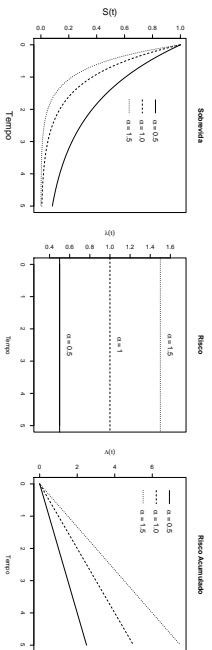
$$S(t) = \exp(-\alpha t)$$

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

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

Algumas exponenciais

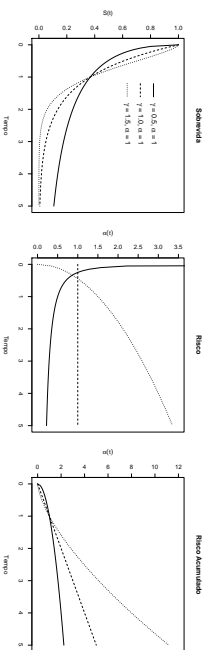
Função de sobrevivida, de risco e de risco acumulado para a distribuição exponencial considerando diferentes valores de α



A distribuição exponencial é conhecida como distribuição exponencial padrão quando $\alpha = 1$.

Algumas Weibull

Função de sobrevivida, de risco e de risco acumulado com parâmetro escala $\alpha = 1$ e diferentes valores do parâmetro de forma γ



Distribuição Weibull

$$T \sim Weibull(\alpha, \gamma)$$

$$f(t) = \gamma \alpha^\gamma t^{\gamma-1} \exp(-(\alpha t)^\gamma)$$

$$S(t) = \exp(-(\alpha t)^\gamma)$$

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

$$\Lambda(t) = -\ln S(t) = (\alpha t)^\gamma - 1$$

Exemplo no R

```
> Leite <- read.table("Leite.txt", header=T)
> Leite
```

crianca	tempo	status
1	1	6
1	1	1
2	2	12
2	2	1
3	3	10
3	3	1
4	4	3
4	4	1
5	5	5
5	5	1
6	6	1
6	1	1
7	7	6
7	6	1
8	8	8
8	1	1
9	9	1
9	1	1
10	10	5
10	1	1
11	11	2
11	1	1
12	12	2
12	1	1
13	13	5
13	1	1
14	14	8
14	1	1
15	15	1
15	1	1

Exemplo no R - Exponential

```
> modeloexp <- survreg(Surv(Leite$tempo, Leite$status)~1,
  data=leite, dist="exponential")
> summary(modeloexp)

Call:
survreg(formula = Surv(Leite$tempo, Leite$status) ~ 1, data = leite,
  dist = "exponential")
      Value Std. Error      z      p
(Intercept)  1.61    0.258  6.23 4.57e-10
Scale fixed at 1
Exponential distribution
Loglik(model)=-39.1  Loglik(intercept only)=-39.1
Number of Newton-Raphson Iterations: 4
n= 15

T ~ Exp(alpha)
alpha = exp(-1.61) = 0.2
lambda(t) = 0.2
S(t) = exp(-0.2t)
E(t) = 1/0.2 = 5.002 -> mean(Leite$time)=5
```

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

89 / 264

Regressão Paramétrica

- Assumimos que o parâmetro da distribuição depende de covariáveis segundo uma função
- Exemplo: $\alpha(\mathbf{x}) = \exp(\mathbf{x}\beta)$
- Regressão Exponencial:

$$S(t|\mathbf{x}) = \exp(-\alpha(\mathbf{x})t) = \exp(-\exp(\mathbf{x}\beta)t^\gamma)$$

$$\lambda(t|\mathbf{x}) = \alpha(\mathbf{x}) = \exp(\mathbf{x}\beta)$$
- Regressão Weibull:

$$S(t) = \exp(-(\alpha(\mathbf{x})t)^\gamma) = \exp(-(\exp(\mathbf{x}\beta)t)^\gamma)$$

$$\lambda(t) = \gamma\alpha(\mathbf{x})^\gamma t^{\gamma-1} = \gamma(\exp(\mathbf{x}\beta))^\gamma t^{\gamma-1}$$

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

91 / 264

Exemplo no R - Weibull

```
> modeloweib <- survreg(Surv(Leite$tempo, Leite$status)~1,
  data=leite, dist="weib")
> summary(modeloweib)

Call:
survreg(formula = Surv(Leite$tempo, Leite$status) ~ 1, data = leite,
  dist = "weib")
      Value Std. Error      z      p
(Intercept)  1.713    0.180  9.54 1.38e-21
Log(scale)   -0.415    0.209 -1.99 4.70e-02
Scale= 0.66
Weibull distribution
Loglik(model)=-37.5  Loglik(intercept only)=-37.5
Number of Newton-Raphson Iterations: 6
n= 15

T ~ Weibull(alpha, gamma)
alpha = exp(-1.713) = 0.18  e  gamma = 1/0.66 = 1.51
S(t) = exp(-(\alpha t)^\gamma) = exp(-(\alpha t)^{1.51})
lambda(t) = \gamma \alpha^\gamma t^{\gamma-1} = 1.51 \times 0.18^{1.51} t^{0.51} = 0.11 t^{0.51}
```

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

90 / 264

Exemplo regressão paramétrica no R

```
> dialise <- read.csv("dialise.csv", header=T)
> regexp<-survreg(formula=Surv(tempo, status) ~ idade,
  data=dialise, dist="exponential")
> summary(regexp)

Call:
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
  dist = "exponential")
      Value Std. Error      z      p
(Intercept)  6.136    0.10585  58.0 0.00e+00
idade        -0.037    0.00174 -21.3 1.36e-100
Scale fixed at 1

lambda(t|idade) = exp(-6.14 + 0.037idade)
lambda(t|a1=70) = exp(beta+30beta) = 0.02886 = 4, 39
lambda(t|a1=30) = exp(beta+70beta) = 0.00657
```

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

92 / 264

```
> regweib<-survreg(formula=Surv(tempo,status) ~ idade,
  data=dialise, dist="weibull")
> summary(regweib)
```

Call:

```
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
  dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	6.7512	0.14693	45.95	0.00e+00
idade	-0.0436	0.00224	-19.50	1.12e-84
Log(scale)	0.1987	0.02083	9.54	1.49e-21
Scale=	1.22			

$$\gamma = 1/1.22 = 0.82$$

$$\lambda(t) = 0.82 \times t^{-0.18} \times (\exp(-6.75 + 0.04idade))^{0.82}$$

Para cada ano a mais de idade, o risco relativo de morrer aumenta 4% ($\exp(0,04) = 1,0408$).

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 **Modelo de Cox**
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Inferência

- Teste de Wald $z = \frac{EP(\hat{\beta})}{\hat{\beta}}$
 - H_0 : parâmetro β da regressão é igual a zero
- Análise da função desvio
 - estatística global do ajuste do modelo $D = 2(l_{\text{modelo}} - l_{\text{null}})$
 - comparar modelos aninhados $D = 2(l_{\text{maior}} - l_{\text{menor}})$

Riscos Proporcionais

O modelo de regressão mais amplamente utilizado para dados de sobrevida ajusta a função de risco $\lambda(t)$, considerando um risco basal $\lambda_0(t)$ e incluindo o vetor de covariáveis \mathbf{x} , de forma que:

$$\begin{aligned} \lambda(t|\mathbf{x}) &= \lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) \\ &= \lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta}) \end{aligned}$$

Ou seja, as covariáveis têm um efeito multiplicativo na função de risco.

Riscos Proporcionais

A razão entre os riscos de ocorrência do evento de dois indivíduos i e j , com covariáveis $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ e $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ é:

$$\frac{\lambda_i(t|\mathbf{x}_i)}{\lambda_j(t|\mathbf{x}_j)} = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{\exp(\mathbf{x}_j\boldsymbol{\beta})}$$

Observe que esta razão de riscos **NÃO** varia ao longo do tempo \rightarrow

Modelo de Riscos Proporcionais

Modelo de Cox

Partindo desta proporcionalidade, é possível estimar os efeitos das covariáveis sem qualquer suposição a respeito da distribuição do tempo de sobrevida, e por isso o modelo de Cox é dito semi-paramétrico.

Não se assume qualquer distribuição estatística para a função de risco basal, $\lambda_0(t)$, apenas que as covariáveis agem multiplicativamente sobre o risco e esta é a parte paramétrica do modelo.

Riscos Proporcionais

O modelo RP também pode ser escrito em termos da função de risco acumulado ou da função de sobrevida:

$$\begin{aligned}\Lambda(t|\mathbf{x}) &= \Lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta}) \\ S(t|\mathbf{x}) &= [S_0(t)]^{\exp(\mathbf{x}\boldsymbol{\beta})}\end{aligned}$$

O risco acumulado basal é $\Lambda_0(t) = \sum_{i: t_i \leq t} \frac{N_i(t)}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j\boldsymbol{\beta})}$ e a sobrevida basal é dada por $S_0(t) = \exp[-\Lambda_0(t)]$

Pressupostos do modelo de Cox

- As covariáveis agem multiplicativamente sobre o risco \rightarrow parte paramétrica do modelo.
- A razão de riscos é constante ao longo de tempo \rightarrow riscos proporcionais.
- Os tempos de ocorrência do evento são independentes.
- Censuras são independentes do evento de interesse

Estimativa dos coeficientes

Para estimar os coeficientes da regressão paramétrica, a função de verossimilhança foi construída a partir da função de densidade de probabilidade calculada nos tempos de ocorrência do evento, multiplicada pela função de sobrevivência calculada nos tempos de censura.

No Modelo de Cox o vetor de parâmetros β é estimado a partir de uma **verossimilhança parcial**.

De forma semelhante ao Kaplan Meier, considera-se apenas, a cada tempo t , a informação dos indivíduos sob risco, estimando os efeitos das covariáveis no tempo de sobrevivência.

Verossimilhança parcial

- Sob o processo de contagem a verossimilhança individual é igual a

$$L_i = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{\sum_{t \geq 0} Y_j(t) \exp(\mathbf{x}_j \boldsymbol{\beta})},$$

- com $Y_j(t)$ igual a 1 se o indivíduo j estiver em risco no tempo t e 0, caso contrário.

Verossimilhança parcial

- Considere m diferentes tempos até a ocorrência de um evento (sem empate), ordenados assim: $t_1 < t_2 < \dots < t_m$.
- A verossimilhança individual, $L_{i,t}$, é a razão entre o risco $\lambda_i(t_k)$ do indivíduo i falhar em t_k e a soma dos riscos de ocorrência de evento de todos os indivíduos em risco:

$$L_{i,t} = \frac{\lambda_i(t_k)}{\sum_{j \in R(t_k)} \lambda_j(t_k)} \\ = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{\sum_{j \in R(t_k)} \exp(\mathbf{x}_j \boldsymbol{\beta})}$$

Verossimilhança Parcial

- A verossimilhança parcial $L(\boldsymbol{\beta})$ = produto das $L_{i,t}$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\mathbf{x}_i \boldsymbol{\beta})}{\sum_j Y_j(t) \exp(\mathbf{x}_j \boldsymbol{\beta})} \right\}^{dN_{i,t}}$$

- $dN_{i,t}$ = diferença entre a contagem de eventos até o instante t e a contagem no momento imediatamente anterior a t .
- Numerador depende apenas da informação dos indivíduos que experimentam o evento
- Denominador utiliza informações a respeito de todos os indivíduos que ainda não experimentaram o evento, incluindo aqueles que serão censurados mais tarde.

Exemplo TMO

- Avaliar os fatores prognósticos associados ao tempo de transplante de medula óssea TMO até o óbito nos pacientes com leucemia mielóide crônica tratados no INCA.
- Covariáveis:
 - sexo,
 - idade,
 - fase da doença no momento do transplante (*fase*),
 - a ocorrência ou não de doença do enxerto contra hospedeiro aguda (*deag*) ou crônica (*decr*).

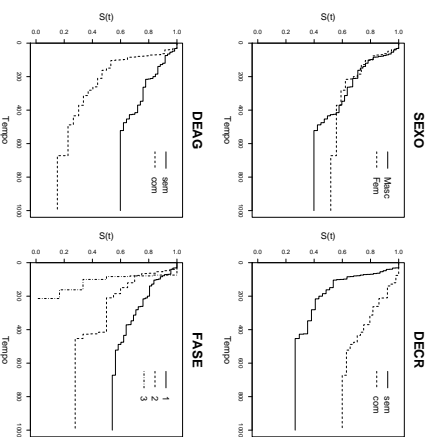
<http://deangue.procc.fiocruz.br/~sobrevida/dados/transplante.html>

Exemplo TMO

```
> tmo<- read.table("tmoclas.dat", header=T, sep=" ")
> head(tmo)
  id sexo idade status  os plaq tempplaq deag tempdeag decr tempdecr fase
1  2  30  30  0 1000  1  9  0  3527  0  3527  0  3527  1
2  2  38  1  39  0  39  0  39  1  28  0  39  1
3  1  23  1  434  1  27  1  23  1  36  1  268  1
4  2  5  1  69  0  69  0  69  1  24  0  69  1
5  2  15  1  672  1  83  1  15  22  1  446  1
6  1  23  1  98  1  98  1  0  1  22  0  98  1
```

Proporcionalidade

Curvas de KM para avaliar o pressuposto de proporcionalidade



No R

```
> tmo <- read.table("tmoclas.dat", header=T, sep=" ")
> mod1 <- coxph(Surv(os, status) ~ idade+factor(sexo), data=tmo, x=TRUE)
> summary(mod1)

Call:
coxph(formula = Surv(os, status) ~ idade + factor(sexo), data = tmo,
      x = TRUE)
n = 96

      coef exp(coef) se(coef)      z      p
idade -0.0186  0.982  0.0141 -1.32  0.19
factor(sexo)2 -0.3299  0.719  0.3219 -1.02  0.31
exp(coef) exp(-coef) lower .95 upper .95
idade 0.982 1.02 0.955 1.01
factor(sexo)2 0.719 1.39 0.383 1.35
```

Selecionando modelos

Teste de Wald

- H_0 parâmetro da regressão β é igual a zero.

$$z = \hat{\beta} / \text{ep}(\hat{\beta}),$$

z segue uma distribuição normal

$\text{ep} =$ erro-padrão

Comparando quatro modelos

```
> anova(mod1, mod2, mod3, mod4, test="Chisq")
```

Analysis of Deviance Table

```
Model 1: Surv(os, status) ~ idade + factor(sexo)
Model 2: Surv(os, status) ~ idade + factor(sexo) + factor(fase)
Model 3: Surv(os, status) ~ idade + factor(sexo) + factor(fase) + deag
Model 4: Surv(os, status) ~ idade + factor(sexo) + factor(fase) + deag +
  decr
Resid. Df Resid. Dev Df Deviance P(<|chi|)
1      94      395.93
2      92      380.78 2      15.14 0.0005146
3      91      366.67 1      14.11 0.0001726
4      90      358.20 1      8.47 0.0036015
```

Selecionando modelos

Análise da Função Desvio ou Teste da Razão de Verossimilhança

- Compara modelos aninhados

$$D = 2(l_{\text{maior}} - l_{\text{menor}}).$$

- l_{maior} = log verossimilhança do modelo com mais covariáveis e l_{menor} = log verossimilhança do modelo com menos covariáveis
- sob H_0 de que não há diferença entre os modelos, segue distribuição χ^2 com graus de liberdade igual à diferença no número de covariáveis dos modelos em questão.

Selecionando Modelos

- A função desvio é assintoticamente semelhante à estatística de Wald quando o número de observações é grande.
- Para número de observações pequenos, a análise da função desvio é mais robusta.
- Outra ressalva a respeito de valores ausentes. Caso eles existam para algumas variáveis incluídas em alguns modelos, mesmo que aninhados, os modelos perdem a comparabilidade.

Medida Global de Ajuste

R^2

- poder explicativo das covariáveis no tempo de ocorrência do evento em estudo.

$$R_{LR}^2 = 1 - \{L(0)/L(\hat{\beta})\}^{2/n}$$

$$= 1 - \exp(2\{l(0) - l(\hat{\beta})\}/n)$$

- Valor mínimo possível de R^2 é zero quando $L(0) = L(\hat{\beta})$
- Valor máximo não é 1 (ou 100%), mas a razão entre as verossimilhanças do modelo saturado e do modelo nulo.

Medida Global de Ajuste

Índice de prognóstico (IP)

- IP é o preditor linear do modelo de Cox, $x\beta$, calculado para cada indivíduo usando as covariáveis observadas e as estimativas dos coeficientes de regressão do modelo ajustado.
- Os indivíduos são estratificados em grupos de tamanhos aproximadamente iguais (grupos de alto, médio e baixo IP, calculados pelos tercís do IP)
- Os valores médios de cada uma das covariáveis dentro de cada grupo são utilizados para obtenção de curvas de sobrevida sob o modelo ajustado.
- Espera-se, se o modelo for razoável, que o gráfico das curvas ajustadas pelo modelo em cada estrato sejam próximas das estimadas por Kaplan-Meier.

Medida Global de Ajuste

Modelo	ln(Verossimil.)	R^2	% Variância Explicada *
Nulo	-199,0424	0,000	0,0%
Saturado	-0,2670	0,984	100,0%
M1: Idade+Sexo	-197,9626	0,022	2,2%
M2: Mod1+Fase	-190,3905	0,165	16,8%
M3: Mod2+deag	-183,3364	0,279	28,4%
M4: Mod3+decr	-179,0992	0,340	34,6%

* $R_{modelo}^2 / R_{saturado}^2$

Medida Global de Ajuste

- Assumindo modelo *mod4*
 - **Indivíduo 1:** sexo masculino (sexo = 0) com 56 anos (idade = 56), na fase intermediária (fase2 = 1 e fase3 = 0), com manifestação de doença do enxerto aguda (deag=1, decr=0)
- $$\beta_{idade} \times 56 = -0,0044 \times 56 = -0,2469$$
- $$\beta_{sexo} \times 0 = -0,2260 \times 0 = 0$$
- $$\beta_{fase2} \times 1 = 0,6413 \times 1 = 0,6413$$
- $$\beta_{fase3} \times 0 = 1,0279 \times 0 = 0$$
- $$\beta_{deag} \times 1 = 1,2530 \times 1 = 1,2530$$
- $$\beta_{decr} \times 0 = -0,9775 \times 0 = 0$$
- Soma = 1,6474

Medida Global de Ajuste

- Assumindo modelo *mod4*
- Indivíduo 2:** sexo feminino (sexo = 1) com 20 anos (idade = 20), na fase avançada (fase2 = 0 e fase3 = 1) com manifestação de doença do enxerto aguda (deag=1, decr=0)

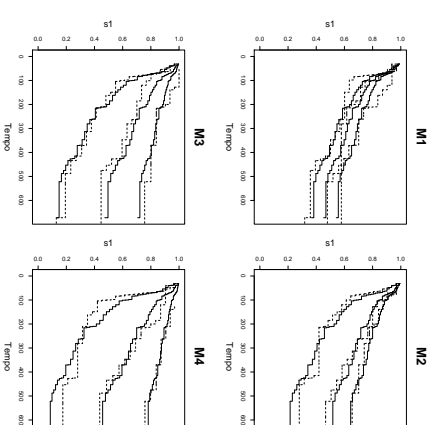
$$\begin{aligned} \beta_{idade} \times 20 &= -0,0044 \times 20 = -0,0882 \\ \beta_{sexo} \times 1 &= -0,2260 \times 1 = -0,2260 \\ \beta_{fase2} \times 0 &= 0,6413 \times 0 = 0 \\ \beta_{fase3} \times 1 &= 1,0279 \times 1 = 1,0279 \\ \beta_{deag} \times 1 &= 1,2530 \times 1 = 1,2530 \\ \beta_{decr} \times 0 &= -0,9775 \times 0 = 0 \\ \hline \text{Soma} &= 1,9667 \end{aligned}$$

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 **Análise de Resíduos**
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Medida Global de Ajuste

Gráfico de sobrevida estratificado por índice prognóstico.



Linha sólida representa o modelo ajustado e linha pontilhada a estimativa de Kaplan-Meier.

Objetivos

Analisar o ajuste e as premissas do modelo de Cox.

- Proporcionalidade dos riscos
- Forma funcional
- Ajuste global
- Pontos influentes e outliers

O resíduo obtido como a diferença entre a resposta observada e a esperada não pode ser usado para os dados de sobrevida: a censuram!!!

Alternativas:

Três tipos de resíduos

- Schoenfeld
- Martingale
- score

Schoenfeld

- Assumindo que não há empates, o resíduo de Schoenfeld é a diferença entre o valor observado de covariável de um indivíduo com tempo de ocorrência do evento $x_k(t_j)$ e o valor esperado da covariável em t_j dado o grupo de risco $R(t_j)$ (e igual a uma média ponderada com pesos iguais ao risco estimado).
- Haverá tantos vetores de resíduos quanto covariáveis ajustadas no modelo, e que estes são definidos somente nos tempos de ocorrência do evento.
- Para cada covariável x_k no tempo de evento t_j , temos

$$r_k(t_j) = x_k(t_j) - \frac{\sum_{t_i \in R_j} x_k(t_i) \exp(\mathbf{x}(t_i)\hat{\beta})}{\sum_{t_i \in R_j} \exp(\mathbf{x}(t_i)\hat{\beta})}$$

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

121 / 264

Schoenfeld

- Suponha que o coeficiente β_k varie com o tempo e seja igual a

$$\beta_k(t) = \beta_k + cU(t)$$
 . $c =$ constante e $U(t) =$ função do tempo

- Podemos padronizar o resíduo de Schoenfeld, $r_k^*(t_j)$, de tal forma que

$$E[r_k^*(t_j) + \beta_k] = \beta_k(t)$$

- Se a premissa de proporcionalidade não é violada esperamos que o gráfico de $r_k^*(t_j)$ versus (t_j) (ou função de (t_j)) apresente uma reta com inclinação zero

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

122 / 264

Schoenfeld no R

```
> residuo <- cox.zph(modelo)
> plot(residuo[1])
> abline(h=modelo$coef[1], lty=2)
#residuo de Schoenfeld padronizado
#gráfico para a primeira covariável
#reta no valor de beta
```

Atenção para a escala do tempo:

- $(1 - S_{km})$ – complementar da função de sobrevida estimada pelo Kaplan-Meier. É a mais utilizada para dados censurados a direita.
 - Tempo t – Útil quando utiliza-se processo de contagem, pode ficar pouco visível se existir uma grande concentração de eventos em um mesmo momento
 - Rank – Ordem dos eventos. Útil quando os tempo observados são muito dispersos ou presença de sobrevida longo
 - Para mudar a escala do tempo veja o help da função `cox.zph()`
- A linha curva é um *Lowess* (regressão não-paramétrica).

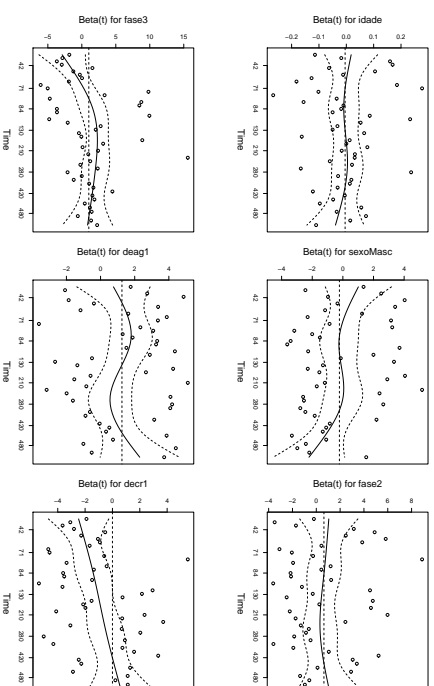
Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

123 / 264

Gráficos de Schoenfeld



Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

124 / 264

Schoenfeld

- Testar H_0 de que a correlação linear entre o resíduo de Schoenfeld e o tempo de sobrevivência é nula
- Equivale a testar H_0 : inclinação igual a zero, ou H_0 : log do risco relativo é constante ao longo do tempo

```
> mod4.zph
              rho      chisq      p
idade      -0.0674    0.2547 0.6138
sexoMasc   -0.2260    2.8393 0.0920
fase2      -0.0317    0.0617 0.8039
fase3       0.2063    2.8416 0.0919
deag1      -0.0147    0.0117 0.9137
decr1       0.3341    6.4078 0.0114
GLOBAL     NA      13.1118 0.0413
```

Resíduos Martingale

É a diferença entre o número observado de eventos para um indivíduo e o número esperado de eventos dado o modelo ajustado.

$$M_i(t) = N_i(t) - E_i(t)$$

em que $E_i = \int_0^{+\infty} Y_i(t) \exp(\mathbf{x}_i \hat{\beta}) d\Lambda_0(t)$

Para o modelo clássico de cox temos

$$M_i = \delta_i - \hat{\Lambda}_0(t_i) \exp(\mathbf{x}_i \hat{\beta})$$

Não proporcionalidade – soluções

- estratificar pela covariável tempo-dependente;
- particionar o eixo do tempo;
- outro tipo de modelo – tempo de vida acelerado, modelos de riscos aditivos;

Resíduos Martingale

Semelhante aos resíduos dos modelos de regressão linear:

- o valor esperado de $M_i = 0$, quando assumimos o valor verdadeiro (mas desconhecido) de β
- o somatório dos resíduos observados $\hat{M}_i = 0$
- os resíduos \hat{M}_i são não correlacionados, mas as estimativas \hat{M}_i são negativamente correlacionadas, ainda que fracamente

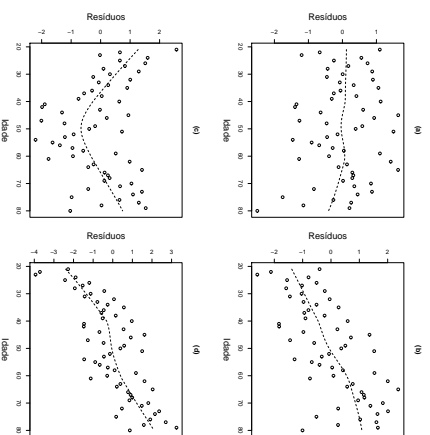
Resíduos Martingale

E diferentes dos resíduos da regressão linear:

- a soma de quadrados dos resíduos não auxilia na avaliação do ajuste global do modelo (o melhor modelo de Cox ajustado não tem a menor soma de quadrados de resíduos martingale);
- a distribuição dos resíduos não é aproximadamente normal;
- o gráfico de resíduos versus valores ajustados não funciona para resíduos martingale pois estes são negativamente correlacionados com os valores ajustados.

Gráficos Martingale

Resíduo do modelo nulo versus idade (dados simulados)

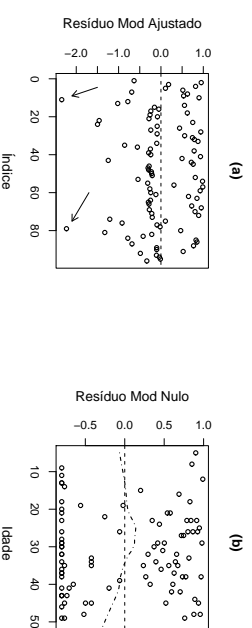


Gráficos Martingale

- M_i versus índice do indivíduo: permite revelar indivíduos mal ajustados pelo modelo;
- M_i do modelo nulo (sem covariáveis) versus covariável com a superposição de uma curva de alisamento: para avaliar a forma funcional da covariável a ser incluída no modelo.

Martingale no R

A função para calcular o resíduo de Martingale é:
`> res <- resid(modelo, type="martingale")`
 em que `modelo` é o objeto que recebeu o modelo de Cox.



(a) modelo `mod4`

(b) covariável X modelo nulo

Ajuste forma funcional não linear

- Incluir uma função de alisamento: *smoothing splines*
- Vantagem sobre polinômios é ser não paramétrica
- São tratadas como covariáveis usuais, inclusive testes de hipótese para não-linearidade
- Permite estimar intervalos de confiança

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

133 / 264

No R

Análise de Resíduos

```
> coxph(formula = Surv(os, status) ~ pspLine(idade, df=0)+
sexo+fase+decr+deag, data=tm0, x=T)
Call: coxph(formula = Surv(os, status) ~ pspLine(idade, df=0)
+ sexo + fase + decr + deag, data = tmocens, x = T)
      coef      se(coef) se2      Chisq DF   p
pspline(idade,df=0),1 -0.0117 0.0157 0.0157 0.56 1.00 0.45000
pspline(idade,df=0),n 12.27 4.63 0.02400
sexo2 -0.2623 0.3445 0.3405 0.58 1.00 0.45000
fase2 0.7428 0.3982 0.3946 3.48 1.00 0.06200
fase3 1.2538 0.5657 0.5593 4.91 1.00 0.02700
decr1 -1.1182 0.3476 0.3444 10.35 1.00 0.00130
deag1 1.4174 0.3592 0.3537 15.57 1.00 0.00008

Iterations: 6 outer, 19 Newton-Raphson      Theta= 0.581
Degrees of freedom for terms= 5,6 1,0 2,0 1,0 1,0
Likelihood ratio test=55.3 on 10.5 df, p=4.38e-08
n= 96
```

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

135 / 264

Smoothing spline

- A idéia é estimar uma curva que produza um bom ajuste aos dados e que ao mesmo tempo não produza uma variação local muito rápida
- Uma maneira de quantificar essa variação é minimizar o seguinte critério

$$\theta \sum_{i=1}^n (y_i - f(x_i))^2 + (1 - \theta) \int [f''(x)]^2 dx$$

- θ é o parâmetro de alisamento
- $\theta \rightarrow 0$ solução converge para uma reta
- $\theta \rightarrow 1$ solução converge para uma curva que passa por todos os pontos (n graus de liberdade)
- O número de graus de liberdade que melhor se adapte pode ser encontrado usando o critério de AIC (Critério de Informação de Akaike)

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

134 / 264

No R

Análise de Resíduos

```
> anova(mod4,mod5, test="Chisq")
Analysis of Deviance Table
Model 1: Surv(os, status) ~
idade + sexo + fase + deag + decr
Model 2: Surv(os, status) ~ pspLine(idade, df = 0)
+ sexo + fase + decr + deag
Resid. Df Resid. Dev Df Deviance P(>|Chi1)
1 90 358.20
2 74 342.74 16 15.46 0.49

Observar:
• os graus de liberdade são fracionários
• o componente linear da idade pspLine(idade,df=0), 1 é não
significativo
• o componente suavizado pspLine(idade,df=0), n tem  $p < 0,05$ 
```

Andreazzi V & Carvalho MS (2008)

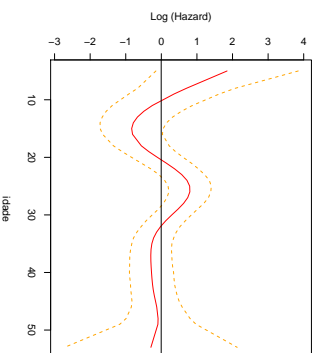
Análise de Sobrevida

Julho/2008

136 / 264

Gráficos de Schoenfeld

Intervalo de confiança praticamente engloba uma a linha reta \rightarrow Não há evidência suficiente para incluir efeito não linear na idade.



Resíduos escore

- Solução

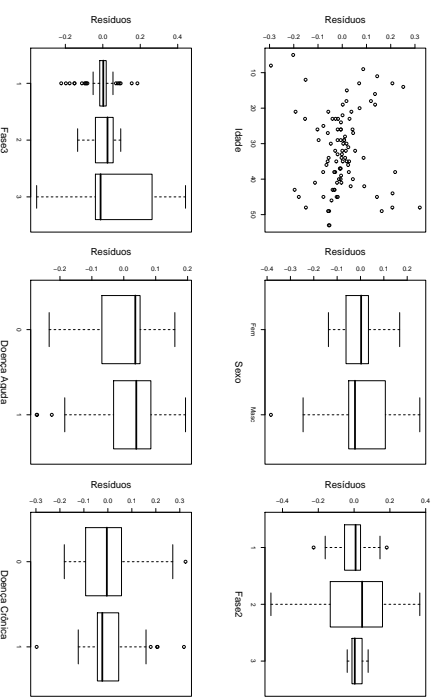
$$\Delta\hat{\beta}_k = (\hat{\beta}_k - \hat{\beta}_k^{(-i)}) \approx \text{resíduo de escore} = DV(\hat{\beta}_k)$$

- D matriz dos resíduos escore
- $V(\hat{\beta}_k)$ matriz de covariância
- O gráfico do resíduo escore para cada covariável $\Delta\hat{\beta}_k$ versus x mostra pontos influentes (alternativa: gráfico dos resíduos de escore versus índice do indivíduo)
- Vantagem – definidos para todos os tempos, mesmo onde não ocorre evento, melhorando a análise de resíduos quando há muita censura

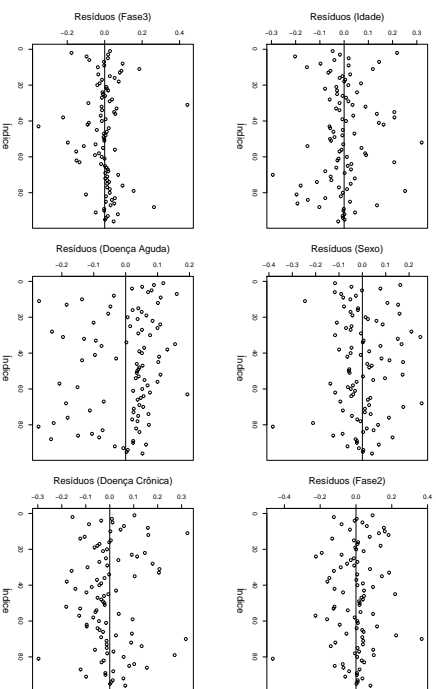
Resíduos escore

- Avalia se uma observação ou um grupo de observações possui uma influência desproporcional an estimativa dos parâmetros
- O ideal:
 - comparar $\hat{\beta}_k$ obtido com todos os dados com $\hat{\beta}_k^{(-i)}$ obtido sem a observação i
 - se $(\hat{\beta}_k - \hat{\beta}_k^{(-i)})$ é próximo de zero, então a observação é pouco influente
 - repetesse esse processo para todas as observações
 - Inviável para dados com muitas observações (implica em estimar $n + 1$ modelos)

Resíduos escore



Resíduos escore



Andreuzzi V & Carvalho MS (2008)

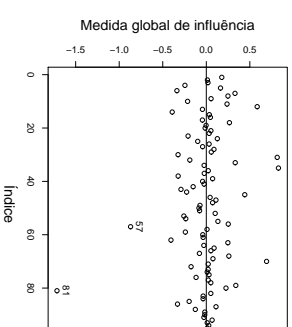
Análise de Sobrevida

Julho/2008

141 / 264

Resíduos escore

Medida Global de Influência \Rightarrow igual a soma dos resíduos escore



Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

142 / 264

Resíduos escore no R

```
> res_esco <- resid(modelo, type="dibetas")
> par(mfrow=c(1,2))
> plot(banco$var1, res_esco[,1], xlab="Var1", ylab="Resíduos")
> plot(banco$var2, res_esco[,2], xlab="Var2", ylab="Resíduos")
```

Observar que o objeto `res_esco` guarda em cada coluna os resíduos referente a cada β das variáveis incluídas no modelo, na ordem em que foram colocadas. Para lembrar

quais são, veja `modelo$call`

Para que o comando `plot()` faça automaticamente o `boxplot`, a covariável tem que estar definida como `factor`

Gráfico versus o índice do indivíduo

```
> plot(1:modelo$n, res_esco[,3], xlab="Índice", ylab="Resíduos")
```

Medida Global de Influência

```
> infl <- apply(mod4_sco2, 1, sum)
> plot(1:modelo$n, infl, xlab="Índice", ylab="Influência")
> abline(h=0)
```

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

143 / 264

Sumário

Para	Fazer
Avaliar o pressuposto de proporcionalidade global	teste de proporcionalidade global fornecido pela função que calcula o resíduo de Schoenfeld
Avaliar o pressuposto de proporcionalidade de cada variável	gráficos do resíduo de Schoenfeld contra o tempo (ou função do tempo)
Estudar a forma funcional da variável	gráficos do resíduo de martingale vs a covariável com a superposição de um alisamento da covariável
Linearizar a forma funcional da variável quando não linear	alisamento <code>spline (pspline())</code> da covariável diretamente no modelo de Cox
Avaliar efeito de valores aberrantes	gráficos de resíduos escore (vs covariável ou índice)

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

144 / 264

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 **Covariável Tempo-dependente**
- 9 Múltiplos Eventos
- 10 Fragilidade

O que muda?

Tudo:

- Idade: terapia antiretroviral, doenças crônicas
- Residência
- Medicamento: crossover, efeitos colaterais
- Hábitos: exercício, alimentação
- Emprego

Objetivos

- Analisar a sobrevida quando as covariáveis mudam ao longo do tempo.
- Construir adequadamente o banco de dados na situação de covariáveis tempo-dependentes.

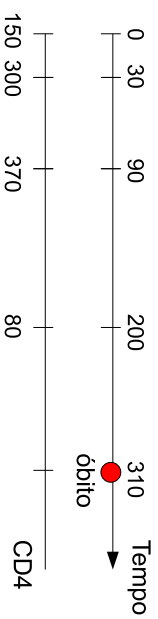
O modelo

$$\lambda(t|\mathbf{x}(t)) = \lambda_0(t) \exp(\mathbf{x}(t)\boldsymbol{\beta})$$

Onde está a diferença?

Exemplo

Indivíduos soropositivos e efeito da contagem de linfócitos CD4 sobre o tempo de sobrevivida.



intervalo	status	CD4
(0,30]	0	150
(30,90]	0	300
(90,200]	0	370
(200, 310]	1	80

Organização dos dados

paciente 1	(0,9+]	(9,1000+]
paciente 2	(0,28+]	(28,39]
paciente 3	(0,27+]	(27,36+]
paciente 4	(0,24+]	(24,69]
paciente 5	(0,22+]	(22,83+]

+ representa censura

(→ intervalo aberto, NÃO inclui o limite inferior
] → intervalo fechado, inclui o limite superior

Organização dos dados

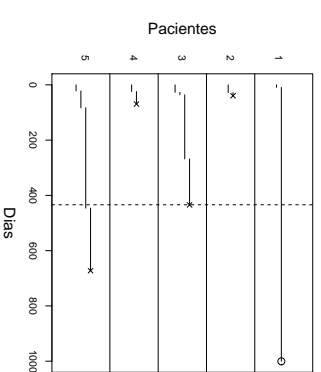
Sobrevida de pacientes submetidos ao TMO

<http://dengue-procc.fiocruz.br/~sobrevida/dados/transplante.html>

id	sexo	idade	status	inicio	fim	deag	decr	recpl	laq	fasegr
1	2	31	0	0	9	0	0	0	0	CP1
1	2	31	0	0	9	1000	0	0	0	CP1
2	2	38	0	0	28	0	0	0	0	CP1
2	2	38	1	1	28	39	1	0	0	CP1
3	1	23	0	0	27	0	0	0	0	CP1
3	1	23	0	0	27	36	0	0	0	CP1
3	1	23	0	0	36	268	1	0	0	CP1
3	1	23	1	1	268	434	1	1	1	CP1
4	2	5	0	0	24	0	0	0	0	CP1
4	2	5	1	1	24	69	1	0	0	CP1
5	2	15	0	0	22	0	0	0	0	CP1
5	2	15	0	0	22	83	1	0	0	CP1
5	2	15	0	0	83	446	1	0	0	CP1
5	2	15	1	1	446	672	1	1	1	CP1

Gráfico da estrutura dos dados de TMO

Quais pacientes estão em risco no tempo=434 dias (linha vertical)?



Estimação

- Os tempos **NÃO** se superpõe.
- A verossimilhança parcial utilizará no máximo uma observação de cada paciente em qualquer momento.
- A soma de indivíduos em risco será feita sobre um conjunto de observações independentes.

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

153 / 264

Exemplo – aids

Estudar o efeito da terapia anti-retroviral de alta potência (Haart) no tempo de sobrevida desde o diagnóstico de Aids até o óbito. Foi registrado a mudança de tratamento (haart = S ou N) ao longo do estudo.

<http://dengue.procc.fiocruz.br/~sobrevida/dados/terapia.html>

Reg	haart	ini	fim	sexo	escol	status	idade	
322010	N	665	804	F	Prim	1	36	
323024	S	1498	1820	M	Univ	0	76	Tempo final da
323024	S	2400	3297	M	Univ	0	76	primeira linha do
325072	N	686	3200	M	Sec	0	33	paciente 323024 é
330476	N	769	1577	M	Sec	0	30	diferente do tempo
330476	S	1577	1597	M	Sec	1	31	inicial da segunda
342863	S	3255	3297	F	Prim	0	52	linha. Por que?
345569	N	1203	1341	M	Prim	1	31	
345815	S	1645	1660	F	Prim	1	32	
348988	N	616	1010	M	Sec	1	31	
351133	N	1191	1627	M	Gin	1	41	

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

154 / 264

Exemplo – aids

```
> muda <- read.table("gafcorr.dat",header=1)
> muda.cox <- coxph(Surv(ini,fim,censura)~haart+idade+
  escol+sexo,data=muda)
> muda.cox
```

Call:

```
coxph(formula = Surv(ini, fim, censura) ~ haart +
  idade + escol + sexo, data = muda)
      coef exp(coef) se(coef)      z      p
haartS    -0.7779    0.459  0.18508  -4.203  2.6e-05
idade     0.0185    1.019  0.00754   2.448  1.4e-02
escolHalf -0.2342    0.791  0.76547  -0.306  7.6e-01
escolGin  0.5364    1.710  0.32688   1.641  1.0e-01
escolPrim 0.7438    2.104  0.31075   2.394  1.7e-02
escolSec  0.3265    1.386  0.33905   0.963  3.4e-01
sexom     0.2253    1.253  0.16929   1.331  1.8e-01
```

Likelihood ratio test=35.1 on 7 df, p=1.08e-05 n= 1377

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

155 / 264

Diagnóstico

Resíduos:

- Schoenfeld:
 - são calculados para os tempos de ocorrência do evento – definição e cálculo sem alteração para processo de contagem
 - valor da covariável utilizado nos cálculos corresponde ao tempo de evento
 - escala *default* é o tempo t (*identity*)
- Martingale:
 - podem ser calculados para cada registro – sem alteração
 - ou para cada indivíduo
- Score: sem alteração.

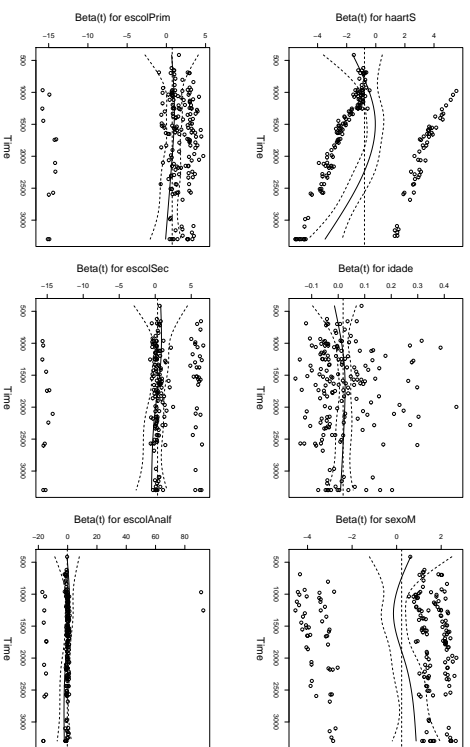
Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

156 / 264

Resíduos Shoenfelf – aids



Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

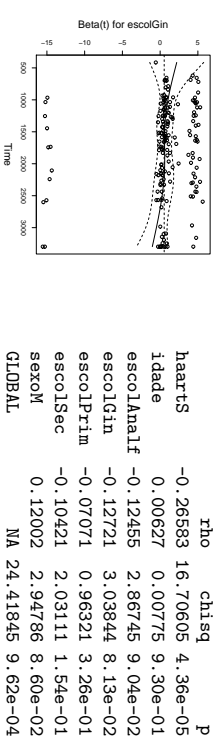
Julho/2008

157 / 264

```
> tmpoc <- read.table("tmpoc.csv", header=T, sep=",")
> tmpoc$sexo <- factor(tmpoc$sexo)
> tmpoc$deag <- factor(tmpoc$deag)
> tmpoc$decr <- factor(tmpoc$decr)
> tmpoc$recplaq <- factor(tmpoc$recplaq)
> tmpoc$recplaq <- relevel(tmpoc$recplaq, "1")
> tmpo.cox <- coxph(Surv(inicio, fim, status)~idade+
  sexo+fasegr+deag+decr+recplaq, data=tmpoc)
> tmpo.cox
```

Exemplo - TMO

Resíduos Shoenfelf – aids



Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

158 / 264

```
Call:
coxph(formula = Surv(inicio, fim, status) ~ idade +
  sexo + fasegr + deag + decr + recplaq,
  data = tmpoc)
      coef exp(coef) se(coef)      Z      p
idade      -0.0208      0.979      0.0141 -1.479 1.4e-01
sexo2       -0.1593      0.853      0.3099 -0.514 6.1e-01
fasegrOther  0.9057      2.474      0.3127  2.897 3.8e-03
deag1       1.0271      2.793      0.2914  3.525 4.2e-04
decr1       0.4392      1.551      0.3859  1.138 2.6e-01
recplaq1    -1.9485      0.142      0.4687 -4.158 3.2e-05
```

Likelihood ratio test=49.2 on 6 df, p=6.68e-09 n= 259

Exemplo - TMO

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

159 / 264

Andreazzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

160 / 264

Resíduos Shoenfeld – TMO

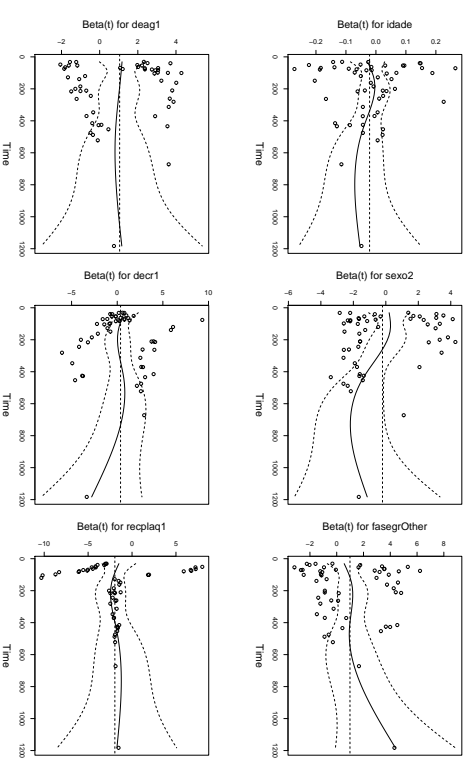
```
> tmo.sch <- cox.zph(tmo.cox)
> tmo.sch
```

	rho	chisq	p
idade	-0.08690	0.45813	0.499
sexo2	-0.25217	3.59294	0.058
fasegrOther	0.15150	1.60183	0.206
deag1	-0.03906	0.07587	0.783
decr1	-0.08257	0.43804	0.508
recplaq1	0.00525	0.00194	0.965
GLOBAL	NA	7.23160	0.300

Resíduos Martingale

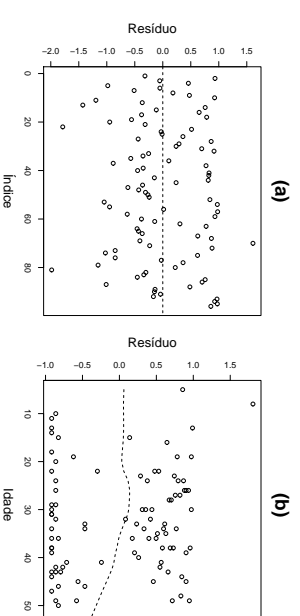
- Podem ser calculados para cada um dos $m > n$ registros
- Ou para cada um dos n indivíduos (resíduo individual = soma dos resíduos do indivíduo em cada intervalo de tempo)
- incluir argumento `collapse=id` para o obter resíduo individual

Resíduos Shoenfeld – TMO



Resíduos Martingale – TMO

Resíduos de martingale para o modelo `tmo.cox` versus índice (a) e para o resíduo do modelo nulo versus idade (b) (covariável contínua).



Resíduos escore

- Permite identificar observações – **períodos de tempo** com valores extremos
- Ou **indivíduos** alavanca
- *collapse=id* – para o indivíduo

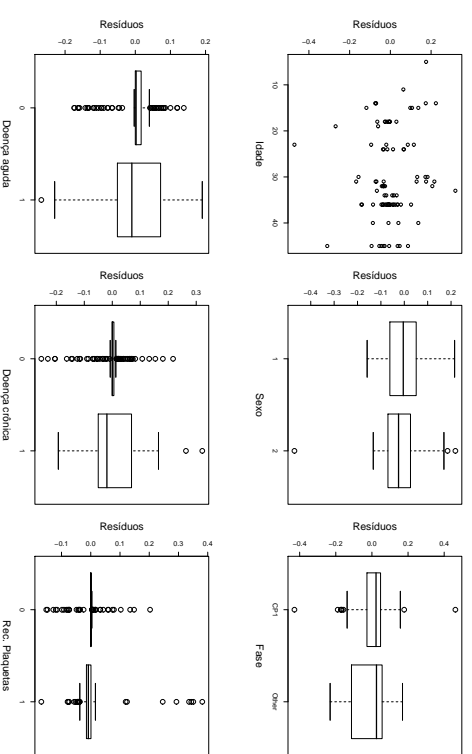
Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

165 / 264

Resíduos escore – TMO



Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

166 / 264

Dados prevalentes e tempo descontínuo

Dados Prevalentes

- Para lidar com intervalos tempo-dependentes precisamos identificar os valores corretos das covariáveis para cada paciente em um dado tempo (construir corretamente o banco de dados)
- O mesmo processo é adotado para dados prevalentes ou truncados à esquerda
- E para representá-lo adequadamente temos que definir como data de referência t_0 a data mais antiga no banco de dados e calcular o tempo de entrada na coorte de cada indivíduo, que terá como limite inferior do seu primeiro intervalo de tempo, o momento de entrada no estudo

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

167 / 264

Dados prevalentes e tempo descontínuo

- Assim cada indivíduo será analisado dentro de sua janela temporal, eliminando o viés potencial da introdução na coorte de **sobreviventes** com tempos mais longos
- E a forma de interpretar os efeitos é condicional – dado que o indivíduo sobreviveu até entrar em observação

Tempo descontínuo

- Podem ocorrer por: ausência de informação, afastamentos por viagem, interrupção, eventos múltiplos (próximo tópico)
- O mesmo mecanismo de registrar os intervalos de tempo (início, fim) permite tratar adequadamente dados de indivíduos com risco descontínuo ao longo do estudo

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

168 / 264

Sumário

Para	Tipo de resíduo	Fazer
Proporcionalidade global	Schoenfeld	<code>cox.zph(modelo)</code>
Proporcionalidade de cada variável	Schoenfeld	gráfico do resíduo de cada variável
Resíduos globais por indivíduo	Martingale de cada indivíduo	<code>resid(model1, type="martingale", collapse=id)</code> ; gráfico contra o índice

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Sumário

Para	Tipo de resíduo	Fazer
Forma funcional	Martingale de cada observação	gráfico do modelo nulo X co-variável
Pontos influentes variável mudando no tempo	Escore de cada observação	<code>resid(model1, type="dfbetas")</code>
Pontos influentes variável que não muda no tempo	Escore de cada indivíduo	<code>resid(model1, type="dfbetas", collapse=id)</code>

Objetivos

- Identificar os diversos modelos existentes para análise de sobrevida com eventos múltiplos.
- Definir conceitualmente os diversos modelos, descrever seus pressupostos e propriedades.
- Indicar o tipo de modelo adequado a cada situação observada.
- Interpretar as estimativas ajustadas em cada modelo.

Exemplos

- Eventos que podem acontecer mais de uma vez para um mesmo indivíduo:
 - gestações,
 - internações,
 - cârries,
 - infartos do miocárdio
 - fraturas.
- Diferentes tipos de eventos decorrentes de um mesmo fator de risco em estudo:
 - efeitos adversos de medicamentos,
 - doenças oportunistas da aids,
 - óbito de um paciente, quando são possíveis várias causas.

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

173 / 264

O que difere

- o registro de mais de um tempo por indivíduo → tempo até cada cârrie, tempo até o óbito, considerando duas possíveis causas competindo
- a utilização direta do modelo de Cox **não** serve, mesmo com processo de contagem, pois os intervalos de tempo de cada indivíduo **não** são independentes entre si
- não havendo independência entre observações → erro na estimação dos parâmetros e dos intervalos de confiança
- composição do grupo de risco → os indivíduos podem (ou não) se manter no grupo de risco mesmo após a ocorrência de algum evento

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

175 / 264

Perguntas

- Quais são os fatores de risco associados aos tempos até as ocorrências de eventos em um mesmo indivíduo?
- Como analisar diferentes eventos como desfecho de uma mesma situação de risco?

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

174 / 264

Estratégias possíveis

- ajustar Poisson → variável dependente é uma contagem dos eventos
 - não diferencia eventos próximos no tempo de distantes
- modelar apenas **um** evento
 - até o primeiro, entre o segundo e o terceiro, ...
 - apenas uma das causas de interesse censurando a outra e vice-versa
 - simplifica, mas pode não ser adequada
- modelos multi-nível (último capítulo)

Modelos Marginais

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

176 / 264

Classificação

O que deve ser observado:

- a população em risco \rightarrow montagem do banco de dados
- o risco basal (λ_0) \rightarrow no modelo semi-paramétrico não é estimado, mas pode ser o mesmo para todos os eventos ou ser diferente conforme o estrato
- a estrutura temporal (ordenação) \rightarrow definida simultaneamente pela forma do banco de dados e pela indicação de presença de estratos ordenados
- a estrutura de dependência entre os eventos \rightarrow métodos robustos para corrigir a variância dos estimadores

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

177 / 264

Eventos ordenados

Combinando definição de grupo sob risco e estrutura do risco basal (λ_0) são três tipos de problemas:

1) **Eventos ordenados independentes** - o risco basal é o mesmo em todos os intervalos de tempo analisados. O indivíduo retorna ao grupo de risco após cada evento.

Ex: IRA. Ordenado, pois não podem ocorrer simultaneamente, e o indivíduo sai do risco na vigência de um episódio.

Modelo de incrementos independentes ou de Andersen-Gill (**AG**)

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

179 / 264

Classificação dos eventos

- **competitivos** \rightarrow só é possível observar o tempo até a ocorrência do primeiro evento que impede a realização de qualquer outro (óbito por diferentes causas e um mesmo fator de risco)
- **paralelos** \rightarrow a ocorrência de um evento não exclui a ocorrência de outro evento e não há qualquer ordem preferencial (doenças oportunistas, efeitos colaterais, perda de dente)
- **ordenados** \rightarrow a sucessão de tempos segue obrigatoriamente uma ordem, que é dada seja pela estrutura de datas de início e fim de cada evento, seja por assumir-se uma ordem nos estratos de risco relacionados a cada desfecho ou ambos

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

178 / 264

Eventos ordenados

2) **Eventos ordenados com risco concomitante** - o indivíduo está sempre em risco de sofrer o primeiro, segundo, terceiro, n-ésimo evento. Cada evento tem um risco basal diferente, dependendo de sua posição na sequência de eventos, mas não do tipo de evento.

Ex. reações adversas (cefaléia, náuseas, tremores e sonolência). O risco basal para cada efeito adverso depende apenas da ordem em que ele aparece, e não do tipo de efeito. O tempo até a ocorrência é contado sempre a partir de zero para todos os indivíduos (não se utiliza a estrutura de contagem).

Modelo **marginal**, também chamado modelo **WLW** (Wei, Lin & Weisfeld)

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

180 / 264

Eventos ordenados

3) **Eventos ordenados com risco condicional** - o indivíduo só entra em risco de sofrer o n -ésimo evento depois que o evento de ordem $n - 1$ tiver ocorrido. O risco basal para um segundo evento é zero até que o primeiro ocorra enquanto que o risco do terceiro evento é zero até que o segundo ocorra.

Ex. IAM. O risco de base se altera à medida em que o indivíduo sofre novos eventos, ou seja, os eventos, além de ordenados no tempo, são ordenados segundo o risco basal.

Modelo condicional, ou modelo **PWP** (Prentice, Williams & Peterson) (todos são marginais!)

Andrezza V & Carvalho MS (2008)

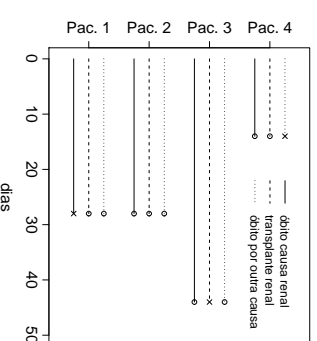
Análise de Sobrevida

Julho/2008

181 / 264

Eventos competitivos

Um dos eventos exclui a possibilidade de ocorrência de outro. Por exemplo, o indivíduo em diálise está em risco para três eventos, a ocorrência de um evento cessa o risco em relação aos demais.



Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

182 / 264

Eventos competitivos – banco de dados

O banco de dados informa quem está em risco a cada momento do tempo e quantos eventos cada indivíduo já experimentou.

id	idade	doença	status	motivo	tempo	endpoint
1	50	outr	1	obrenal	27	1
1	50	outr	0	obrenal	27	2
1	50	outr	0	obrenal	27	3
2	51	hiper	0	censura	28	1
2	51	hiper	0	censura	28	2
2	51	hiper	0	censura	28	3
3	30	diab	0	transpl	18	1
3	30	diab	0	transpl	18	2
3	30	diab	1	transpl	18	3
4	33	diab	0	oboutra	14	1
4	33	diab	1	oboutra	14	2
4	33	diab	0	oboutra	14	3

Andrezza V & Carvalho MS (2008)

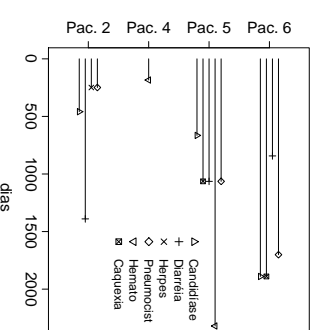
Análise de Sobrevida

Julho/2008

183 / 264

Eventos paralelos

Tempo de sobrevivência de três pacientes portadores do vírus HIV. O tempo foi mensurado desde o acompanhamento do paciente até a ocorrência de algumas doenças oportunistas.



Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

184 / 264

Eventos paralelos – banco de dados

Neste modelo não há ordenação, ou não há interesse em considerar a ordenação entre os eventos.

reg	sex	esc	idade	udi	sexual	ini	oport	fim	status	tempo
2	M	2	38	1	0	2884	Candida	3343	1	459
2	M	2	38	1	0	2884	Diarreia	4274	1	1390
2	M	2	38	1	0	2884	Herpes	3133	1	249
2	M	2	38	1	0	2884	Pneumo	3133	1	249
4	M	3	43	0	0	3780	Hemato	3964	1	184
5	F	3	32	0	1	2923	Candida	3588	1	665
5	F	3	32	0	1	2923	Caquex	3987	1	1064
5	F	3	32	0	1	2923	Diarreia	3987	1	1064
5	F	3	32	0	1	2923	Hemato	5243	1	2320
5	F	3	32	0	1	2923	Pneumo	3987	1	1064
6	M	2	33	0	0	2938	Candida	4828	1	1890
6	M	2	33	0	0	2938	Caquex	4828	1	1890
6	M	2	33	0	0	2938	Diarreia	3785	1	847
6	M	2	33	0	0	2938	Pneumo	4639	1	1701

Andrezza V & Carvalho MS (2008)

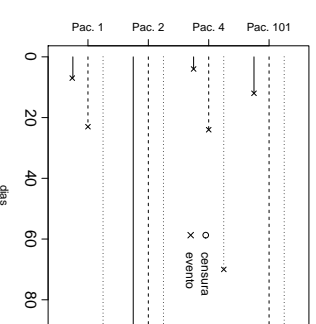
Análise de Sobrevida

Julho/2008

185 / 264

Eventos ordenados

Tempo até o surgimento de efeitos colaterais de medicamentos em quatro pacientes



Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

186 / 264

Eventos ordenados – BD

Ninguém sai do grupo de risco, exceto se 3 eventos ocorrerem.

Linhas do banco de dados = indivíduos X número de eventos

reg	ttto	enum	status	tempo	idade
1	1	1	1	8	25
1	1	1	2	24	25
1	1	3	0	90	25
2	1	1	0	90	29
2	1	2	0	90	29
2	1	3	0	90	29
4	1	1	1	5	26
4	1	2	1	25	26
4	1	3	1	71	26
101	2	1	1	13	24
101	2	2	0	90	24
101	2	3	0	90	24

Andrezza V & Carvalho MS (2008)

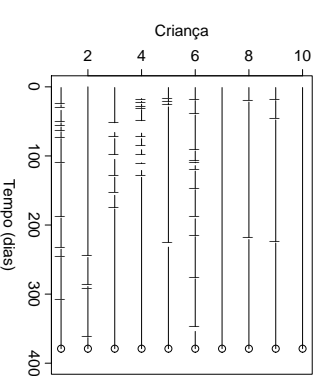
Análise de Sobrevida

Julho/2008

187 / 264

Eventos ordenados

Tempos observados entre episódios recorrentes de diarreia



Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

188 / 264

Eventos ordenados – BD

```

numcri grupo sexo idade ini fim enum status
2 pla fem 48 1 244 1 1
2 pla fem 48 247 286 2 1
2 pla fem 48 289 291 3 1
2 pla fem 48 293 361 4 1
2 pla fem 48 363 379 5 0
5 pla fem 42 1 16 1 1
5 pla fem 42 18 21 2 1
5 pla fem 42 23 26 3 1
5 pla fem 42 29 225 4 1
5 pla fem 42 232 379 5 0

```

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

189 / 264

Modelos marginais

- Modela-se a resposta média como função das covariáveis
- Interpretação igual
- Estimativa robusta da variância por *Jackknife*
- Etapas:
 - identificar conceitualmente o modelo
 - definir está sob risco em cada momento → construir banco de dados
 - ajustar modelo de Cox simples
 - modelar usando mais de uma estrutura

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

191 / 264

Duração do evento

- pontual: sem duração mensurável → o tempo do indivíduo sob risco é contínuo, sem interrupção
- IAM: o indivíduo pode ainda não ter se recuperado de um infarto, e já estar em risco de outro
- com duração: o evento tem início e fim → o tempo do indivíduo sob risco exclui o tempo de duração (diarria)
- evitar a fragmentação do evento

Andrezza V & Carvalho MS (2008)

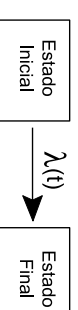
Análise de Sobrevida

Julho/2008

190 / 264

Modelos multiestados

A - Modelo de Cox



Andrezza V & Carvalho MS (2008)

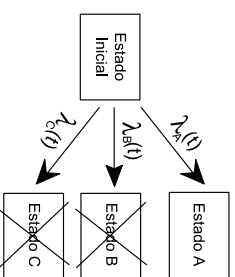
Análise de Sobrevida

Julho/2008

192 / 264

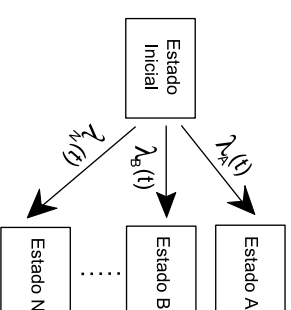
Modelos multiestados

B - Eventos Competitivos



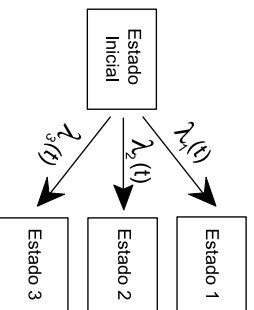
Modelos multiestados

C - Eventos Paralelos



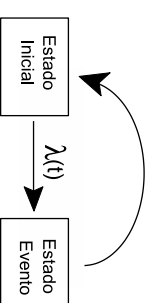
Modelos multiestados

D - Marginal ou WLW



Modelos multiestados

E - Incrementos Independentes ou AG



Modelos multiestados

E - Condicional ou PWP



Eventos Competitivos – diálise

- Três possíveis desfechos:
 - óbito pela doença renal
 - óbito por outra causa
 - transplante
- Será que há interesse em estimar o efeito das variáveis para estes desfechos?
- Ou talvez só para os óbitos por diferentes causas?

Modelo Competitivo – diálise

Call:

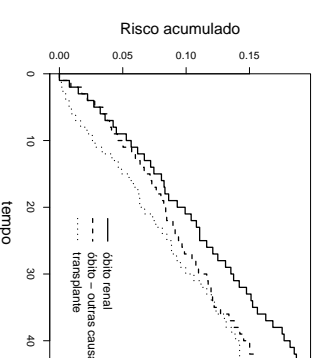
```
coxph(formula = Surv(tempo, status) ~ idade + doenca +
      cluster(id) + strata(endpoint), data = compete)
```

	coef	exp(coef)	se(coef)	robust se	z	p
idade	0.0113	1.011	0.00285	0.00297	3.796	0.00015
doencacong	0.1626	1.177	0.23711	0.22998	0.707	0.48000
doencadiab	0.5089	1.663	0.11609	0.11938	4.263	0.00002
doencaotr	-0.0602	0.942	0.12731	0.12549	-0.479	0.63000
doencarim	-0.1346	0.874	0.09478	0.09407	-1.431	0.15000

Likelihood ratio test=53.8 on 5 df, p=2.28e-10
 n=7359 (12 observations deleted due to missing)

Modelo Competitivo – diálise

Risco acumulado basal para óbito por causa renal, óbito por outras causas e transplante, em pacientes submetidos à hemodiálise



Modelo Competitivo – diálise

```
Call:
coxph(formula = Surv(tempo, status) ~ idade +
      doenca + cluster(id) + strata(endpoint),
      data = compete, subset = endpoint < 3)
```

	coef	exp(coef)	se(coef)	robust se	z	p
idade	0.0319	1.032	0.00330	0.00333	9.566	0.0e+00
doencacong	-0.0300	0.970	0.28540	0.27459	-0.109	9.1e-01
doencaclab	0.5921	1.808	0.12270	0.12658	4.678	2.9e-06
doencaoutr	-0.0853	0.918	0.15155	0.14993	-0.569	5.7e-01
doencarim	-0.3501	0.705	0.12025	0.11953	-2.929	3.4e-03

```
Likelihood ratio test=171 on 5 df, p=0
n=4906 (8 observations deleted due to missing)
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

201 / 264

Interação – diálise

Interação entre a covariável de interesse e o desfecho

```
Call:
coxph(formula = Surv(tempo, status) ~ obrenal.diab +
      oboutr.diab + transp.diab + cluster(id) +
      strata(motivo), data = compete)
```

	coef	exp(coef)	se(coef)	robust se	z	p
obrenal.diab	0.592	1.81	0.155	0.148	4.010	6.1e-05
oboutr.diab	0.365	1.44	0.172	0.151	2.417	1.6e-02
transp.diab	0.259	1.30	0.455	0.343	0.757	4.5e-01

```
[...]
Rsquare= 0.002 (max possible= 0.647 )
[...]
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

202 / 264

Interação – diálise

Interação entre a covariável de interesse e o desfecho: construindo a interação

```
> attach(competete)
> obrenal.diab<-(doenca=="diab")*(motivo=="obrenal")
> oboutr.diab<-(doenca=="diab")*(motivo=="oboutra")
> transp.diab<-(doenca=="diab")*(motivo=="transpl")
> compete<-data.frame(competete,obrenal.diab,
      oboutr.diab,transp.diab)
> compete.fit3<-coxph(Surv(tempo,status)~obrenal.diab+
      oboutr.diab + transp.diab + cluster(id)+
      strata(motivo),data=competete)
> summary(compete.fit3)
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

203 / 264

Eventos paralelos – aids

- o termo *cluster* é necessário para corrigir a variância da estimativa
- para preservar a não ordenação no tempo o modelo deve ser *Surv(tempo, status)*, nunca usando *Surv(Início, fim, status)*
- pode-se ou não estratificar, dependendo do risco de base para cada evento ser igual ou não:

$$\lambda_A(t) = \lambda_B(t)$$

$$\lambda_A(t) \neq \lambda_B(t)$$

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

204 / 264

Banco de dados – eventos paralelos

reg	sex	esc	idade	udi	sexual	ini	oport	fim	status	tempo
2	M	2	38	1	0	2884	Candida	3343	1	459
2	M	2	38	1	0	2884	Diarreia	4274	1	1390
2	M	2	38	1	0	2884	Herpes	3133	1	249
2	M	2	38	1	0	2884	Pneumo	3133	1	249
4	M	3	43	0	0	3780	Hemato	3964	1	184
5	F	3	32	0	1	2923	Candida	3588	1	665
5	F	3	32	0	1	2923	Caquex	3987	1	1064
5	F	3	32	0	1	2923	Diarreia	3987	1	1064
5	F	3	32	0	1	2923	Hemato	5243	1	2320
5	F	3	32	0	1	2923	Pneumo	3987	1	1064
6	M	2	33	0	0	2938	Candida	4828	1	1890
6	M	2	33	0	0	2938	Caquex	4828	1	1890
6	M	2	33	0	0	2938	Diarreia	3785	1	847
6	M	2	33	0	0	2938	Pneumo	4639	1	1701

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

205 / 264

Modelo eventos paralelos

Pode-se ajustar um modelo para cada doença oportunista, estimando os diferentes riscos para cada uma delas:

```
> hem_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Hemato"))
> her_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Herpes"))
> can_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Candida"))
> caq_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Caquex"))
> dia_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Diarreia"))
> pne_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Pneoclist"))
> tub_fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex +
  factor(esc), data=op, subset=(oport=="Tuberc"))
```

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

206 / 264

Efeitos de covariáveis – eventos paralelos

Efeitos de covariáveis sobre o risco de desenvolver algumas doenças

oportunistas	Covs	hem.fit	her.fit	can.fit	caq.fit	dia.fit	pne.fit	tub.fit
udi	0.820	0.458	0.982	0.628	1.075	0.527	2.578	
sexual	0.746	0.957	0.806	0.732	1.335	0.709	0.799	
idade	1.004	1.011	1.012*	1.020*	1.015	0.995	1.017	
sex	1.374*	0.730	1.090	0.939	0.668*	1.193	1.086	
esc1	0.731	0.443	0.745	1.644	0.777	2.222*	4.776	
esc2	0.739	0.517	0.757	1.308	0.689	2.950*	4.622	
esc3	0.722	0.534	0.632	1.494	0.653	3.256*	6.289	
esc4	0.509	0.453	0.622	NA	0.620	NA	1.262	

*Efeito significativo ao nível de 10%

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

207 / 264

Modelo para eventos paralelos

```
Call:
coxph(formula = Surv(tempo, status) ~ udi +
  sexual + idade + sex + factor(esc) +
  cluster(reg), data = op)
```

```
[...]
          exp(coef) exp(-coef) lower .95 upper .95
udi          0.800      1.250      0.586      1.09
sexual       0.829      1.206      0.627      1.10
idade        1.008      0.992      0.997      1.02
sexM         1.097      0.911      0.906      1.33
factor(esc)1 0.892      1.120      0.463      1.72
factor(esc)2 0.894      1.118      0.470      1.70
factor(esc)3 0.861      1.162      0.452      1.64
factor(esc)4 0.613      1.632      0.314      1.19
```

```
Rsquare= 0.025 (max possible= 1 )
[...]
```

Andreuzzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

208 / 264

Eventos paralelos – comparando

- efeito médio de todas as covs não significativo, por que?
- diferentes linhas de base?
- modelo para cada doença separadamente não permite comparar riscos
- ajustar cada doença como um estrato

Modelo marginal - WLW

- eventos são estratificados considerando a ordenação entre eles
- o indivíduo muda de estado com risco $\lambda_1(t)$ de sofrer o primeiro evento, $\lambda_2(t)$ de sofrer o segundo evento e assim sucessivamente
- o tempo é contado sempre a partir do instante zero para cada indivíduo
- a formulação é a mesma dos eventos paralelos, sendo necessário incluir o termo para a correção da variância (`cluster()`), pois os tempos até a ocorrência de cada evento se superpõem

Eventos paralelos – diferentes estratos

```
Call:
coxph(formula = Surv(tempo, status) ~ udi+sexual+idade+sex+
      factor(esc)+cluster(reg)+strata(oport),data = oport)
[...]
```

	exp(coef)	exp(-coef)	lower	.95	upper	.95
udi	0.786	1.273	0.572	1.08	1.08	
sexual	0.811	1.234	0.618	1.06	1.06	
idade	1.007	0.993	0.997	1.02	1.02	
sexM	1.114	0.897	0.934	1.33	1.33	
factor(esc)1	0.930	1.076	0.434	1.99	1.99	
factor(esc)2	0.924	1.082	0.435	1.96	1.96	
factor(esc)3	0.900	1.111	0.424	1.91	1.91	
factor(esc)4	0.620	1.613	0.287	1.34	1.34	

```
Rsquare= 0.027 (max possible= 1 )
[...]
```

WLW – efeitos colaterais

```
Call:
coxph(formula = Surv(tempo, status) ~ idade+factor(ttto)+
      cluster(reg)+ strata(enum),data = colateral)
n= 600
```

	coef	exp(coef)	se(coef)	robust se	z	p
idade	-0.0168	0.983	0.0189	0.0239	-0.703	0.4800
factor(ttto)2	-0.5162	0.597	0.1323	0.1802	-2.864	0.0042

```
exp(coef) exp(-coef) lower .95 upper .95
idade      0.983      1.02      0.938      1.03
factor(ttto)2 0.597      1.68      0.419      0.85
Rsquare= 0.027 (max possible= 0.981 )
[...]
```

Incrementos independentes – AG

- o indivíduo retorna ao estado inicial sempre com o mesmo risco $\lambda(t)$ de sofrer qualquer evento
- pressuposto de risco igual para qualquer evento assume que o histórico do indivíduo não afeta seu risco presente
- Ex: risco de engravidar, diarreias, IRAs.
- a formulação é a mesma dos eventos paralelos, sendo necessário incluir o termo para a correção da variância (`cluster()`), pois os tempos até a ocorrência de cada evento se superpõem
- não se usa *strata* – a ordenação é dada pelo `Surv(ini, fim, status)`

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

213 / 264

Modelo condicional ou PWP

- condicional apenas no nome, porque considera-se que o evento de ordem j somente pode acontecer se o de ordem $j - 1$ tiver acontecido
- não é condicional na forma de estimar – também é marginal

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

215 / 264

Incrementos independentes – diarreia

```
Call:
coxph(formula = Surv(ini, fim, status) ~ grupo+sexo+idade+
      cluster(numcri), data = diarreia)
n= 5592

      coef exp(coef) se(coef) robust se      z      p
grupopla  0.1621    1.176  0.02911  0.06394  2.536  0.011
sexomasc  0.0597    1.061  0.02914  0.06398  0.933  0.350
idade    -0.0339    0.967  0.00122  0.00256 -13.265  0.000

exp(coef) exp(-coef) lower .95 upper .95
grupopla  1.176    0.850    1.037    1.333
sexomasc  1.061    0.942    0.936    1.203
idade     0.967    1.035    0.962    0.971
Rsquare= 0.135 (max possible= 1 )
```

Para avaliar se este modelo é adequado, comparar **se(coef)** e **robust se**. Se forem muito diferentes é porque deve haver uma estrutura de correlação entre os tempos, logo não seriam incrementos independentes.

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

214 / 264

Modelo PWP – diarreia

```
Call:
coxph(formula = Surv(ini, fim, status) ~ grupo+sexo+idade+
      strata(enum)+cluster(numcri), data = diarreia)
n= 5592

      coef exp(coef) se(coef) robust se      z      p
grupopla  0.0470    1.048  0.02991  0.03574  1.31  0.19
sexomasc  0.0529    1.054  0.02974  0.03571  1.48  0.14
idade    -0.0160    0.984  0.00132  0.00153 -10.48  0.00

exp(coef) exp(-coef) lower .95 upper .95
grupopla  1.048    0.954    0.977    1.124
sexomasc  1.054    0.948    0.983    1.131
idade     0.984    1.016    0.981    0.987

Rsquare= 0.027 (max possible= 0.999 )
Likelihood ratio test= 155 on 3 df, p=0
Wald test = 112 on 3 df, p=0
Score (logrank) test = 154 on 3 df, p=0, Robust = 105 p=0
```

Andrezzi V & Carvalho MS (2008)

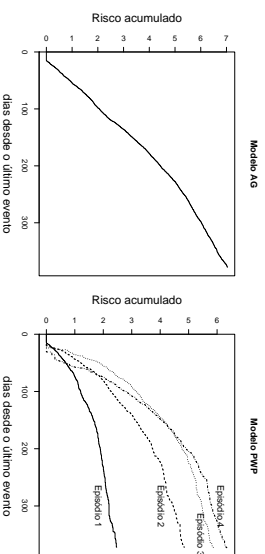
Análise de Sobrevida

Julho/2008

216 / 264

Modelo PWP – diarreia

Risco acumulado basal para ocorrência de novo episódio diarreico



Ajuste e resíduos

Tudo igual

Não existe modelo perfeito!

- Pensar bem sobre o modelo conceitual para o problema (estratificação, covariáveis tempo-dependentes, etc).
- Montar o banco de dados é trabalhoso e sujeito a erros, defina cuidadosamente qual o modelo que será ajustado.
- Sempre verificar no banco:
 - a estrutura temporal,
 - a variável indicativa de censura,
 - o tempo que indica término do estudo,
 - as mudanças de covariáveis no tempo,
 - a variável indicativa de estrato.
- Ajustar o modelo de Cox simples.
- Corrigir a estimativa de variância, calculando uma variância robusta, caso necessário.
- Introduzir a estratificação.

Sumário

Principais características dos modelos de eventos não ordenados.

Modelo	Desfecho	Risco de base
Eventos com-petitivos	Desfechos são mutuamente exclusivos	Não varia de um desfecho para o outro
Eventos par-alelos	Desfechos independentes, em número variável	Pode variar de um desfecho para o outro

Sumário

Principais características dos modelos de eventos ordenados.

Modelo	Desfecho	Risco de base
Marginal	Desfechos ordenados tratados como se fossem um problema de eventos competitivos	Varia de um evento para o outro
Incrementos independentes	Supõe observações mutuamente pendentes para um indivíduo	Não varia de um evento para o outro
AG		
Condicional	Assume estrutura de dependência entre os eventos	Associa a cada evento, um diferente estrato, em que a função de risco basal varia
PWP		

Banco de Dados – eventos não ordenados

Evento	ind	inicio	fim	tempo	desfecho	status
Competitivo	1	1	10	9	óbito	1
	1	1	10	9	transplante	0
	2	1	15	14	óbito	0
	2	1	15	14	transplante	1
	3	8	20	12	óbito	0
	3	8	20	12	transplante	0
	Pararolo	1	1	10	9	doença1
	1	1	12	11	doença2	1
	1	1	16	15	doença3	1
	2	1	15	14	doença1	1
	2	1	15	14	doença5	1
	3	8	20	12	doença7	1

Banco de Dados – eventos ordenados

Modelo	ind	inicio	fim	status	enum
AG (incremento independente)	1	0	12	1	1
	1	12	33	1	1
	1	33	54	1	1
	2	3	20	1	1
PWP (condicional)	2	20	43	0	0
	1	0	12	1	1
	1	12	33	1	1
	1	33	54	1	1
WLLW (marginal)	2	3	20	1	1
	2	20	43	0	0
	1	0	12	1	1
	1	0	33	1	1
	1	0	54	1	1
	2	0	20	1	1
	2	0	43	1	1
	2	0	43	1	1

Grupo de risco

Quem está sob risco no dia 38 quando um indivíduo experimentava o segundo evento?

- AG: todos os indivíduos em observação no dia 38.
- PWP: todos os indivíduos em observação no dia 38 que não tenham experimentado o segundo evento e tenham experimentado o primeiro evento.
- WLLW: todos os indivíduos que estão em observação no dia 38 e não tenham ainda experimentado o terceiro evento.

Sintaxe no R

- AG:
`coxph(Surv(inicio, fim, status) ~ covariáveis+ cluster(id), data=banco`
- PWP:
`coxph(Surv(inicio, fim, status) ~ covariáveis+ cluster(id)+strata(enum), data=banco`
- WLW:
`coxph(Surv(tempo, status) ~ covariáveis+ cluster(id)+strata(enum), data=banco`

Programa

- 1 Introdução
- 2 O Tempo
- 3 Funções de Sobrevida
- 4 Estimação Não-Paramétrica
- 5 Modelagem Paramétrica
- 6 Modelo de Cox
- 7 Análise de Resíduos
- 8 Covariável Tempo-dependente
- 9 Múltiplos Eventos
- 10 Fragilidade

Objetivos

- Conceituar fragilidade ou efeito aleatório.
- Indicar adequadamente a inclusão de efeitos aleatórios na modelagem da sobrevida.
- Identificar as distribuições mais utilizadas e as formas de estimação do efeito aleatório.
- Ajustar os modelos utilizando o R.
- Interpretar as estimativas obtidas no ajuste de modelo com efeitos aleatórios.

Algo não medido....

- Características importantes não medidas
- Características não mensuráveis
- Diferenças nos serviços de saúde prestados (utilizados)

heterogeneidade, correlação

Exemplo – Hemodiálise

Estudar a sobrevivência de pacientes em hemodiálise tratados no RJ

Indivíduo

- Sexo, idade
- Causa da doença renal
- Características sócio-econômicas (reflete acesso aos serviços de saúde)
- Características nutricionais

Centros de diálise

- Oferta de serviço
- Público x Privado

Efeitos Aleatórios

Como levar em consideração na análise: **heterogeneidade** devido a informação não observada e **correlação intra-grupo**?

No exemplo da hemodiálise:

- Características sócio-econômicas e nutricionais não foram observadas
- Heterogeneidade dos pacientes ⇒ Modelos com superdispersão

O que fazer?

Exemplo – Hemodiálise

DADOS: Sistema de Informação Ambulatorial de Alto Custo (APAC), 1998 a 2001 (mensal), 44 meses no Estado do Rio de Janeiro

Indivíduo

Idade e sexo
Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro

No de máquinas: 11 a 19 ou 20 e mais
Número de pacientes: até 50 e maior
Proporção de pacientes idosos, oferta de diálise cídica, ...

Efeitos Aleatórios

O que fazer?

- Obter mais informação (nem sempre possível)
- Considerar **efeitos aleatórios individuais** para tratar a superdispersão intra-grupo
- Considerar **efeito aleatório de grupo** para tratar da correlação intra-grupo
 - não melhoraram a compreensão dos mecanismos que influenciam a sobrevivência
 - mas aumentam a confiança nas estimativas sobre o que foi possível medir

Outros Exemplos

- Estudos multicêntricos
- Dados agrupados (mesma família, gêmeos, hospitais)
- Eventos recorrentes
- Estudo pareado

Modelo de fragilidade

- Z é uma variável aleatória desconhecida que reflete a fragilidade (peculiaridade) do indivíduo ou do grupo
- média igual a 1 e variância ξ
- ξ grande \Rightarrow variabilidade não atribuível a covariável observada
- ξ pequeno \Rightarrow pouca heterogeneidade, não é necessário incluir fragilidade

Efeitos Aleatórios

- Atuam multiplicativamente sobre o risco de base, da mesma forma que as covariáveis.
- É possível incluir dois níveis de efeito aleatórios, por exemplo, o do indivíduo e o da unidade, mas a estimação fica complicada.

Modelo de fragilidade

sendo
 $\mathbf{x} \Rightarrow$ covariáveis
 $\beta \Rightarrow$ efeitos fixos
 $Z = z \Rightarrow$ efeitos aleatórios

$$\lambda(t|\mathbf{x}) = z\lambda_0(t)\exp(\mathbf{x}\beta)$$

- Fragilidade $z_i > 1 \Rightarrow$ indivíduo (ou grupo) i tendem a experimentar o evento com uma taxa mais rápida que no modelo de Cox básico.
- Variância ξ se a aproxima de 0 \Rightarrow modelo de Cox.

Quem é Z – Variável aleatória com distribuição "sensata"

Distribuição da fragilidade precisa ter plausibilidade biológica:

- *gamma* – a distribuição da fragilidade entre os que sofreram o evento em um dado tempo e entre os sobreviventes também é *gamma* (resultado teórico demonstrável)
- *Lognormal* – escolha quando fragilidades relacionadas a covariáveis perdidas x_{m_i} : dado $Z = \exp(x_{m_i}\beta_{m_i})$, ao aplicarmos o teorema do limite central a $x_{m_i}\beta_{m_i}$, teremos uma distribuição lognormal para Z .
- Ambas distribuições muito flexíveis, caracterizadas por dois parâmetros, variam desde forma exponencial até forma de sino similar à curva normal

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

237 / 264

Quem é Z – *gamma*

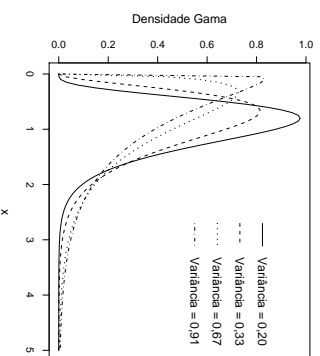


Figura: Distribuição gama com valor esperado 1 e diferentes variâncias ξ

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

239 / 264

Quem é Z – *gamma*

- γ é parâmetro de forma
- α é parâmetro de escala

$$E(Z) = \gamma\alpha$$

$$Var(Z) = \gamma\alpha^2.$$

Sendo

$$\gamma = 1/\xi$$

$$\alpha = \xi$$

Então

$$E(Z) = 1$$

$$Var(Z) = \xi$$

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

238 / 264

Quem é Z – *lognormal*

- μ é parâmetro de locação
- σ^2 é parâmetro de escala

$$E(Z) = \exp(\mu + \sigma^2/2)$$

$$Var(Z) = \exp(\sigma^2) - 1$$

Se

$$\mu = -\sigma^2/2e$$

$$\sigma^2 = \log(1 + \xi)$$

Então

$$E(Z) = 1$$

$$Var(Z) = \xi$$

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

240 / 264

Quem é Z – lognormal

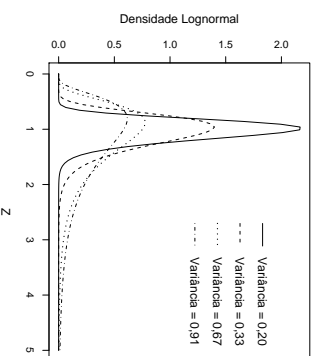


Figura: Distribuição lognormal com valor esperado 1 e diferentes variâncias ξ

Algoritmo EM

- Utiliza-se algoritmo Esperança-Maximização (EM) ou inferência bayesiana.
- Fragilidade tratada como dados não observáveis \Rightarrow estimadas no passo **E** do algoritmo
- $E(Z|t, \delta)$ como estimador.
- No passo **M** são obtidos os valores dos coeficientes de regressão que maximizam a verossimilhança parcial:

$$L_k(\beta) = \left(\frac{z_i \exp(\mathbf{x}_i \beta)}{\sum_{j \in R(t_i)} z_j \exp(\mathbf{x}_j \beta)} \right)^{\delta_i}$$

Quem é Z

- *Lognormal* tem a cauda direita mais pesada do que a *gamma* \Rightarrow proporção maior de indivíduos com riscos muito altos
- cauda esquerda é mais leve que *gamma* \Rightarrow menor a proporção de indivíduos com riscos muito baixos

Algoritmo EM

- Não implementado no R (nem em qq outro)
- Problemas na estimação da variância ξ
- Bom para estimar fragilidade individual

Verossimilhança Parcial Penalizada

- Implementado no R
- Somente para fragilidade compartilhada (ou multinível)

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\omega})$$

- \mathbf{x} e \mathbf{U} são matrizes de covariáveis
- $\boldsymbol{\beta}$ correspondem aos p efeitos fixos
- $\boldsymbol{\omega}$ é um vetor que contém q efeitos aleatórios
- o elemento ω_{ij} da matriz \mathbf{U} é igual a 1 se o indivíduo i é um membro do grupo j , e é igual a zero caso contrário

Ajuste no R

- θ pode ser fixado diretamente
- pode-se especificar graus de liberdade para o termo dos efeitos aleatórios de 0 (máxima penalidade) a q (nenhuma penalidade, sendo q o número máximo de grupos)
- critério de informação de Akaike (AIC)
- a variância da fragilidade gama pode ser estimada usando-se verossimilhança perfilada, com resultado idêntico ao do algoritmo EM
- fragilidades lognormais pode ser estimada com base numa equação de verossimilhança restrita aproximada

Estimação

- Maximização da função de logverossimilhança parcial penalizada em função de $\boldsymbol{\beta}$ e $\boldsymbol{\omega}$.

$$LPP = l(\boldsymbol{\beta}, \boldsymbol{\omega}) - g(\boldsymbol{\omega}, \theta)$$

- g é a função que impõe penalidades a valores de $\boldsymbol{\omega}$ menos desejáveis.
- θ é um parâmetro de ajustamento de g
- Quando:
 - $g(\boldsymbol{\omega}, \theta) = (1/\theta) \sum_i [\omega_i - \exp(\omega_i)] \Rightarrow$ modelo gama de efeitos aleatórios
 - $g(\boldsymbol{\omega}, \theta) = (1/2\theta) \sum_i (\omega_i^2) \Rightarrow$ modelo lognormal, sendo θ a variância dos efeitos $\omega_{i,j}$.

Testes de Hipóteses

- Teste de Wald
 - Teste da Razão de Verossimilhança
- ambas estatísticas seguem uma distribuição qui-quadrado

Fragilidade no R

- Tudo igual, incluindo o termo `frailty`

```
modelo <- coxph(Surv(inicio, fim, status) ~ covariáveis +
  frailty(grupo), data=dados)
```

- `grupo` ⇒ variável que indica o segundo nível, para o qual a fragilidade será estimada

Hemodíalise no R

```
> dial <- read.table("dialmenor.dat", header=T)
> head(dial)

  unidade idade sexo inicio fim status tempo grande causa
1      128   52   1    26  45     0    19     1     1     out
2      128   76   0    32  33     0     2     1     1     out
3      128   61   1    22  24     0     2     1     1     out
4      128   35   0     7  13     0     6     1     1     out
5      128   42   0     2  13     0    11     1     1     out
6      128   44   1     6  30     0    24     1     1     hip
7      128   41   1     1   6     1     5     1     1     out
8      128   39   1    10  13     0     3     1     1     out
9      128   57   0     7  45     0    38     1     1     out
10     128   71   1    16  33     0    17     1     1     out

> dial$causa <- relevel(dial$causa, ref="hip")
```

Clínicas de hemodíalise

Objetivo Avaliar qualidade dos centros de hemodíalise

Dados APAÇ, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis Indivíduo Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais

Número de pacientes: até 50 X 50 ou mais

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Hemodíalise no R

```
> uni.cox <- coxph(Surv(inicio, fim, status) ~ idade+causa,
  data=dial)
> summary(uni.cox)

Call:
coxph(formula = Surv(inicio, fim, status) ~ idade + causa, data = dial)
n = 861

      exp(coef) exp(-coef) Lower .95 upper .95
idade      1.036      0.965      1.025      1.05
causacon   0.366      2.733      0.147      0.91
causadia   1.887      0.530      1.335      2.67
causaout   1.088      0.919      0.684      1.73
causaren   1.422      0.703      0.956      2.12

Rsquare= 0.085 (max possible= 0.929 )
Likelihood ratio test= 76.5 on 5 df, p=4.55e-15
Wald test = 71.6 on 5 df, p=4.82e-14
Score (logrank) test = 76.4 on 5 df, p=4.66e-15
```

Hemodíalise no R

Pacientes podem ter um risco comum associado à unidade de diálise em que são assistidos, a unidade foi incluída como variável *dummy*.

```
> summary(coxph(Surv(inicio, fim, status) ~
  idade+causa+factor(unidade), data=dial))

exp(coef) exp(-coef) lower .95 upper .95
idade      1.03e+00  9.68e-01  1.0220  1.04
causacon   4.41e-01  2.27e+00  0.1354  1.44
causadia   1.78e+00  5.62e-01  1.2371  2.56
causacout  1.20e+00  8.34e-01  0.6694  2.15
causaren   1.22e+00  8.19e-01  0.8039  1.86
factor(unidade) 217  8.84e-01  1.13e+00  0.0995  7.85
[...]
factor(unidade) 641  1.93e+00  5.17e-01  0.6901  5.42
factor(unidade) 741  5.19e+00  1.93e-01  1.4148  19.05
[...]
factor(unidade) 5681  8.70e-07  1.15e+06  0.0000  Inf
factor(unidade) 5692  3.34e+00  3.00e-01  0.2820  39.49
Rsquare= 0.22 (max possible= 0.929 )
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

253 / 264

Hemodíalise no R

Incluindo fragilidade

```
> summary(uni.gama<- coxph(Surv(inicio, fim, status) ~
  idade + causa + frailty(unidade), data=dial))

Call:
coxph(formula = Surv(inicio, fim, status) ~ idade + causa +
  frailty(unidade), data = dial)

n = 861

      coef      se(coef)      se2      Chisq  DF  p
idade      0.0325  0.00555  0.00552  34.26  1.0  4.8e-09
causacon   -0.9333  0.51228  0.48190   3.32  1.0  6.8e-02
causadia    0.5980  0.18320  0.18191  10.66  1.0  1.1e-03
causacout   0.1674  0.29820  0.29033   0.32  1.0  5.7e-01
causaren    0.2462  0.21095  0.20909   1.36  1.0  2.4e-01
frailty(unidade)
123.10 15.6 0.0e+00
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

255 / 264

Hemodíalise no R

- Observe unidade com limite superior igual a infinito (sem ocorrência de óbito)

```
Warning message:
Loglik converged before variable 23 ; beta may be infinite.
in: filter(X, Y, strats, offset, init, control,
weights = weights,
```

- *Rsquare* maior que o modelo sem as unidades (de 0.085 para 0.22)
- Neste modelo não podemos testar a hipótese de que a unidade número 741 tem um risco 5 vezes maior que a primeira unidade (referência) devido a proporção elevada de idosos, pois não é possível incluir covariáveis no nível do centro. (pois os valores não mudariam em relação a variável unidade)

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

254 / 264

Hemodíalise no R

Incluindo fragilidade (continuação)

```
[...]
exp(coef) exp(-coef) lower .95 upper .95
idade      1.033  0.968  1.022  1.04
causacon   0.393  2.543  0.144  1.07
causadia   1.819  0.550  1.270  2.60
causacout  1.182  0.846  0.659  2.12
causaren   1.279  0.782  0.846  1.93

Iterations: 10 outer, 27 Newton-Raphson

Variance of random effect= 0.734  I-likelihood = -1059.8
Degrees of freedom for terms= 1.0  3.8  15.6
Rsquare= 0.216 (max possible= 0.929 )
Likelihood ratio test= 210 on 20.4 df, p=0
Wald test = 54.4 on 20.4 df, p=6.09e-05
```

Andrezzi V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

256 / 264

Hemodiálise no R

Interpretação

`uni.cox`

estima o efeito **médio populacional** da doença de base no risco de morte por insuficiência renal.

compara-se duas amostras aleatórias, por exemplo, pacientes cuja doença de base é diabetes e os hipertensos.

`uni.gama`

estima o efeito da doença de base dentro de uma **unidade típica**.

estima o risco relativo da diabetes dado que os pacientes são assistidos na **mesma unidade típica**, isto é, que possui efeito aleatório igual a zero.

Comparando os modelos

H_0 : efeitos aleatórios = 0

- Estatística de Wald segue distribuição qui-quadrado (`frailty(unidade) = 123.10` com 15.6 gl):
- Razão de verossimilhança segue distribuição qui-quadrado com um grau de liberdade

$$2 \times \loglik(\text{modelo.frail}) - \loglik(\text{modelo.nofrail}) \\ 2 \times (-1059,8 - (-1099,963)) = 80,33$$

`loglik(modelo.frail)` = logverossimilhança parcial integrada fora os termos de fragilidade (`I-likelihood`)
`loglik(modelo.nofrail)` = logverossimilhança do modelo sem fragilidade

- Medida global de ajuste:

`Likelihood ratio test = 210` on `20.4 df`, $p=0$
 Não disponível para modelos Gauss

Estimação

- *Iterations: 10 outer, 27 Newton-Raphson*
 27 iterações necessárias para a variância dos efeitos aleatórios em cada iteração do efeitos fixos (total 10)
- Modelo não é bem estimado quando...

Warning message:

Inner loop failed to converge for iterations 6 in:

```
coxpenal.fit(X, Y, strats, offset, intt = intt,
control, weights = weights,
```

- Porque utiliza-se como padrão o argumento `sparse=TRUE`, que utiliza somente a diagonal da matriz de informação para facilitar os cálculos matriciais. Neste caso, recomenda-se ajustar o modelo sem usar essa opção, com a seguinte sintaxe:

```
> uni.gama<- coxph(Surv(inicio, fim, status) ~
idade+cause+frailty(unidade, sparse= FALSE),
data=dial)
```

Estimando os modelos

- distribuição gama \Rightarrow verossimilhança perfilada para θ (`method=EM`)
- distribuição normal \Rightarrow máxima verossimilhança restrita aproximada (`method=REML`)
- minimização AIC (Critério de Akaike)
- **Não** há método para identificar se uma distribuição é melhor que a outra, ou um método de estimação é melhor que o outro
- Avaliar consistência:
 - mesmas unidades com risco diferenciado
 - estimativas dos efeitos fixos e intervalos de confiança semelhantes

Comparando os modelos

Efeito Gama (EM)			Efeito Gama (AIC)			
	$exp(\beta)$.95Inf	.95Sup	$exp(\beta)$.95Inf	.95Sup
idade	1.033	1.022	1.04	1.033	1.022	1.04
causacon	0.393	0.144	1.07	0.388	0.144	1.04
causadia	1.819	1.270	2.60	1.827	1.277	2.61
causout	1.182	0.659	2.12	1.174	0.656	2.10
causaren	1.279	0.846	1.93	1.293	0.856	1.95
Efeito Gauss (REML)			Efeito Gauss (AIC)			
	$exp(\beta)$.95Inf	.95Sup	$exp(\beta)$.95Inf	.95Sup
idade	1.033	1.022	1.04	1.033	1.022	1.04
causacon	0.419	0.157	1.12	0.415	0.157	1.09
causadia	1.814	1.267	2.60	1.820	1.273	2.60
causout	1.168	0.660	2.07	1.161	0.658	2.05
causaren	1.247	0.824	1.89	1.254	0.829	1.90

Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

261 / 264

Comparando os ajustes

Medidas de ajuste	Cox Classico	Gama EM	Gama AIC	Gauss EM	Gauss AIC
Variancia do efeito aleatorio	—	0,734	0,55	0,697	0,55
Wald Test ^{a,c}	—	123,10 (15,6)	118,81 (14,8)	128,26 (14,7)	123,82 (14,1)
Rsquare	0,085	0,216	0,215	0,216	0,215
Log-verossimil, I-likelihood	-1100	-1033	-1034	-1033	-1034
Qualidade do Ajuste ^{b,c}	76,5 (5)	-1059,8 (20,4)	-1059,9 (19,5)	208 (19,5)	210 (18,8)

^a Teste para a fragilidade (graus de liberdade entre parênteses)^b Teste da razão de verossimilhança (graus de liberdade entre parênteses)^c p-valor <0,001

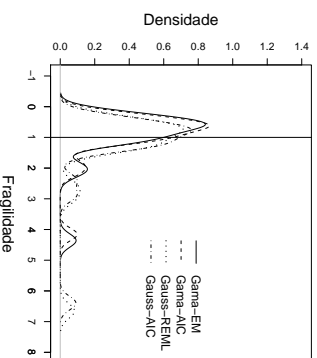
Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

262 / 264

Comparando os efeitos



Observar o escopo da fragilidade – gama chega perto do zero, gauss com valores muito altos e flutuação

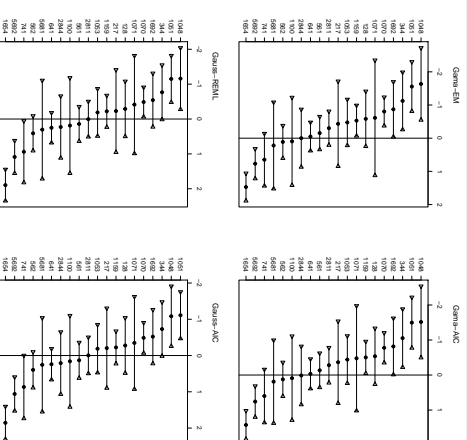
Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

263 / 264

Comparando os efeitos



Andrezza V & Carvalho MS (2008)

Análise de Sobrevida

Julho/2008

264 / 264