

Análise de Sobrevida

Covariável Tempo-dependente

Valeska Andreozzi¹

valeska.andreozzi@fc.ul.pt

&

Marilia Sá Carvalho²

cavalho@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

Programa

1 Covariável Tempo-dependente

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 que muda?

Tudo:

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

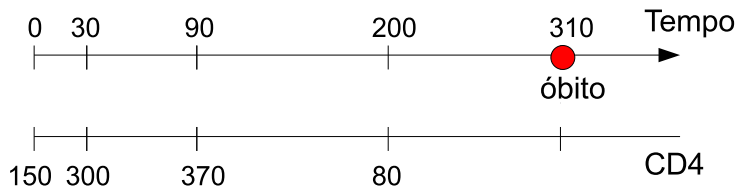
O modelo

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

Onde está a diferença?

Exemplo

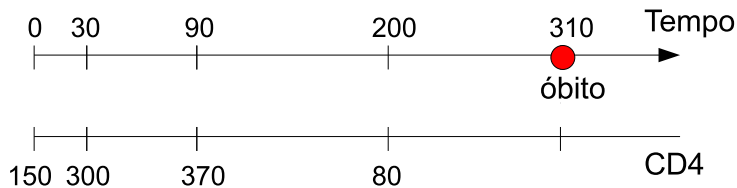
Indivíduos soropositivos e efeito da contagem de linfócitos CD4 sobre o tempo de sobrevivida.



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

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

Sobrevida de pacientes submetidos ao TMO

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

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

Organização dos dados

paciente 1	(0,9+]	(9,1000+]		
paciente 2	(0,28+]	(28,39]		
paciente 3	(0,27+]	(27,36+]	(36,268+]	(268,434]
paciente 4	(0,24+]	(24,69]		
paciente 5	(0,22+]	(22,83+]	(83,446+]	(446,672]

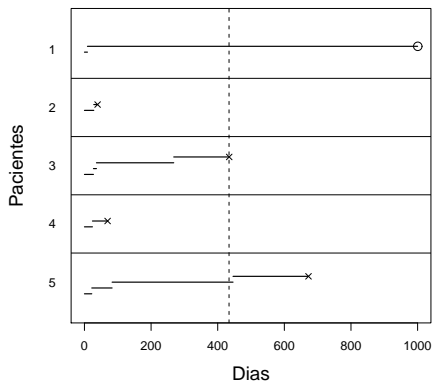
+ representa censura

(→ intervalo aberto, NÃO inclui o limite inferior

] → intervalo fechado, inclui o limite superior

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.

Exemplo – aids

Estudar o efeito da terapia anti-retroviral de alta potência (Haart) no tempo de sobrevivência 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
323024	S	2400	3297	M	Univ	0	76
325072	N	686	3200	M	Sec	0	33
330476	N	769	1577	M	Sec	0	30
330476	S	1577	1597	M	Sec	1	31
342863	S	3255	3297	F	Prim	0	52
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

Tempo final da primeira linha do paciente 323024 é diferente do tempo inicial da segunda linha. Por que?

Exemplo – aids

```
> muda <- read.table("gafcorr.dat",header=T)
> 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
escolAnalf -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
```

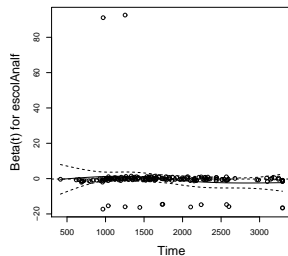
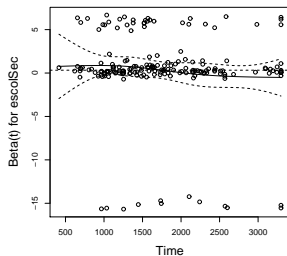
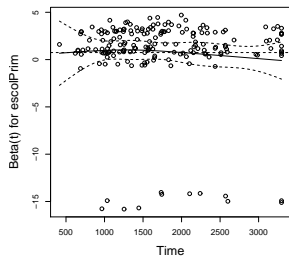
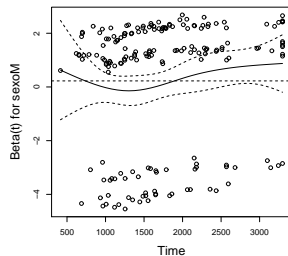
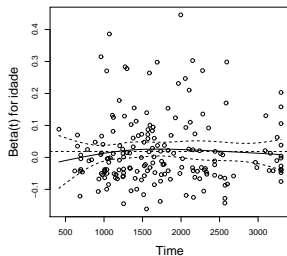
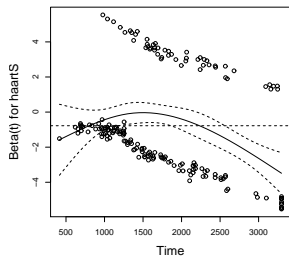
Likelihood ratio test=35.1 on 7 df, p=1.08e-05 n= 1377

Diagnóstico

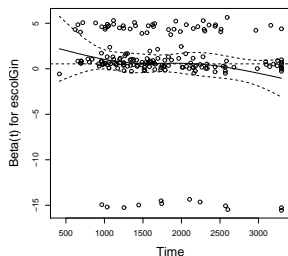
Resíduos:

- 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
- Escore: sem alteração.

Resíduos Shoefeld – aids



Resíduos Shoefield – aids



	rho	chisq	p
haartS	-0.26583	16.70605	4.36e-05
idade	0.00627	0.00775	9.30e-01
escolAnalf	-0.12455	2.86745	9.04e-02
escolGin	-0.12721	3.03844	8.13e-02
escolPrim	-0.07071	0.96321	3.26e-01
escolSec	-0.10421	2.03111	1.54e-01
sexoM	0.12002	2.94786	8.60e-02
GLOBAL	NA	24.41845	9.62e-04

Exemplo - TMO

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

Exemplo - TMO

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ idade +
      sexo + fasegr + deag + decr + recplaq,
      data = tmopc)
```

	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
fasegr0ther	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

Resíduos Schoenfeld – TMO

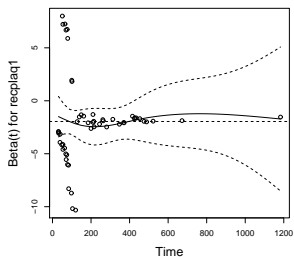
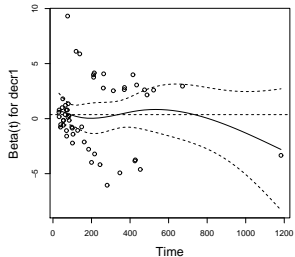
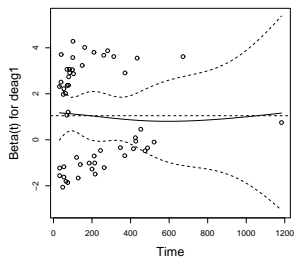
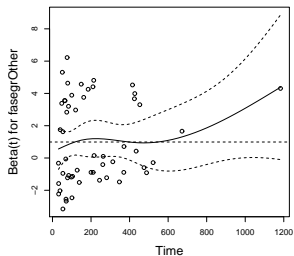
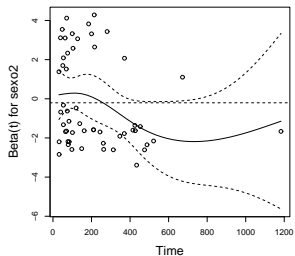
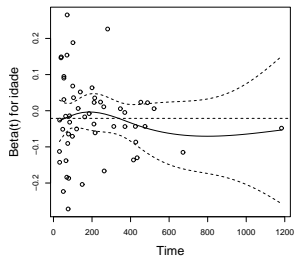
```

> 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
fasegr0ther	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 Shoefield – TMO

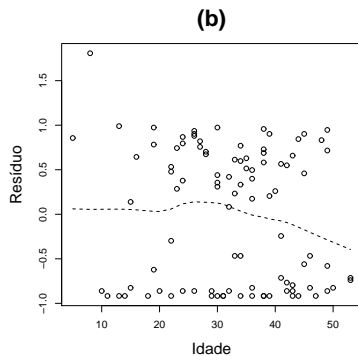
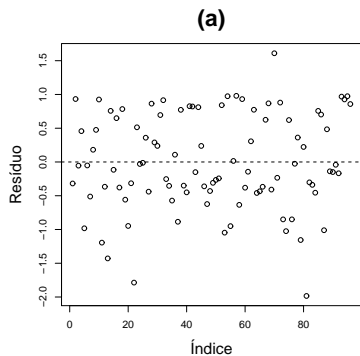


Resíduos Martingale

- 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 Martingale – TMO

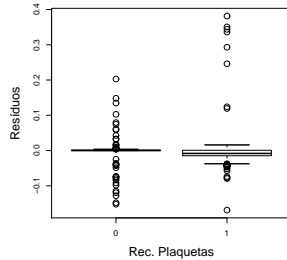
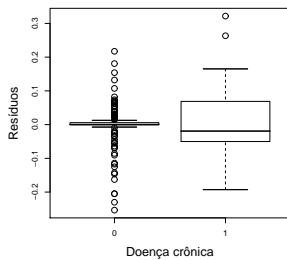
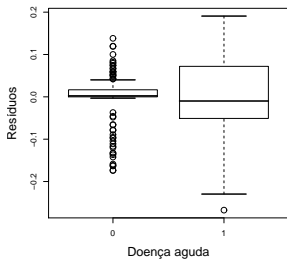
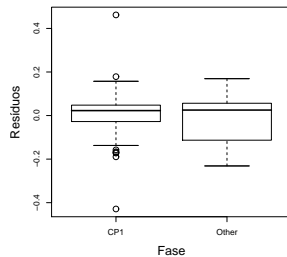
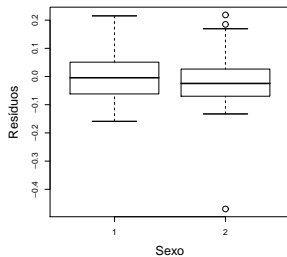
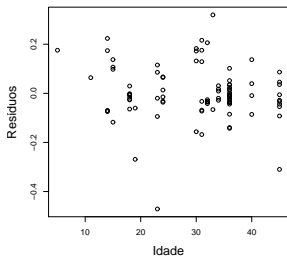
Resíduos de martingale para o modelo `tmo.cox` versus índice (a) e para o resíduo do modelo nulo versus idade (b) (covariável contínua).



Resíduos score

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

Resíduos escore – TMO



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

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

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	Martigale de cada indivíduo	<code>resid(model, type="martingale", collapse=id);</code> gráfico contra o índice

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(model, type="dfbetas")</code>
Pontos influentes variável que não muda no tempo	Escore de cada indivíduo	<code>resid(model, type="dfbetas", collapse=id)</code>