

# Outline

- 1 Cap 1 – Introdução
- 2 Cap 2 – O tempo
- 3 Cap 3 – Funções de Sobrevida
- 4 Cap 4 – Não-Paramétrica
- 5 Cap 8 – Análise de Resíduos**

# Objetivos

Analisar o ajuste e as premissas do modelo de Cox.

São três tipos de resíduos:

- Schoenfeld
- Martingale
- escore

# Pressupostos

- proporcionalidade: a relação entre variável resposta e variável independente do tempo.
- linearidade (log-linearidade, pois a função de risco  $\lambda(t)$  tem uma estrutura log-linear): a razão de riscos entre um indivíduo de 45 anos e um de 50 anos é idêntica àquela entre um indivíduo de 80 anos e um de 85 anos.
- Efeito de pontos influentes (ou de alavanca).
- O resíduo obtido como a resposta observada menos a esperada não pode ser usado para os dados de sobrevida: a censura!!!

# Schoenfeld

$$r_i(\boldsymbol{\beta}) = \mathbf{x}_i - \frac{\sum_{j \in R(t_i)} \mathbf{x}_j \exp(\mathbf{x}_j \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j \boldsymbol{\beta})}$$

sendo  $j$  cada indivíduo e  $i$  ( $i = 1, \dots, m$ ) o índice dos tempos observados de eventos.

O resíduo de Schoenfeld é a diferença entre os valores observados de covariáveis de um indivíduo com tempo de ocorrência do evento  $t_i$  e os valores esperados em  $t_i$  dado o grupo de risco  $R(t_i)$ . Haverá tantos vetores de resíduos quanto covariáveis ajustadas no modelo, e que estes são definidos somente nos tempos de ocorrência do evento.

# Schoenfeld

Suponha um coeficiente  $\beta_k$  ( $k$  é cada covariável) que varia com o tempo  $t$ .  $\beta_k$  pode ser dividido em duas partes:

- uma média constante –  $E[r_i(\beta_k)|R(t_i)]$ , com variância  $V(\beta_k)$
- e uma função  $U(t)$  – que varia no tempo

O resíduo padronizado de Schoenfeld em  $t_i$  pode ser obtido por:

$$r_i^*(\beta_k) = \frac{r_i(\beta_k)}{V(\beta_k)}.$$

O valor esperado deste resíduo padronizado  $r_i^*(\beta_k)$  para cada grupo em risco  $R(t_i)$  é aproximadamente igual à parte de  $\beta_k$  que varia no tempo – a função  $U(t)$  – GRÁFICO.

## Schoenfeld no $R$

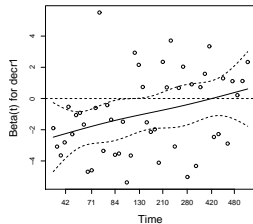
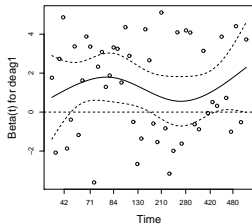
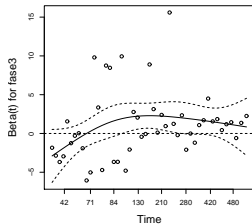
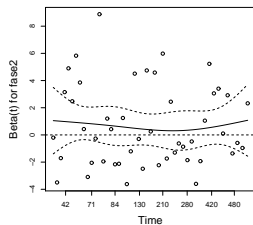
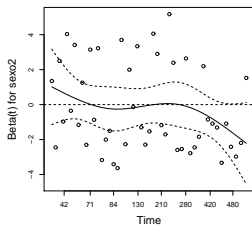
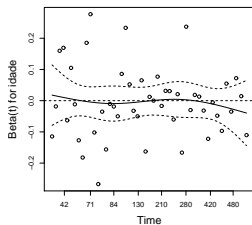
```
> residuo <- cox.zph(modelo)
> plot(residuo[1])
> abline(h=0,lty=2)
```

Atenção para a escala do tempo:

- Kaplan-Meier – nos tempos de falha
- Calendário – bom quando ajuste usando processo de contagem, pode ficar pouco visível se concentra grande quantidade de eventos em um mesmo momento
- Rank – ordem dos eventos

A linha curva é um *lowess*.

## Gráficos de Schoenfeld



# Não proporcionalidade – soluções

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



# Resíduos Martingale

É a diferença entre o número observado de eventos para um indivíduo e o esperado dado o modelo ajustado, o tempo de seguimento e o percurso observado de quaisquer covariáveis tempo-dependentes.

# Resíduos Martingale

Semelhante aos resíduos dos modelos de regressão linear:

- o valor esperado = 0
- o somatório dos resíduos observados = 0
- os resíduos  $M_i$  são não correlacionados, mas as estimativas  $\hat{M}_i$  são negativamente correlacionadas, ainda que fracamente

# Resíduos Martingale

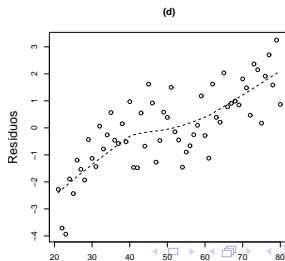
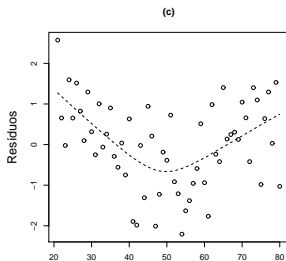
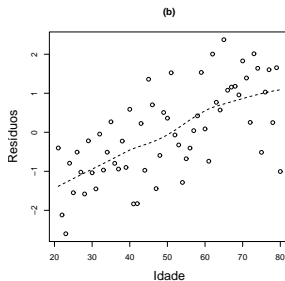
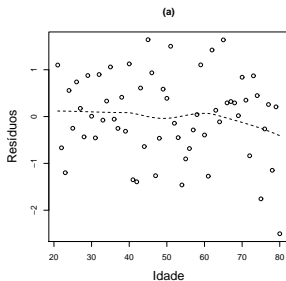
E diferentes dos resíduos da regressão linear:

- 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

- $M_i$  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.

# Gráficos Martingale



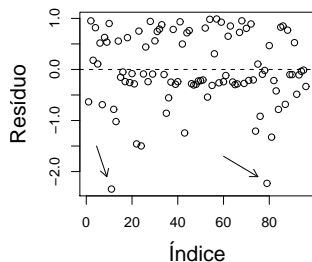
# Martingale no $R$

A função para calcular o resíduo de Martingale é:

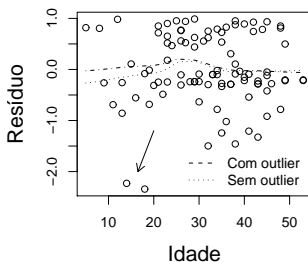
```
> res <- resid(modelo, type="martingale")
```

em que *modelo* é o objeto que recebeu o modelo de Cox.

(a)



(b)



(a) resíduo por indivíduo (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

## Smoothing spline

- O objetivo é estimar  $\beta$  de tal forma que se obtenha, simultaneamente, o menor número possível de nós e a menor soma dos resíduos quadráticos para a covariável em questão
- Parâmetro  $\theta$  indica afastamento da reta:
  - $\theta \rightarrow 0$ , a solução converge para uma reta
  - $\theta \rightarrow 1$  a curva passa por todos os pontos
- Número de pontos pelos quais a curva passará são os graus de liberdade acrescentados ao modelo
- Escolha pelo critério de informação de Akaike (*Akaike Information Criteria* – AIC)



## No R

```
> coxph(formula = Surv(os, status) ~ pspline(idade,df=0)+
  sexo+fase+decr+deag, data=tmocens, 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),l	-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
```

## No R

```
> 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(> Chi )
1	90	358.20			
2	74	342.74	16	15.46	0.49

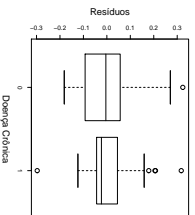
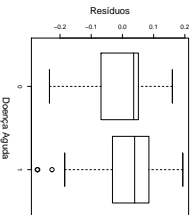
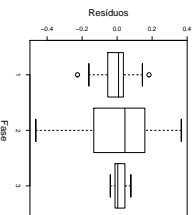
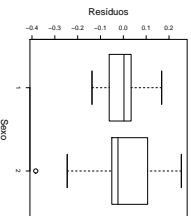
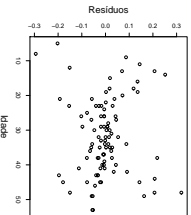
Observar:

- os graus de liberdade são fracionários
- o componente linear da idade  $pspline(idade, df=0)$ ,  $1$  é não significativo
- o componente suavizado tem  $p < 0,05$
- $summary(mod5)$  – indica 18 nós para idade

## Resíduos score

- Verifica a influência de cada observação no ajuste do modelo
- Permite a estimação robusta da variância dos coeficientes de regressão (útil para dados em cluster)
- A influência de cada observação deve ser proporcional à  $(x_i - \bar{x}) \times$  resíduo
- O gráfico do resíduo score para cada covariável  $\Delta\beta_k$  versus  $x$  mostra pontos de alavanca
- Vantagem – definidos para todos os tempos, mesmo onde não ocorre evento, melhorando a análise quando há muita censura

## Resíduos escore



## Resíduos score no R

```
> res.esco <- resid(modelo,type="dfbetas")  
> 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 as variáveis incluídas no modelo, na ordem em que foram colocadas. Para lembrar quais são, veja `modelo$call`

# 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
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 <i>spline</i> ( <code>pspline()</code> ) da covariável diretamente no modelo de Cox
Avaliar efeito de valores aberrantes	gráficos de resíduos score