

# Modelos de Riscos Proporcionais

Suely Ruiz Giolo<sup>1</sup>

Departamento de Estatística  
Universidade Federal do Paraná

## Resumo

Modelos de análise de sobrevivência têm por objetivo estudar dados de experimentos em que, geralmente, a variável resposta é o tempo até a ocorrência de um evento de interesse. A principal característica dos dados de sobrevivência é a presença de censuras, que é a observação parcial da resposta. Enquanto comum na prática, a censura impede o uso de procedimentos estatísticos convencionais tais como a análise de regressão e a análise de planejamento de experimentos, uma vez que tais procedimentos não permitem que as informações contidas nos dados censurados sejam incorporadas. Em consequência, métodos nessa área têm sido desenvolvidos e, o interesse pelo assunto, crescido significativamente após a publicação do artigo de Cox (1972). Neste material, os modelos de riscos proporcionais, que modelam a função de risco na presença de covariáveis são apresentados. Dois exemplos são analisados e seus resultados interpretados e discutidos. As funções utilizadas no *software* R (<http://www.r-project.org>) para obtenção dos resultados são também fornecidas nos anexos A e B.

**Palavras-chave:** Sobrevivência, modelo de Cox, riscos proporcionais, censura.

---

<sup>1</sup>Endereço da autora: Universidade Federal do Paraná, Departamento de Estatística, Centro Politécnico, CP 19081, 81531-990 Curitiba, Paraná, Brasil. Tel.: (41) 361-3264 / 361-3141; E-mail: [suely@est.ufpr.br](mailto:suely@est.ufpr.br)

# SUMÁRIO

|   | Página |
|---|--------|
| 1 INTRODUÇÃO . . . . .  | 1      |
| 2 MODELOS DE RISCOS PROPORCIONAIS . . . . .                         | 2      |
| 2.1 Modelo de regressão de Cox . . . . .                            | 2      |
| 2.1.1 Estimação dos parâmetros no modelo de Cox . . . . .           | 3      |
| 2.1.2 Estimação de funções relacionadas com $\alpha_0(t)$ . . . . . | 4      |
| 2.1.3 Interpretação dos parâmetros no modelo de Cox . . . . .       | 5      |
| 2.1.4 Suposição de riscos proporcionais no modelo de Cox . . . . .  | 5      |
| 2.1.5 Os resíduos no modelo de Cox . . . . .                        | 6      |
| 2.1.5.1 Resíduos de Cox-Snell . . . . .                             | 6      |
| 2.1.5.2 Resíduos de Schoenfeld . . . . .                            | 7      |
| 2.1.5.3 Resíduos Martingale . . . . .                               | 7      |
| 2.2 Exemplos . . . . .  | 8      |
| 2.2.1 Exemplo 1 . . . . .   | 8      |
| 2.2.2 Diagnóstico do modelo de Cox ajustado . . . . .               | 10     |
| 2.3 Exemplo 2 . . . . .   | 12     |
| 2.3.1 Funções de sobrevivência e risco estimadas . . . . .          | 16     |
| 2.4 Modelos de riscos proporcionais paramétricos . . . . .          | 18     |
| 2.4.1 Modelo exponencial de riscos proporcionais . . . . .          | 18     |
| 2.4.2 Modelo Weibull de riscos proporcionais . . . . .              | 19     |
| BIBLIOGRAFIA . . . . .  | 20     |
| ANEXOS . . . . .  | 21     |

# 1 INTRODUÇÃO

A análise de sobrevivência reúne um conjunto de métodos úteis na análise de experimentos em que, geralmente, a variável resposta é o tempo até a ocorrência de um evento de interesse. Os indivíduos sob estudo podem ser animais, seres humanos, plantas, equipamentos etc. Por outro lado, o evento de interesse pode ser, por exemplo:

- nos experimentos biomédicos: morte, remissão de uma doença, reação a uma droga, cura de uma doença etc.
- nos experimentos industriais: quebra de um componente eletrônico, queima de uma lâmpada etc.
- em demografia: casamento, divórcio etc.
- na área de seguros: sinistro (roubo, colisão etc).

Dados de sobrevivência é, portanto, um termo usado para descrever experimentos que medem o tempo até a ocorrência de um evento. A principal característica dos dados de sobrevivência é a presença de censuras, que é a observação parcial da resposta. Isso refere-se a circunstâncias em que alguns indivíduos encontram-se livres do evento por terem, por exemplo, sido retirados mais cedo do estudo ou pelo término do experimento. Isto significa que toda informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado.

Enquanto comum na prática, a censura impede o uso de procedimentos estatísticos convencionais uma vez que tais procedimentos não permitem que as informações contidas nos dados censurados sejam incorporadas. Em consequência, métodos nessa área têm sido desenvolvidos e, o interesse pelo assunto, crescido significativamente após a publicação do artigo de Cox (1972).

Neste material são, portanto, apresentados os modelos de riscos proporcionais, que modelam a função de risco na presença de covariáveis. Embora os exemplos utilizados sejam da área biomédica, os modelos são válidos para quaisquer outras áreas de pesquisa.

## 2 MODELOS DE RISCOS PROPORCIONAIS

### 2.1 Modelo de regressão de Cox

O modelo de regressão de Cox, proposto por Cox em 1972, é sem dúvida um dos mais populares na análise de dados de sobrevivência. Este modelo permite que a análise dos tempos de vida até a ocorrência de um evento seja realizada considerando-se as covariáveis de interesse. Cox, assim como outros autores, propõe modelar dados de sobrevivência, na presença de covariáveis, por meio da função de risco.

Assume-se, nesse modelo, que os tempos  $t_i$ ,  $i = 1, \dots, n$ , são independentes e que o risco do indivíduo  $i$  é dado por:

$$\alpha_i(t \mid \mathbf{x}_i) = \alpha_0(t) \exp\{\boldsymbol{\beta}'\mathbf{x}_i\} \quad (1)$$

em que  $\alpha_0$  é conhecida como a função de risco base, ou seja, o risco de um indivíduo com covariáveis  $\mathbf{x} = \mathbf{0}$ ,  $\boldsymbol{\beta}$  é o vetor de dimensão  $p$  de coeficientes de regressão desconhecidos e  $\mathbf{x}_i$  é o vetor de dimensão  $p$  de covariáveis observadas para o indivíduo  $i$ .

O modelo de Cox, definido em (1), é conhecido como sendo semi-paramétrico por assumir que as covariáveis atuam multiplicativamente no risco pela relação  $g(\mathbf{x}, \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}'\mathbf{x}_i\}$  e por considerar  $\alpha_0(t)$  arbitrária, ou seja, por não ser assumida nenhuma forma paramétrica para  $\alpha_0(t)$ . A denominação desse modelo como sendo de riscos proporcionais se deve ao fato de a razão entre as funções de risco de dois indivíduos,

$$\frac{\alpha(t \mid \mathbf{x}_i)}{\alpha(t \mid \mathbf{x}_\ell)} = \frac{\alpha_0(t) \exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}{\alpha_0(t) \exp\{\boldsymbol{\beta}'\mathbf{x}_\ell\}} = \exp\left\{\boldsymbol{\beta}'(\mathbf{x}_i - \mathbf{x}_\ell)\right\},$$

( $i, \ell = 1, \dots, n$  e  $i \neq \ell$ ), não depender de  $t$ .

Observe que a constante  $\beta_0$  não aparece no componente paramétrico  $\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}$ . Isto ocorre devido à presença do componente não-paramétrico que absorve este termo constante.

A suposição básica para o uso do modelo de riscos proporcionais de Cox é, portanto, que as taxas de falha sejam proporcionais. Em outras palavras, isto significa que se um indivíduo apresenta no início do estudo um risco de falha igual a três vezes o risco de um outro indivíduo, então esta taxa será a mesma para qualquer tempo  $t$  do período de acompanhamento.

Pela presença do componente não-paramétrico, o que o torna bastante flexível, o modelo de regressão de Cox é utilizado com muita frequência em estudos médicos.

As funções de taxa de falha base acumulada bem como a correspondente função de sobrevivência são também de interesse e estas relacionam-se com a função de risco base por, respectivamente,

$$H_0(t) = \int_0^t \alpha_0(u) du$$

e

$$S(t | \mathbf{x}) = \exp\{-H_0(t) \exp(\boldsymbol{\beta}'\mathbf{x})\} = \left(S_0(t)\right)^{\exp\{\boldsymbol{\beta}'\mathbf{x}\}}$$

em que,

$$S_0(t) = \exp\{-H_0(t)\}.$$

### 2.1.1 Estimação dos parâmetros no modelo de Cox

O modelo de Cox caracteriza-se pelos coeficientes  $\boldsymbol{\beta}$  que devem ser estimados a partir das observações amostrais. A presença do componente não-paramétrico inviabiliza, contudo, o uso do método da máxima verossimilhança.

Para a estimação do vetor de parâmetros desconhecidos  $\boldsymbol{\beta}$ , Cox (1975) propôs, então, a função de verossimilhança parcial,  $L(\boldsymbol{\beta})$ . Para a construção desta função de verossimilhança, considere uma amostra de  $n$  indivíduos em que  $k \leq n$  falhas distintas ocorrem nos tempos  $t_1, \dots, t_k$  e que a probabilidade condicional da  $i$ -ésima observação vir a falhar no tempo  $t_i$ , conhecendo as observações sob risco em  $t_i$ , é:

$$\frac{\alpha_i(t_i | \mathbf{x}_i)}{\sum_{j \in R(t_i)} \alpha_j(t_i | \mathbf{x}_j)} = \frac{\alpha_0(t_i) \exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}{\sum_{j \in R(t_i)} \alpha_0(t_i) \exp\{\boldsymbol{\beta}'\mathbf{x}_j\}} = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}{\sum_{j \in R(t_i)} \exp\{\boldsymbol{\beta}'\mathbf{x}_j\}}$$

em que  $R(t_i) = R_i$  é o conjunto de todos os indivíduos ainda em risco no tempo  $t_i$ .

Assim, a função de verossimilhança a ser utilizada é expressa por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left( \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}{\sum_{j \in R_i} \exp\{\boldsymbol{\beta}'\mathbf{x}_j\}} \right)^{\delta_i}.$$

Os indivíduos censurados entram na função de verossimilhança parcial  $L(\boldsymbol{\beta})$  através do conjunto de risco  $R_{(t_i)}$  e contribuem, para esta função, somente enquanto permanecem

em risco. Os parâmetros  $\beta$  são estimados maximizando-se o logaritmo da função  $L(\beta)$ , isto é, resolvendo-se o sistema de equações escores

$$0 = \frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^n \delta_i \left[ x_{ik} - \frac{\sum_{j \in R_i} x_{jk} \exp\{\beta' \mathbf{x}_j\}}{\sum_{j \in R_i} \exp\{\beta' \mathbf{x}_j\}} \right] \quad k = 1, \dots, p$$

em que  $\ell(\beta) = \log L(\beta)$ .

As propriedades assintóticas dos estimadores de máxima verossimilhança parcial, necessárias para a construção de intervalos de confiança e testes de hipóteses, foram estudadas (Andresen e Gill, 1982) e estes são, consistentes e assintoticamente normais sob certas condições de regularidade. Pode-se, desse modo, usar as estatísticas de Wald e da razão de verossimilhanças para fazer inferências no modelo de Cox.

### 2.1.2 Estimação de funções relacionadas com $\alpha_0(t)$

Os coeficientes de regressão  $\beta$  são quantidades de grande interesse no modelo de Cox. No entanto, funções relacionadas com  $\alpha_0(t)$ , tais como as funções  $H_0(t)$  e  $S_0(t)$  apresentadas na seção 2.1, são também importantes.

Se  $\alpha_0(t)$  fosse especificado parametricamente, seria possível estimá-lo por meio da função de verossimilhança. Como visto, entretanto, na construção da função de verossimilhança parcial,  $\alpha_0(t)$  é eliminado completamente desta função e, desse modo, os estimadores propostos para essas quantidades serão de natureza não-paramétrica.

Um estimador simples proposto para a função de risco base acumulada  $H_0(t)$ , referenciado como estimador de Nelson-Aalen-Breslow, ou simplesmente estimador de Breslow, é uma função escada, com saltos nos distintos tempos de falha, e expresso por:

$$\widehat{H}_0(t_i) = \sum_{i: t_i \leq t} \frac{d(t_i)}{\sum_{j \in R_j} \exp(\hat{\beta}' \mathbf{x}_j)}$$

em que  $d(t_i)$  é o número de falhas em  $t_i$ .

Note que se não existirem covariáveis, então

$$\widehat{H}_0(t_i) = \sum_{i: t_i \leq t} \frac{d(t_i)}{n(t_i)}$$

que é o estimador de Nelson-Aalen.

Uma vez estimada a função de risco base acumulada, pode-se obter a função de sobrevivência base estimada, isto é,

$$\hat{S}_0(t) = \exp \left\{ -\hat{H}_0(t) \right\}$$

e a partir desta, obter a função de sobrevivência estimada que, para um indivíduo com vetor de covariáveis  $\mathbf{x}_i$ , é expressa por:

$$\hat{S}(t | \mathbf{x}_i) = \left( \hat{S}_0(t) \right)^{\exp \{ \hat{\beta}' \mathbf{x}_i \}}$$

e será, também, uma função escada decrescendo com o tempo.

### 2.1.3 Interpretação dos parâmetros no modelo de Cox

Os coeficientes  $\beta$  no modelo de regressão de Cox, medem os efeitos das covariáveis sobre a taxa de falha sendo que, uma covariável pode acelerar, ou desacelerar, a função de risco.

A propriedade de riscos proporcionais é usada para interpretar os coeficientes estimados. Observe que a taxa de falha para dois indivíduos ( $i$  e  $l$ ) que apresentam os mesmos valores para as covariáveis, exceto para a  $p$ -ésima delas, é dada por:

$$\frac{\alpha(t | \mathbf{x}_i)}{\alpha(t | \mathbf{x}_l)} = \frac{\exp \{ \beta_p x_{ip} \}}{\exp \{ \beta_p x_{lp} \}} = \exp \{ \beta_p (x_{ip} - x_{lp}) \}.$$

Esta razão é, na realidade, uma razão de riscos e, assim, se, por exemplo,  $x_p$  é a covariável dicotômica sexo em que  $x_{ip} = 1$  (masculino) e  $x_{lp} = 0$  (feminino), tem-se que o risco de falha dos indivíduos do sexo masculino é  $\exp(\beta_p)$  vezes o risco de falha dos indivíduos do sexo feminino, mantendo-se fixas as demais covariáveis.

### 2.1.4 Suposição de riscos proporcionais no modelo de Cox

Como mencionado anteriormente, o modelo de Cox é adequado para situações em que a suposição de riscos proporcionais é válida, isto é, para situações em que as funções de risco não se cruzam.

Para verificar tal suposição pode-se fazer uso, dentre outras maneiras, de uma técnica gráfica que consiste basicamente em:

- i) dividir os dados em  $j$  estratos de acordo com as  $j$  categorias de alguma covariável. Por exemplo, em dois estratos ( $j = 2$ ) se a covariável considerada é sexo e, então,
- ii) estimar  $\widehat{H}_{0j}(t_i)$  para cada estrato  $j$  obtendo-se as curvas de  $\log \widehat{H}_{0j}(t)$  versus  $t$ , ou  $\log(t)$ .

Se a suposição for válida, as curvas de  $\log \widehat{H}_{0j}(t)$  versus  $t$ , ou  $\log(t)$ , devem apresentar diferenças constantes no tempo, ou seja, devem ser aproximadamente paralelas. Estas curvas podem ser obtidas para cada covariável em estudo podendo-se, desse modo, encontrar indícios de qual covariável estaria causando a possível violação da suposição de riscos proporcionais. Se a covariável for de natureza contínua é sugerido agrupá-la em um pequeno número de categorias.

### 2.1.5 Os resíduos no modelo de Cox

Em regressão linear é usual verificar a adequação do modelo ajustado por meio da inspeção de gráficos dos resíduos. Em análise de sobrevivência, devido a presença de censuras bem como pelos resíduos não seguirem a distribuição Normal e poderem ser altamente assimétricos, diversos tipos de resíduos têm sido propostos na literatura.

#### 2.1.5.1 Resíduos de Cox-Snell

Para os  $n$  indivíduos ( $i = 1, \dots, n$ ) em estudo, os resíduos de Cox-Snell (1968), para o modelo de Cox, são definidos por:

$$e_i = \hat{H}_0(t_i) \exp \left\{ \sum_{k=1}^p x_{ip} \hat{\beta}_k \right\}, \quad i = 1, \dots, n$$

e podem ser usados para verificar a qualidade de ajuste do modelo de Cox. Se o modelo estiver bem ajustado, os  $e_i$ 's podem ser olhados como uma amostra censurada de uma distribuição exponencial padrão e, então, o gráfico de, por exemplo,  $\hat{H}(e_i)$  versus  $e_i$  deveria ser aproximadamente uma reta.

Os resíduos de Cox-Snell são úteis para examinar o ajuste global do modelo de Cox. Uma desvantagem do seu uso, contudo, é que eles não indicam, quando o gráfico de  $\hat{H}(e_i)$  versus  $e_i$  apresentar-se não linear, o tipo de falha detectado pelo modelo. Outros tipos de resíduos são preferidos nessas situações.



Os resíduos de Cox-Snell deveriam, ainda, ser usados com cuidado pois a distribuição exponencial dos mesmos mantém-se somente quando os verdadeiros valores dos parâmetros são usados. Quando as estimativas dessas quantidades são usadas para o cálculo dos resíduos, como feito aqui, falhas em detectar a distribuição exponencial podem ser devidas, parcialmente, à incerteza no processo de estimação de  $\beta$  e  $H_0$ . Essa incerteza é maior na cauda direita da distribuição e para amostras pequenas. Crowley e Storer (1983) mostraram, usando simulação, que estes resíduos não são tão apropriados para avaliar a adequação do modelo quanto se pensava.

### 2.1.5.2 Resíduos de Schoenfeld

Considerando os  $k$  ( $k \leq n$ ) tempos distintos de falha  $t_1, \dots, t_k$ , um vetor de resíduos de Schoenfeld é obtido em cada tempo observado de falha. Assim, se o indivíduo  $i$  é observado falhar, o correspondente resíduo é definido por:

$$r_i = x_i - \frac{\sum_{j \in R(t_i)} x_j e^{\hat{\beta}x_j}}{\sum_{j \in R(t_i)} e^{\hat{\beta}x_j}}. \quad (2)$$

Note que este é um vetor com um componente para cada covariável. Condicional a uma única falha no conjunto de risco  $R(t_i)$ , o valor esperado da covariável, desta falha, é expresso pelo termo  $\frac{\sum_{j \in R(t_i)} x_j e^{\hat{\beta}x_j}}{\sum_{j \in R(t_i)} e^{\hat{\beta}x_j}}$  apresentado em (2) e, desse modo, a interpretação de  $r_i$  como um resíduo é apropriada. Também,  $\sum_i r_i = 0$ , como usual para resíduos e, se o modelo é apropriado, não deveria existir tendências no tempo. Os resíduos de Schoenfeld são, em geral, usados para verificar a suposição de riscos proporcionais.

Uma forma escalonada dos resíduos de Schoenfeld é definida por:

$$r_i^* = d \times i(\beta)^{-1} \times r_i$$

em que  $i(\beta)$  é a informação e  $d$  é o número de falhas observadas.

### 2.1.5.3 Resíduos Martingale

O resíduo Martingale resulta de uma modificação do resíduo de Cox-Snell apresentado anteriormente. Assim, quando os dados apresentam censuras à direita e todas as

covariáveis são fixadas no início do estudo, ou seja, não forem dependentes do tempo, o resíduo martingale é definido por:

$$M_i = \delta_i - \hat{H}_0(t_i) \exp \left\{ \sum_{k=1}^p x_{ip} \hat{\beta}_k \right\} = \delta_i - e_i, \quad i = 1, \dots, n$$

Esses resíduos podem ser usados para verificar a adequação do modelo mas, na prática, devido à sua distribuição assimétrica, o que torna sua interpretação difícil, eles são usados para verificação da forma funcional das covariáveis, isto é, se estas deveriam ser usadas no modelo como  $\log(x_i)$ ,  $x_i^2$ , e assim por diante, em vez de  $x_i$ .

## 2.2 Exemplos

### 2.2.1 Exemplo 1

Considere os tempos de sobrevida, em semanas, de 17 pacientes com leucemia aguda que apresentaram Ag positivo, bem como outro grupo de 16 pacientes, também com leucemia aguda, mas com Ag negativo (Louzada-Neto et al., 2002) apresentados na Tabela 1.

Tabela 1. Sobrevida de dois grupos de pacientes com leucemia aguda.

| Ag+    |        |                         | Ag-    |        |                         |
|--------|--------|-------------------------|--------|--------|-------------------------|
| Tempos | WBC    | $\log_{10}(\text{WBC})$ | Tempos | WBC    | $\log_{10}(\text{WBC})$ |
| 65     | 2300   | 3,36                    | 56     | 4400   | 3,64                    |
| 156    | 750    | 2,88                    | 65     | 3000   | 3,48                    |
| 100    | 4300   | 3,63                    | 17     | 4000   | 3,60                    |
| 134    | 2600   | 3,41                    | 7      | 1500   | 3,18                    |
| 16     | 6000   | 3,78                    | 16     | 9000   | 3,95                    |
| 108    | 10000  | 4,02                    | 22     | 5300   | 3,72                    |
| 121    | 10000  | 4,00                    | 3      | 10000  | 4,00                    |
| 4      | 17000  | 4,23                    | 4      | 19000  | 4,28                    |
| 39     | 5400   | 3,73                    | 2      | 27000  | 4,43                    |
| 143    | 7000   | 3,85                    | 3      | 28000  | 4,45                    |
| 56     | 9400   | 3,97                    | 8      | 31000  | 4,49                    |
| 26     | 32000  | 4,51                    | 4      | 26000  | 4,41                    |
| 22     | 35000  | 4,54                    | 3      | 21000  | 4,32                    |
| 1      | 100000 | 5,00                    | 30     | 79000  | 4,90                    |
| 1      | 100000 | 5,00                    | 4      | 100000 | 5,00                    |
| 5      | 52000  | 4,72                    | 43     | 100000 | 5,00                    |
| 65     | 100000 | 5,00                    |        |        |                         |

Para todos os pacientes, a covariável contagem de glóbulos brancos (WBC) foi registrada na data do diagnóstico estando, seus respectivos logaritmos na base 10, apresentados também na referida tabela. As covariáveis presentes nesse conjunto de dados são, portanto:  $X_1$  = logaritmo da contagem de glóbulos brancos e  $X_2$  = grupos (Ag+ ou Ag-).

Utilizando-se o modelo de Cox para a análise desses dados, foram obtidos os resultados apresentados na Tabela 2.

Tabela 2. Estimativas dos parâmetros para os dados de leucemia aguda.

| Modelo | Covariáveis no modelo | estimativas             | log verossimilhança parcial |
|--------|-----------------------|-------------------------|-----------------------------|
| 1      | nenhuma               | -                       | $l_1 = -85,05447$           |
| 2      | $X_1$ : ag            | $\hat{\beta}_1 = 1,180$ | $l_2 = -80,92388$           |
| 3      | $X_1$ : ag            | $\hat{\beta}_1 = 1,069$ |                             |
|        | $X_2$ : log(wbc)      | $\hat{\beta}_2 = 0,844$ | $l_3 = -77,25004$           |

Para testar a significância das covariáveis  $X_1$  e  $X_2$ , isto é, testar as hipóteses nulas  $H_0$ :  $\beta_1 = 0$  e  $H_0$ :  $\beta_2 = 0$  foi utilizado o teste da razão de verossimilhanças obtendo-se, respectivamente,  $RV = 8,26$  ( $p = 0,004$ , g.l = 1) e  $RV = 7,35$  ( $p = 0,0067$ , g.l. = 1). Ambas as covariáveis são, portanto, significativas permanecendo, desse modo, no modelo de Cox que, nesse exemplo, é expresso por:

$$\hat{\alpha}_i(t | \mathbf{x}_i) = \hat{\alpha}_0(t) \exp\{1,069 \text{ ag} + 0,844 \log(\text{wbc})\}$$

em que  $\text{ag} = 0$  se o paciente  $i$  pertence ao grupo Ag+ ou  $\text{ag} = 1$  se este pertence ao grupo Ag- e  $\log(\text{wbc})$  é o correspondente logaritmo, na base 10, da contagem de glóbulos brancos deste paciente.

As estimativas das funções base  $S_0(t)$ ,  $\alpha_0(t)$  e  $H_0(t)$ , para os dados desse estudo, encontram-se apresentadas na Tabela 3 sendo, portanto, a função de sobrevivência  $S(t | \mathbf{x})$  estimada, expressa por:

$$\hat{S}(t | \mathbf{x}) = [\hat{S}_0(t)]^{\exp\{1,069 \text{ ag} + 0,844 \log_{10}(\text{wbc})\}}$$

em que  $\text{ag} = 0$  se Ag+ ou  $\text{ag} = 1$  se Ag- e  $\log_{10}(\text{wbc})$  é o logaritmo, na base 10, da contagem de glóbulos brancos e, sua representação gráfica, para dois pacientes com Ag+,

Tabela 3. Estimativas de  $S_0(t)$ ,  $\alpha_0(t)$  e  $H_0(t)$ .

| Tempo | $\hat{S}_0(t)$ | $\hat{\alpha}_0(t)$ | $\hat{H}_0(t)$ |
|-------|----------------|---------------------|----------------|
| 1     | 0,99914        | 0,00086             | 0,00086        |
| 2     | 0,99869        | 0,00045             | 0,00131        |
| 3     | 0,99719        | 0,00150             | 0,00281        |
| 4     | 0,99470        | 0,00250             | 0,00531        |
| 5     | 0,99395        | 0,00075             | 0,00607        |
| 7     | 0,99317        | 0,00079             | 0,00685        |
| 8     | 0,99236        | 0,00082             | 0,00767        |
| 16    | 0,99050        | 0,00188             | 0,00955        |
| 17    | 0,98950        | 0,00101             | 0,01056        |
| 22    | 0,98731        | 0,00222             | 0,01277        |
| 26    | 0,98610        | 0,00123             | 0,01400        |
| 30    | 0,98482        | 0,00130             | 0,01530        |
| 39    | 0,98315        | 0,00170             | 0,01699        |
| 43    | 0,98141        | 0,00177             | 0,01876        |
| 56    | 0,97570        | 0,00584             | 0,02460        |
| 65    | 0,96262        | 0,01350             | 0,03810        |
| 100   | 0,95554        | 0,00738             | 0,04548        |
| 108   | 0,94719        | 0,00878             | 0,05426        |
| 121   | 0,93601        | 0,01187             | 0,06613        |
| 134   | 0,91913        | 0,01820             | 0,08433        |
| 143   | 0,89472        | 0,02692             | 0,11124        |
| 156   | 0,81940        | 0,08794             | 0,19918        |

e outros dois pacientes com Ag-, que apresentaram  $\log(\text{wbc}) = 3,0$  e  $4,0$  no diagnóstico, são mostradas na Figura 1.

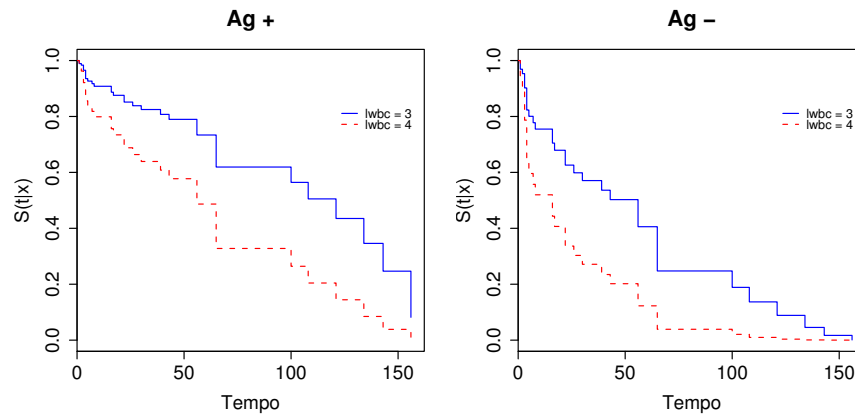


Figura 1: Sobrevivências estimadas pelo modelo de Cox.

### 2.2.2 Diagnóstico do modelo de Cox ajustado

Para verificar se o modelo ajustado é adequado aos dados sob estudo será usado os resíduos de Schoenfeld definidos anteriormente.

Utilizando-se, então, os resíduos escalonados de Schoenfeld obtiveram-se os resultados apresentados na Tabela 4 e Figura 2.

Tabela 4. Testes para tendências.

|        | rho     | chisq  | p     |
|--------|---------|--------|-------|
| ag     | 0.1315  | 0.6892 | 0.406 |
| lwbc   | -0.0455 | 0.0522 | 0.819 |
| GLOBAL | NA      | 0.7098 | 0.701 |

A partir da Tabela 4, em que a tendência dos resíduos sobre o tempo é testada, observa-se, pelos valores  $p$ , não haver evidências de tendências acentuadas para as co-variáveis em estudo. Nos gráficos apresentados na Figura 2, é possível visualizar tais tendências não acentuadas ao longo do tempo. A suposição de riscos proporcionais parece ser, portanto, razoável para a análise desses dados.

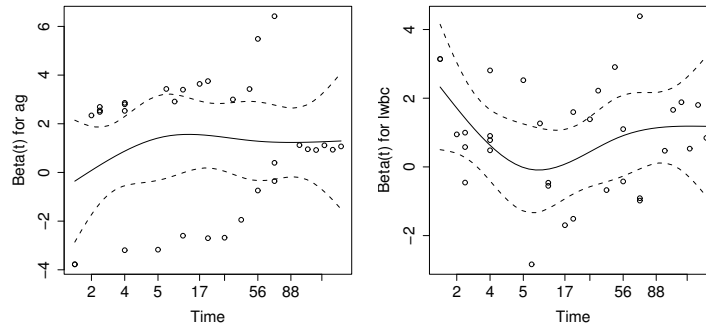


Figura 2: Resíduos escalonados de Schoenfeld.

De um modo geral, pode-se concluir para os dados aqui analisados, ressaltando-se o fato de ter-se uma amostra relativamente pequena, que o modelo de riscos proporcionais de Cox apresentou um ajuste razoável aos dados utilizados.

Assim, por exemplo, para dois pacientes  $i$  e  $l$  com  $x_2 = \log_{10}(\text{wbc}) = 3,0$ , mas um deles com  $x_1 = \text{ag} = 1$  e o outro com  $x_1 = \text{ag} = 0$ , tem-se a razão de riscos:

$$\frac{\alpha_i(t \mid \mathbf{x}_i)}{\alpha_l(t \mid \mathbf{x}_l)} = \exp \{1,069(1 - 0)\} = 2,91.$$

Isto significa, portanto, que o risco de falha dos pacientes do grupo Ag- ( $x_1 = 1$ ) é 2,91 vezes maior do que o risco de falha dos pacientes do grupo Ag+ ( $x_1 = 0$ ). Observe que se

fosse considerado, para ambos os pacientes,  $\log_{10}(\text{wbc}) = 4,0$  em vez de  $\log_{10}(\text{wbc}) = 3,0$ , obteria-se também, para a razão de riscos,  $\exp\{1,069\} = 2,91$ .

Ainda, para dois pacientes do grupo Ag+, mas um deles com  $x_2 = 4,0$  e o outro com  $x_2 = 3,0$ , tem-se a razão de riscos:

$$\frac{\alpha_i(t \mid \mathbf{x}_i)}{\alpha_l(t \mid \mathbf{x}_l)} = \exp\{0,844(4,0 - 3,0)\} = 2,32$$

o que implica que o risco de falha dos pacientes com  $x_2 = 4,0$  é 2,32 vezes maior do que o risco de falha dos pacientes com  $x_2 = 3,0$ . Observe que, se ambos os pacientes fossem do grupo Ag-, a razão de riscos resultante desses dois pacientes seria também  $\exp\{0,844(4,0 - 3,0)\} = 2,32$ .

As respectivas funções de risco estimadas  $\hat{\alpha}(t \mid \mathbf{x})$  para os pacientes citados encontram-se apresentadas na Figura 3.

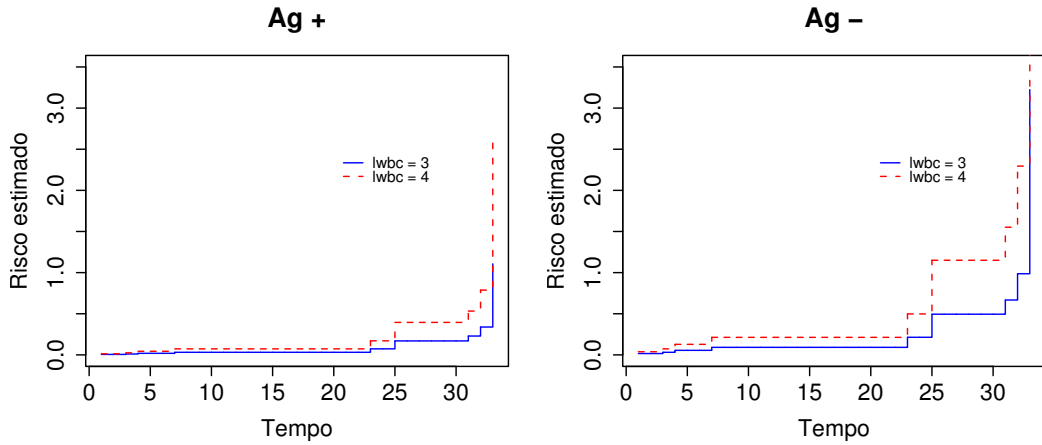


Figura 3: Funções de risco estimadas pelo modelo de Cox para dois pacientes do grupo Ag+, e dois outros do grupo Ag-, com  $\log(\text{wbc}) = 3,0$  e  $4,0$ .

## 2.3 Exemplo 2

Neste exemplo os dados considerados referem-se a um estudo, descrito em Klein e Moeschberger (1997), realizado com 90 pacientes do sexo masculino diagnosticados no período de 1970-1978 com câncer de laringe e que foram acompanhados até 01/01/1983. Para cada paciente foram registrados, no diagnóstico, a idade (em anos) e o estágio da

doença (I = tumor primário, II = envolvimento de nódulos, III = metástases e IV = combinações dos 3 estágios anteriores) bem como seus respectivos tempos de falha ou censura (em meses). Os estágios encontram-se ordenados pelo grau de seriedade da doença (menos sério para mais sério).

Utilizando-se o modelo de Cox para a análise desses dados, foram ajustados diversos modelos seqüências cujos resultados encontram-se apresentados na Tabela 5.

Tabela 5. Estimativas obtidas para os modelos ajustados aos dados de câncer de laringe.

| Modelo | Covariáveis no modelo | estimativas               | log da verossimilhança parcial |
|--------|-----------------------|---------------------------|--------------------------------|
| 1      | nenhuma               | -                         | $l_1 = -196,8635$              |
| 2      | $X_1$ : estágio (II)  | $\hat{\beta}_1 = 0,0647$  | $l_2 = -188,6208$              |
|        | (III)                 | $\hat{\beta}_2 = 0,6146$  |                                |
|        | (IV)                  | $\hat{\beta}_3 = 1,7353$  |                                |
| 3      | $X_1$ : estágio (II)  | $\hat{\beta}_1 = 0,1399$  | $l_3 = -187,7074$              |
|        | (III)                 | $\hat{\beta}_2 = 0,6422$  |                                |
|        | (IV)                  | $\hat{\beta}_3 = 1,7062$  |                                |
|        | $X_2$ : idade         | $\hat{\beta}_4 = 0,0190$  |                                |
| 4      | $X_1$ : estágio (II)  | $\hat{\beta}_1 = -8,0837$ | $l_4 = -184,5311$              |
|        | (III)                 | $\hat{\beta}_2 = -0,1640$ |                                |
|        | (IV)                  | $\hat{\beta}_3 = 0,8253$  |                                |
|        | $X_2$ : idade         | $\hat{\beta}_4 = -0,0029$ |                                |
|        | $X_1 * X_2$ (II* id)  | $\hat{\beta}_5 = 0,1223$  |                                |
|        | (III* id)             | $\hat{\beta}_6 = 0,1203$  |                                |
|        | (IV* id)              | $\hat{\beta}_7 = 0,0142$  |                                |

A partir da Tabela 5 tem-se, para o teste da razão de verossimilhança parcial associado à interação entre estágio e idade, o resultado  $RV = 2(187,7074 - 184,5311) = 6,3526$  ( $p = 0,095$ , g.l. = 3). Por outro lado, os resultados dos testes locais dessa interação, apresentados na Tabela 6, mostram evidências de que pelo menos um dos  $\beta$ 's associados à referida interação difere significativamente de zero, no caso  $\beta_5$  com p-valor = 0,02.

Em consequência dos resultados encontrados decidiu-se pela realização dos diagnósticos dos resíduos de Schoenfeld dos modelos de Cox com, e sem, a presença da interação para, então, proceder a escolha de um desses dois modelos.

Tabela 6. Testes locais dos parâmetros associados à interação.

| parâmetro                     | estimativa | erro padrão | z      | p-valor      |
|-------------------------------|------------|-------------|--------|--------------|
| $\beta_5 = \text{idade:est2}$ | 0,12236    | 0,0525      | 2,3295 | <b>0,020</b> |
| $\beta_6 = \text{idade:est3}$ | 0,01203    | 0,0375      | 0,3206 | 0,750        |
| $\beta_7 = \text{idade:est4}$ | 0,01422    | 0,0359      | 0,3958 | 0,690        |

Desse modo, e utilizando-se dos resíduos escalonados de Schoenfeld do modelo de Cox com a interação, foram obtidos os resultados apresentados na Tabela 7 e Figura 4.

Tabela 7. Testes para tendências - modelo com interação.

|                     | rho     | chisq   | p     |
|---------------------|---------|---------|-------|
| estágio II          | 0,1082  | 0,65419 | 0,419 |
| estágio III         | 0,0277  | 0,05724 | 0,811 |
| estágio IV          | 0,0110  | 0,00715 | 0,933 |
| idade               | 0,0891  | 0,64045 | 0,424 |
| estágio II * idade  | -0,1058 | 0,63314 | 0,426 |
| estágio III * idade | -0,0596 | 0,26696 | 0,605 |
| estágio IV * idade  | -0,0273 | 0,04446 | 0,833 |
| GLOBAL              | NA      | 6,13594 | 0,524 |

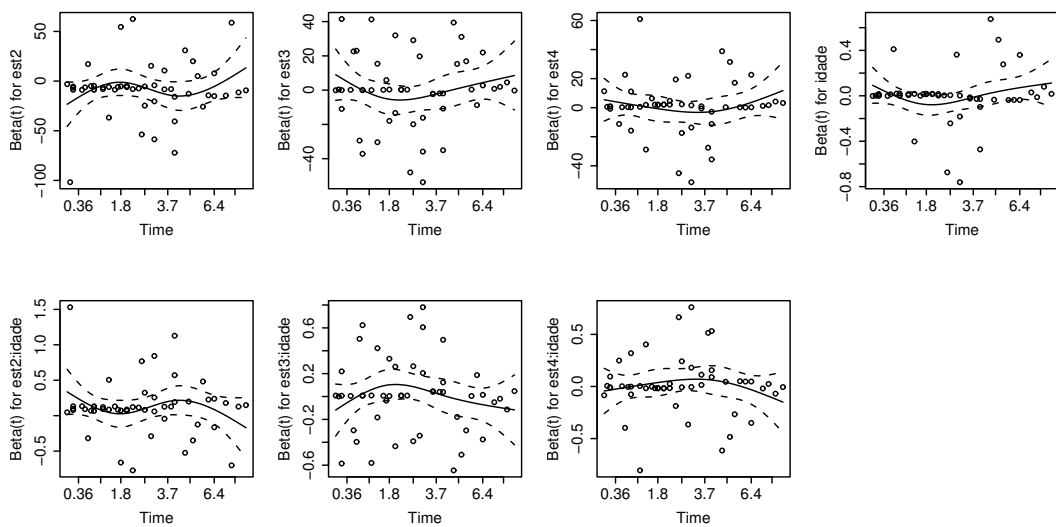


Figura 4: Resíduos escalonados de Schoenfeld do modelo de Cox com interação.



Dos resultados apresentados pode-se, portanto, observar que a suposição de riscos proporcionais não é rejeitada, uma vez que tendências ao longo do tempo não são evidentes. Para o modelo de Cox sem a interação, foram obtidos os resultados apresentados na Tabela 8 e Figura 5.

Tabela 8. Testes para tendências - modelo sem interação.

|             | rho     | chisq  | p      |
|-------------|---------|--------|--------|
| estágio II  | -0,0158 | 0,0132 | 0,9084 |
| estágio III | -0,2599 | 3,2313 | 0,0722 |
| estágio IV  | -0,1105 | 0,5424 | 0,4615 |
| idade       | 0,0891  | 0,8646 | 0,3525 |
| GLOBAL      | NA      | 4,6755 | 0,3222 |

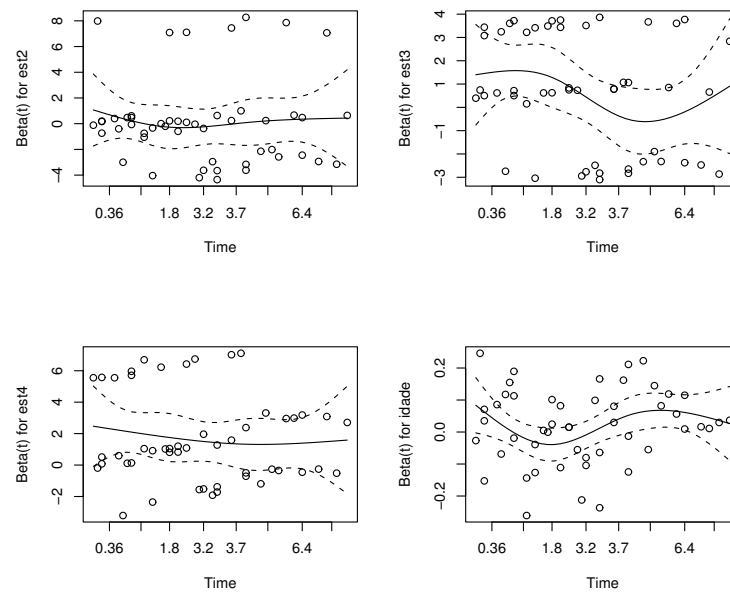


Figura 5: Resíduos escalonados de Schoenfeld do modelo de Cox sem interação.

Comparando-se os resultados de ambos os diagnósticos, decidiu-se pelo modelo com a presença da interação.

### 2.3.1 Funções de sobrevivência e risco estimadas

As funções de sobrevivência e risco estimadas são, portanto, expressas, respectivamente, por:

$$\begin{aligned}
 \hat{S}(t | \mathbf{x}) &= \left[ \hat{S}_0(t) \right]^{\exp\{\hat{\beta}_4 x_2\}} && \text{se estágio I} \\
 &= \left[ \hat{S}_0(t) \right]^{\exp\{\hat{\beta}_1 + (\hat{\beta}_4 + \hat{\beta}_5) x_2\}} && \text{se estágio II} \\
 &= \left[ \hat{S}_0(t) \right]^{\exp\{\hat{\beta}_2 + (\hat{\beta}_4 + \hat{\beta}_6) x_2\}} && \text{se estágio III} \\
 &= \left[ \hat{S}_0(t) \right]^{\exp\{\hat{\beta}_3 + (\hat{\beta}_4 + \hat{\beta}_7) x_2\}} && \text{se estágio IV}
 \end{aligned}$$

e,

$$\begin{aligned}
 \hat{\alpha}(t | \mathbf{x}) &= \hat{\alpha}_0(t) \exp\{\hat{\beta}_4 x_2\} && \text{se estágio I} \\
 &= \hat{\alpha}_0(t) \exp\{\hat{\beta}_1 + (\hat{\beta}_4 + \hat{\beta}_5) x_2\} && \text{se estágio II} \\
 &= \hat{\alpha}_0(t) \exp\{\hat{\beta}_2 + (\hat{\beta}_4 + \hat{\beta}_6) x_2\} && \text{se estágio III} \\
 &= \hat{\alpha}_0(t) \exp\{\hat{\beta}_3 + (\hat{\beta}_4 + \hat{\beta}_7) x_2\} && \text{se estágio IV}
 \end{aligned}$$

em que  $x_2 = \text{idade}$ .

Na Figura 6 encontram-se representadas as curvas de sobrevivência estimadas, para pacientes com idades de 50 e 65 anos, em cada um dos 4 estágios da doença.

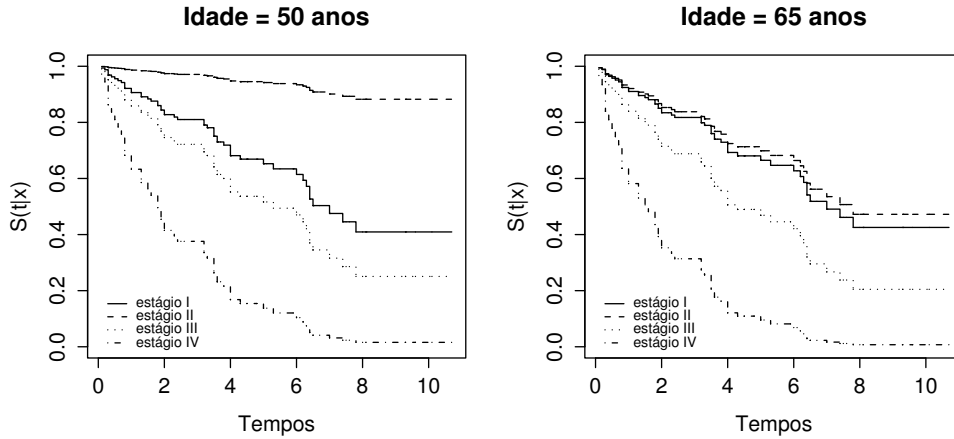


Figura 6: Sobrevivências estimadas pelo modelo de Cox - dados laringe.

Pela Figura 6 pode-se observar que as curvas de sobrevivência estimadas para os estágios I, III e IV não apresentam diferenças muito acentuadas quando comparadas para as idades de 50 e 65 anos. No estágio II, contudo, observa-se um decréscimo expressivo desta curva quando comparada para pacientes de 50 e 65 anos de idade. Este fato justifica,

assim, a presença da interação entre o estágio e idade no modelo, em especial entre idade e o estágio II.

Na Figura 7 encontram-se representadas as correspondentes curvas dos riscos estimados para pacientes com idades de 50 e 65 anos, em cada um dos 4 estágios da doença.

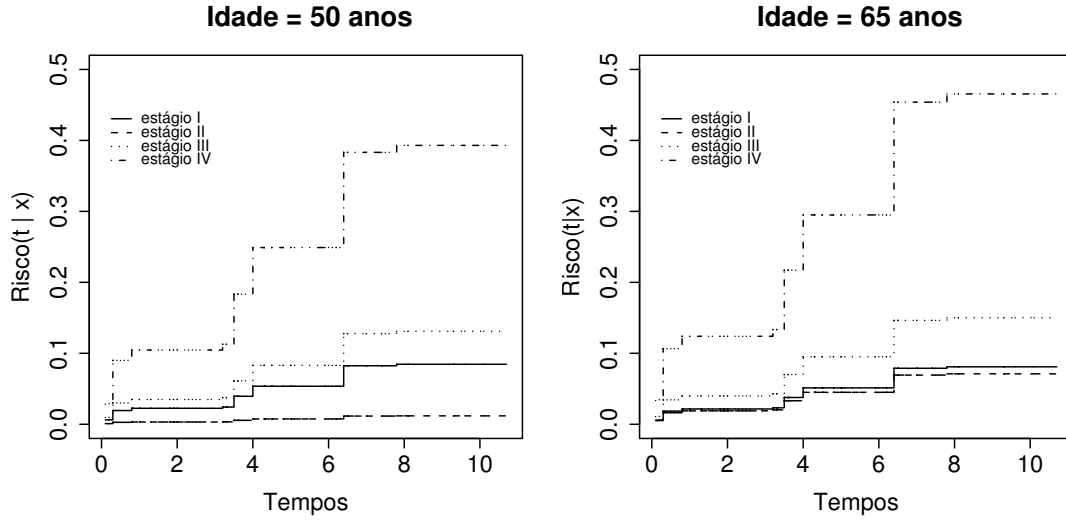


Figura 7: Riscos estimados pelo modelo de Cox - dados laringe.

Assim, por exemplo, para os pacientes  $i$  e  $l$  em que ambos encontram-se no estágio II da doença, mas um deles apresenta idade de 65 anos e outro de 50 anos, tem-se que a razão de riscos entre eles é:

$$\begin{aligned} \frac{\alpha_i(t | \mathbf{x}_i)}{\alpha_l(t | \mathbf{x}_l)} &= \frac{\exp \{ \hat{\beta}_1 + (\hat{\beta}_4 + \hat{\beta}_5) * 65 \}}{\exp \{ \hat{\beta}_1 + (\hat{\beta}_4 + \hat{\beta}_5) * 50 \}} \\ &= \exp \{ (\hat{\beta}_4 + \hat{\beta}_5) * (65 - 50) \} = 5,9983, \end{aligned}$$

o que significa que o risco de falha de pacientes com 65 anos de idade e no estágio II da doença, é de aproximadamente 6 vezes maior do que o risco de falha de pacientes com 50 anos e no mesmo estágio da doença.

Por outro lado, tem-se, por exemplo, para os pacientes  $j$  e  $k$  em que ambos têm 50 anos de idade mas um deles encontra-se no estágio IV da doença e outro no estágio III, que a razão de riscos entre eles é de:

$$\frac{\alpha_j(t | \mathbf{x}_j)}{\alpha_k(t | \mathbf{x}_k)} = \frac{\exp \{ \hat{\beta}_3 + (\hat{\beta}_4 + \hat{\beta}_7) * 50 \}}{\exp \{ \hat{\beta}_2 + (\hat{\beta}_4 + \hat{\beta}_6) * 50 \}} = 3,0006.$$

Desse modo, tem-se que o risco de falha de pacientes com 50 anos de idade e no estágio IV da doença é de 3 vezes maior do que o risco de falha de pacientes também com 50 anos de idade mas que encontram-se no estágio III da doença.

Razões de risco para todas as demais comparações de interesse podem ser obtidas e discutidas de forma análoga a apresentada.

## 2.4 Modelos de riscos proporcionais paramétricos

Para os modelos paramétricos assume-se que a função de risco do indivíduo  $i$  é dada por:

$$\alpha_i(t \mid \mathbf{x}_i) = \alpha_0(t) c(\boldsymbol{\beta}'\mathbf{x}_i) \quad (3)$$

em que,  $\alpha_0$  é a função de risco base e  $c$  é uma função que relaciona  $\boldsymbol{\beta}$ , o vetor de dimensão  $p$  de coeficientes de regressão desconhecidos, e  $\mathbf{x}_i$ , o vetor de dimensão  $p$  de covariáveis observadas para o indivíduo  $i$ . Ao contrário, contudo, do modelo semi-paramétrico de riscos proporcionais de Cox, uma forma paramétrica para  $\alpha_0(t)$  é assumida.

A função de sobrevivência para um indivíduo  $i$  condicional ao vetor de covariáveis  $\mathbf{x}_i$  é, por sua vez, expressa por:

$$S(t \mid \mathbf{x}_i) = \left[ S_0(t) \right]^{c(\boldsymbol{\beta}'\mathbf{x}_i)}.$$

Também para os modelos de riscos proporcionais paramétricos, a razão de riscos de dois indivíduos com distintos valores de  $\mathbf{x}$  são proporcionais, isto é,

$$\frac{\alpha_i(t \mid \mathbf{x}_i)}{\alpha_\ell(t \mid \mathbf{x}_\ell)} = \frac{c(\boldsymbol{\beta}'\mathbf{x}_i)}{c(\boldsymbol{\beta}'\mathbf{x}_\ell)}$$

a qual é uma constante que não depende do tempo.

### 2.4.1 Modelo exponencial de riscos proporcionais

Para a distribuição exponencial cuja função de risco base é  $\alpha_0(t) = 1/\lambda$  tem-se:

$$\alpha(t \mid \mathbf{x}_i) = \frac{1}{\lambda} c(\boldsymbol{\beta}'\mathbf{x}_i)$$

e, desse modo, quando o modelo de Cox é usado para a função  $c$ , tem-se, respectivamente, para as funções de risco e de sobrevivência condicionais,

$$\alpha(t \mid \mathbf{x}_i) = \frac{1}{\lambda} \exp\{\boldsymbol{\beta}' \mathbf{x}_i\}$$

e

$$S(t \mid \mathbf{x}_i) = \left[ \exp \left\{ -\frac{t}{\lambda} \right\} \right]^{\exp\{\boldsymbol{\beta}' \mathbf{x}_i\}}$$

#### 2.4.2 Modelo Weibull de riscos proporcionais

Analogamente, tem-se para a distribuição Weibull que apresenta função de risco base  $\alpha_0(t) = \frac{\gamma}{\lambda^\gamma} t^{\gamma-1}$ ,

$$\alpha(t \mid \mathbf{x}_i) = \frac{\gamma}{\lambda^\gamma} t^{\gamma-1} c(\boldsymbol{\beta}' \mathbf{x}_i)$$

que, quando o modelo de Cox é usado para a função  $c$ , tem-se, respectivamente, para as funções de risco e de sobrevivência condicionais,

$$\alpha(t \mid \mathbf{x}_i) = \frac{\gamma}{\lambda^\gamma} t^{\gamma-1} \exp\{\boldsymbol{\beta}' \mathbf{x}_i\}$$

e

$$S(t \mid \mathbf{x}_i) = \left[ \exp \left\{ -\left( \frac{t}{\lambda} \right)^\gamma \right\} \right]^{\exp\{\boldsymbol{\beta}' \mathbf{x}_i\}}.$$

Obs: Para qualquer outra distribuição (gamma, log-normal etc), tem-se por analogia aos modelos de riscos proporcionais exponencial e Weibull apresentados, suas respectivas funções de risco e de sobrevivência condicionais.

# BIBLIOGRAFIA

- COLOSIMO, E. A. **Análise de Sobrevivência Aplicada**. Piracicaba: Minicurso apresentado na 46a. RBRAS e 9o. SEAGRO, 2001.
- COX, D.R. Regression Models and life tables. **Journal of Royal Statistical Society**. Series B, v.39, p.1-38, 1972.
- COX, D.R. Partial likelihood. **Biometrika**, v. 62, n.2, p.269-276, Mar. 1975.
- GIOLO, S. R. **Modelos de Análise de Sobrevivência para Experimentos Dose-Resposta**. Campinas: Dissertação de Mestrado, 1994.
- GIOLO, S. R. **Variáveis Latentes em Análise de Sobrevivência e Curvas de Crescimento**. Piracicaba: Tese de Doutorado, 2003.
- HOUGAARD, P. **Analysis of Multivariate Survival Data**. New York: Springer Verlag, 2000.
- KAPLAN, E. L.; MEIER, P. Non-parametric estimation from incomplete observations. **Journal of American Statistical Association**, v.53, p.457-481, 1958.
- KLEIN, J. P, MOESCHBERGER, M. **Survival Analysis**. New York: Springer Verlag, 1997.
- LINDSEY, J.C.; RYAN, L.M. Tutorial in bioestatistics: methods for interval-censored data. **Statistics in Medicine**. v.17, p.219-238, 1998.
- LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. New York: John Wiley & Sons, 1981.
- LEE, Elisa. **Statistical Methods for Survival Data Analysis**. Belmont: Lifetime Learning Publications, 1980.
- LOUZADA NETO, F., MAZUCHELI, J., ACHCAR, J.A. **Introdução à Análise de Sobrevivência e Confiabilidade**. Maringá: Minicurso apresentado na III Jornada Regional de Estatística, 2002.

# ANEXOS

## Anexo A. Usando o *software R* para os dados do exemplo 1

### 1. Entrando com os dados diretamente no R

```
require(survival)
temp1<-c(65,156,100,134,16,108,121,4,39,143,56,26,22,1,1,5,65)
cens1<-rep(1,17)
lwbc1<-c(3.36,2.88,3.63,3.41,3.78,4.02,4.00,4.23,3.73,3.85,3.97,4.51,4.54,5.00,5.00,4.72,5.00)
temp2<-c(56,65,17,7,16,22,3,4,2,3,8,4,3,30,4,43)
cens2<-rep(1,16)
lwbc2<-c(3.64,3.48,3.6,3.18,3.95,3.72,4.0,4.28,4.43,4.45,4.49,4.41,4.32,4.90,5.0,5.0)
temp<-c(temp1,temp2)
cens<-c(cens1,cens2)
lwbc<-c(lwbc1,lwbc2)
ag <-c(rep(0,17),rep(1,16))
leuc<-as.data.frame(cbind(temp,cens,ag,lwbc))
leuc
```

### 2. Estimando os coeficientes de regressão

```
fit1<-coxph(Surv(temp,cens) ~ ag, data=leuc)
summary(fit1)
names(fit1)
fit1$loglik
fit2<-coxph(Surv(temp,cens) ~ ag + lwbc, data=leuc)
summary(fit2)
fit2$loglik
```

### 3. Obtendo $S_0(t)$ , $H_0(t)$ e $\alpha_0(t)$

```
ss<-(survfit(coxph(Surv(temp,cens) ~ ag + lwbc), newdata = rbind(c(0,0))))
s0<-round(ss$surv,digits=5) # S0(t) #
H0<- -log(s0) # H0(t) #
x<-as.matrix(H0)
n<-nrow(x)
a0<-rep(0,n)
for(i in 1:n){a0[i]<-H0[i+1] - H0[i]}
alpha0<-c(H0[1],a0[1:(n-1)])
alpha0<-c(H0[1],a0[1:(n-1)]) # a0(t) #
round(cbind(ss$time,s0,alpha0,H0),digits=5)
```

### 4. Resíduos de Schoenfeld e Martingale

```
residuals.coxph(fit2,type="scaledsch")
residuals.coxph(fit2,type="schoenfeld")
residuals.coxph(fit2,type="martingale")
cox.zph(fit2)
par(mfrow=c(1,2))
plot(cox.zph(fit2))
```

### 5. Gráficos das funções de sobrevivência estimadas

```
par(mfrow=c(1,2))
plot.survfit(survfit(coxph(Surv(temp,cens) ~ ag + lwbc), newdata= rbind(c(0,3),c(0,4))),
  lty=1:2,col=c(4,2),xlab="Tempo",ylab="S(t|x)")
title("Ag +")
legend(110,0.85,lty=c(1,2),c("lwbc = 3","lwbc = 4"),lwd=1, bty="n", cex=0.7)
plot.survfit(survfit(coxph(Surv(temp,cens) ~ ag + lwbc), newdata= rbind(c(1,3),c(1,4))),
  lty=1:2,col=c(4,2),xlab="Tempo",ylab="S(t|x)")
title("Ag -")
legend(110,0.85,lty=c(1,2),c("lwbc = 3","lwbc = 4"),lwd=1, bty="n", cex=0.7)
```

## 6. Gráficos das funções de risco estimadas pelo modelo de Cox

```

tt<-sort(temp)
aux1<-as.matrix(tt)
n<-nrow(aux1)
aux2<-as.matrix(cbind(ss$time,alpha0))
alpha00<-rep(min(aux2[,2]),n)
  for(i in 1:n){
    if(tt[i]> min(aux2[,1])){
      i1<- aux2[,1]<= tt[i]
      alpha00[i]<-max(aux2[i1,2])}
  }
talpha0<-cbind(tt,alpha00)
b<-fit2$coefficients
b1<-b[1]
b2<-b[2]
risco1<-alpha00*exp((b1*0) + (b2*3))          # risco estimado para x1=1 e x2=3 #
risco2<-alpha00*exp((b1*0) + (b2*4))          # risco estimado para x1=1 e x2=4 #
plot(risco1,type="s",lty=1,ylim=range(c(0,3.5)),col=4,xlab="Tempo",ylab="Risco estimado")
lines(risco2,type="s",lty=2,col=2)
title("Ag +")
legend(20,2.5,lty=c(1,2),col=c(4,2),c("lwbc = 3","lwbc = 4"),lwd=1, bty="n", cex=0.7)
risco11<-alpha00*exp((b1*1) + (b2*3))          # risco estimado para x1=2 e x2=3 #
risco12<-alpha00*exp((b1*1) + (b2*4))          # risco estimado para x1=2 e x2=3 #
plot(risco11,type="s",lty=1,col=4, ylim=range(c(0,3.5)),xlab="Tempo",ylab="Risco estimado")
lines(risco12,type="s",lty=2,col=2)
title("Ag -")
legend(20,2.5,lty=c(1,2),col=c(4,2),c("lwbc = 3","lwbc = 4"),lwd=1, bty="n", cex=0.7)

```

## Anexo B. Usando o *software R* para os dados do exemplo 2

### 1. Entrando com os dados diretamente no R

```

tempos<-c(0.6,1.3,2.4,3.2,3.3,3.5,3.5,4.4,4.3,5.3,6.6,6.4,6.5,7.4,2.5,3.2,3.3,4.5,4.5,5.5,5.9,5.9,
6.1,6.2,6.5,6.7,7.7,7.4,8.1,8.1,9.6,10.7,0.2,1.8,2,3.6,4,6.2,7,2.2,2.6,3.3,3.6,4.3,4.3,
5,7.5,7.6,9.3,0.3,0.3,0.5,0.7,0.8,1,1.3,1.6,1.8,1.9,1.9,3.2,3.5,5,6.3,6.4,7.8,3.7,4.5,
4.8,4.8,5,5.1,6.5,8,9.3,10.1,0.1,0.3,0.4,0.8,0.8,1,1.5,2,2.3,3.6,3.8,2.9,4.3)
cens<-c(rep(1,15),rep(0,18),rep(1,7),rep(0,10),rep(1,17),rep(0,10),rep(1,11),0,0)
idade<-c(77,53,45,58,76,43,60,52,63,86,81,75,77,67,68,57,51,63,48,68,70,47,58,77,64,79,61,66,73,
56,73,58,68,86,64,63,70,81,74,62,71,67,51,72,47,64,66,50,53,61,49,71,57,79,82,49,60,64,
74,53,72,54,81,59,70,65,68,52,66,54,63,49,69,65,78,69,51,65,71,76,65,78,41,68,69,62,71,
84,74,48)
estagio<-c(rep(1,33),rep(2,17),rep(3,27),rep(4,13))
laringe<-as.data.frame(cbind(tempos,cens,idade,estagio))
i<-order(laringe$tempos)
laringe <-laringe[i,]
laringe
est<-as.factor(laringe$estagio)

```

### 2. Estimando os coeficientes de regressão

```

require(survival)
fit1<-coxph(Surv(tempos,cens) ~ est , data=laringe)
summary(fit1)
fit1$loglik

fit2<-coxph(Surv(tempos,cens) ~ est + idade, data=laringe)
summary(fit2)
fit2$loglik

fit3<-coxph(Surv(tempos,cens) ~ est + idade + est*idade, data=laringe)
summary(fit3)
fit3$loglik

```



3. Obtendo  $S_0(t)$ ,  $H_0(t)$  e  $\alpha_0(t)$ 

```

ss<-survfit(fit3)
round(ss$surv,digits=5)
tt<-sort(tempos)
aux1<-as.matrix(tt)
n<-nrow(aux1)
x<-matrix(0,n,7)
  for(i in 1:n){
    x[i,1]<-ifelse(estagio[i]==2,1,0)
    x[i,2]<-ifelse(estagio[i]==3,1,0)
    x[i,3]<-ifelse(estagio[i]==4,1,0)
    x[i,4]<-idade[i]
    x[i,5]<-x[i,1]*x[i,4]
    x[i,6]<-x[i,2]*x[i,4]
    x[i,7]<-x[i,3]*x[i,4]
  }
x
nx<-ncol(x)
xbar<-matrix(0,nx,1)
  for(i in 1:nx){
    xbar[i]<-mean(x[,i])
  }
xbar
b<-fit3$coefficients
b<-as.vector(b)
embx<-exp(-sum(b*xbar))
s0<- (ss$surv)^embx # So(t) #
H0<- -log(s0) # Ho(t) #
x1<-as.matrix(H0)
n<-nrow(x1)
a0<-rep(0,n)
  for(i in 1:n){a0[i]<-H0[i+1] - H0[i]}
  alpha0<-c(H0[1],a0[1:(n-1)])
alpha0<-c(H0[1],a0[1:(n-1)]) # ao(t) #
round(cbind(ss$time,s0,alpha0,H0),digits=5)

```

## 4. Resíduos de Schoenfeld e Martingale

```

residuals.coxph(fit3,type="scaledsch")
residuals.coxph(fit3,type="schoenfeld")
residuals.coxph(fit3,type="martingale")
cox.zph(fit3)
par(mfrow=c(2,4))
plot(cox.zph(fit3))

par(mfrow=c(2,2))
plot(cox.zph(fit2))

```

## 5. Gráficos das funções de sobrevivência estimadas

```

tt<-sort(tempos)
aux1<-as.matrix(tt)
n<-nrow(aux1)
aux2<-as.matrix(cbind(ss$time,s0))
S00<-rep(max(aux2[,2]),n)
  for(i in 1:n){
    if(tt[i]> min(aux2[,1])){
      i1<- aux2[,1]<= tt[i]
      S00[i]<-min(aux2[i1,2])}
  }
ts0<-cbind(tt,S00)

```

```

b<-fit3$coefficients
id<-50

st1<- S00^(exp(b[4]*id))          # S(t|x) para estágio I   e idade = 50 anos #
st2<- S00^( exp(b[1]+(( b[4]+b[5])*id))) # S(t|x) para estágio II  e idade = 50 anos #
st3<- S00^( exp(b[2]+(( b[4]+b[6])*id))) # S(t|x) para estágio III e idade = 50 anos #
st4<- S00^( exp(b[3]+(( b[4]+b[7])*id))) # S(t|x) para estágio IV  e idade = 50 anos #

id<- 65

st11<- S00^(exp(b[4]*id))          # S(t|x) para estágio I   e idade = 65 anos #
st21<- S00^(exp(b[1]+((b[4]+b[5])*id))) # S(t|x) para estágio II  e idade = 65 anos #
st31<- S00^(exp(b[2]+((b[4]+b[6])*id))) # S(t|x) para estágio III e idade = 65 anos #
st41<- S00^(exp(b[3]+((b[4]+b[7])*id))) # S(t|x) para estágio IV  e idade = 65 anos #

par(mfrow=c(1,2))
plot(tt,st1,type="s",ylim=range(c(0,1)),xlab="Tempos", ylab="S(t|x)", lty=1)
lines(tt,st2,type="s",lty=2)
lines(tt,st3,type="s",lty=3)
lines(tt,st4,type="s",lty=4)
legend(0,0.2, lty=c(1,2,3,4),c("estágio I","estágio II", "estágio III", "estágio IV"),lwd=1, bty="n", cex=0.7)
title(" Idade = 50 anos")

plot(tt,st11,type="s",ylim=range(c(0,1)),xlab="Tempos", ylab="S(t|x)", lty=1)
lines(tt,st21,type="s",lty=2)
lines(tt,st31,type="s",lty=3)
lines(tt,st41,type="s",lty=4)
legend(0,0.2, lty=c(1,2,3,4),c("estágio I","estágio II", "estágio III", "estágio IV"),lwd=1, bty="n", cex=0.7)
title(" Idade = 65 anos")

```

## 6. Gráficos das funções de risco estimadas pelo modelo de Cox

```

tt<-sort(tempos)
aux1<-as.matrix(tt)
n<-nrow(aux1)
aux2<-as.matrix(cbind(ss$time,alpha0))
alpha00<-rep(min(aux2[,2]),n)

for(i in 1:n){
  if(tt[i]> min(aux2[,1])){
    i1<- aux2[,1]<= tt[i]
    alpha00[i]<-max(aux2[i1,2])}
}
talpha0<-cbind(tt,alpha00)

b<-fit3$coefficients
id<-50

rt1<- alpha00*(exp(b[4]*id))          # S(t|x) para estágio I   e idade = 50 anos #
rt2<- alpha00*( exp(b[1]+( ( b[4]+b[5] )*id) ) ) # S(t|x) para estágio II  e idade = 50 anos #
rt3<- alpha00*( exp(b[2]+( ( b[4]+b[6] )*id) ) ) # S(t|x) para estágio III e idade = 50 anos #
rt4<- alpha00*( exp(b[3]+( ( b[4]+b[7] )*id) ) ) # S(t|x) para estágio IV  e idade = 50 anos #

id<- 65

rt11<- alpha00*(exp(b[4]*id))          # S(t|x) para estágio I   e idade = 65 anos #
rt21<- alpha00*(exp(b[1]+( ( b[4]+b[5] )*id) ) ) # S(t|x) para estágio II  e idade = 65 anos #
rt31<- alpha00*(exp(b[2]+( ( b[4]+b[6] )*id) ) ) # S(t|x) para estágio III e idade = 65 anos #
rt41<- alpha00*(exp(b[3]+( ( b[4]+b[7] )*id) ) ) # S(t|x) para estágio IV  e idade = 65 anos #

par(mfrow=c(1,2))
plot(tt,rt1,type="s",ylim=range(c(0,0.5)),xlab="Tempos", ylab="Risco(t | x)", lty=1)
lines(tt,rt2,type="s",lty=2)
lines(tt,rt3,type="s",lty=3)
lines(tt,rt4,type="s",lty=4)
legend(0,0.45, lty=c(1,2,3,4),c("estágio I","estágio II","estágio III","estágio IV"),lwd=1, bty="n", cex=0.7)
title(" Idade = 50 anos")

```

```
plot(tt,rt11,type="s",ylim=range(c(0,0.5)),xlab="Tempos", ylab="Risco(t|x)", lty=1)
lines(tt,rt21,type="s",lty=2)
lines(tt,rt31,type="s",lty=3)
lines(tt,rt41,type="s",lty=4)
legend(0,0.45, lty=c(1,2,3,4),c("estágio I","estágio II","estágio III","estágio IV"),lwd=1, bty="n", cex=0.7)
title(" Idade = 65 anos")
```

## 7. Razão de riscos

```
i) Comparando pacientes no estágio II com 65 e 50 anos
b<-fit3$coefficients
a1<-exp(b[1]+(b[4]+b[5])*65)
a2<-exp(b[1]+(b[4]+b[5])*50)
RR<-a1/a2
RR

ii) Comparando pacientes de 50 anos nos estágios IV e III
a1<-exp(b[3]+(b[4]+b[7])*50)
a2<-exp(b[2]+(b[4]+b[6])*50)
RR<-a1/a2
RR
```