



**UFSM**

**Monografia de Especialização**

**APLICAÇÃO DO MODELO DE REGRESSÃO DE COX**

---

**Maria Helena Rigão**

**PGEMQ**

**Santa Maria, RS, Brasil**

**2004**

# **APLICAÇÃO DO MODELO DE REGRESSÃO DE COX**

---

**por**

**Maria Helena Rigão**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa**

**PGEMQ**

**Santa Maria, RS, Brasil**

**2004**

**Universidade Federal de Santa Maria  
Centro de Ciências Naturais e Exatas  
Curso de Pós-Graduação em Estatística e Modelagem  
Quantitativa**

A Comissão Examinadora, abaixo assinada,  
aprova a Monografia de Especialização

**APLICAÇÃO DO MODELO DE REGRESSÃO DE COX**

elaborada por  
**Maria Helena Rigão**

como requisito parcial para a obtenção do grau de  
**Especialista em Estatística e Modelagem Quantitativa**

---

**COMISSÃO EXAMINADORA:**

---

**Anaelena Bragança de Moraes Ethur**  
(Presidente/Orientadora)

---

**Roselaine Ruviano Zanini**  
(Co-Orientadora)

---

**Adriano Mendonça Souza**

Santa Maria, 4 de junho de 2004

*É preciso amar, amar, amar e nada mais.*

*Pe. Pio de Pietrelcina*

*A*  
*Aurora*

v

Agradeço a todos que contribuíram na elaboração deste trabalho e, particularmente, Anaelena Bragança de Moraes Ethur, orientadora, Roselaine Ruviaro Zanini, co-orientadora, Adriano Mendonça Souza, Ari Edson Poloni de Quevedo, Edith Maria Formentini Chiesa, Maria de Loudes Zanini e Rubem Pereira dos Santos.

## SUMÁRIO

|                                                                   |      |
|-------------------------------------------------------------------|------|
| <b>LISTA DE TABELAS</b> .....                                     | ix   |
| <b>LISTA DE FIGURAS</b> .....                                     | xi   |
| <b>RESUMO</b> .....                                               | xii  |
| <b>ABSTRACT</b> .....                                             | xiii |
| <br>                                                              |      |
| <b>1 INTRODUÇÃO</b> .....                                         | 1    |
| 1.1 Objetivos .....                                               | 3    |
| 1.2 Estrutura do trabalho .....                                   | 3    |
| <br>                                                              |      |
| <b>2 REVISÃO DA LITERATURA</b> .....                              | 4    |
| 2.1 Tempo de sobrevivência, $t$ .....                             | 4    |
| 2.2 Censura .....                                                 | 5    |
| 2.3 Função de sobrevivência $S(t)$ .....                          | 7    |
| 2.4 Função taxa de falha ou taxa de risco, $h(t)$ .....           | 8    |
| 2.5 Estimativa de $S(t)$ e $H(t)$ .....                           | 11   |
| 2.5.1 Função de sobrevivência empírica para dados sem censura.... | 11   |
| 2.5.2 Tabelas de vida.....                                        | 12   |
| 2.5.3 Estimador de Kaplan-Meier.....                              | 13   |
| 2.5.4 Modelo de regressão de Cox.....                             | 14   |
| 2.5.4.1 Estimativa dos coeficientes $\beta$ do modelo de Cox..... | 16   |
| 2.5.4.2 Testes de significância dos coeficientes .....            | 18   |

|          |                                                             |           |
|----------|-------------------------------------------------------------|-----------|
| 2.5.4.3  | Teste para a comparação de modelos .....                    | 20        |
| 2.5.4.4  | Estimação de funções relacionadas com $h_0(t)$ .....        | 20        |
| 2.5.4.5  | Técnicas para verificar a proporcionalidade de riscos ...   | 21        |
| 2.6      | Comparação de curvas de sobrevida .....                     | 23        |
| <b>3</b> | <b>METODOLOGIA .....</b>                                    | <b>26</b> |
| <b>4</b> | <b>RESULTADOS.....</b>                                      | <b>28</b> |
| 4.1      | Descrição da amostra.....                                   | 28        |
| 4.2      | Análise de sobrevida usando o estimador de Kaplan-Meier     | 29        |
| 4.2.1    | Análise de sobrevida por tipo de cirurgia.....              | 30        |
| 4.2.2    | Análise de sobrevida por estágio da doença.....             | 31        |
| 4.2.3    | Análise de sobrevida por tipo de tratamento cirúrgico ..... | 32        |
| 4.3      | Modelo de regressão de Cox.....                             | 33        |
| 4.3.1    | Modelo 1 - Com a covariável cirurgia.....                   | 34        |
| 4.3.2    | Modelo 2 - Com a covariável estágio.....                    | 35        |
| 4.3.3    | Modelo 3 - Com a covariável idade .....                     | 36        |
| 4.3.4    | Modelo 4 - Com a covariável tratamento .....                | 37        |
| 4.3.5    | Modelo 5 - Com todas as covariáveis significativas.....     | 39        |
| 4.3.6    | Modelo 6 - Com as covariáveis cirurgia e estágio.....       | 40        |
| 4.3.7    | Modelo final .....                                          | 41        |
| 4.3.8    | Teste gráfico para a proporcionalidade .....                | 42        |
| 4.3.9    | Teste de covariável dependente do tempo.....                | 43        |
| <b>5</b> | <b>CONCLUSÃO .....</b>                                      | <b>46</b> |
| <b>6</b> | <b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>                     | <b>49</b> |
| <b>7</b> | <b>APÊNDICE.....</b>                                        | <b>50</b> |

## LISTA DE TABELAS

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| TABELA 01 - Tabela de contingência gerada no tempo $t_j$ .....                       | 24 |
| TABELA 02 - Distribuição dos pacientes.....                                          | 29 |
| TABELA 03 - Distribuição por cirurgia .....                                          | 34 |
| TABELA 04 - Modelo 1 de regressão de Cox - covariável cirurgia..                     | 34 |
| TABELA 05 - Teste da razão de verossimilhança do modelo 1 .....                      | 35 |
| TABELA 06 - Distribuição por estágio da doença.....                                  | 36 |
| TABELA 07 - Modelo 2 de regressão de Cox - covariável estágio...                     | 36 |
| TABELA 08 - Teste da razão de verossimilhança do modelo 2.....                       | 36 |
| TABELA 09 - Distribuição por idade.....                                              | 37 |
| TABELA 10 - Modelo 3 de regressão de Cox - covariável idade .....                    | 37 |
| TABELA 11 - Teste da razão de verossimilhança do modelo 3.....                       | 37 |
| TABELA 12 - Distribuição por tratamento.....                                         | 38 |
| TABELA 13 - Modelo 4 de regressão de Cox - covariável tratamen-<br>to .....          | 38 |
| TABELA 14 - Teste da razão de verossimilhança do modelo 4.....                       | 38 |
| TABELA 15 - Distribuição por cirurgia, estágio e tratamento.....                     | 39 |
| TABELA 16 - Modelo 5 de regressão de Cox - análise múltipla das<br>covariáveis ..... | 39 |
| TABELA 17 - Teste da razão de verossimilhança do modelo 5.....                       | 40 |
| TABELA 18 - Distribuição por cirurgia e estágio .....                                | 40 |

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| TABELA 19 - Modelo 6 de regressão de Cox - covariáveis cirurgia e estágio ..... | 41 |
| TABELA 20 - Teste da razão de verossimilhança do modelo 6.....                  | 41 |
| TABELA 21 - Modelo de regressão de Cox final, com a covariável estágio .....    | 42 |
| TABELA 22 - Modelo com a covariável dependente do tempo.....                    | 43 |
| TABELA 23 - Banco de dados.....                                                 | 51 |

## LISTA DE FIGURAS

|                                                                  |    |
|------------------------------------------------------------------|----|
| FIGURA 1 - Curva de sobrevida.....                               | 8  |
| FIGURA 2 - Curva de sobrevida global.....                        | 29 |
| FIGURA 3 - Curvas de sobrevida por tipo de cirurgia .....        | 30 |
| FIGURA 4 - Curvas de sobrevida por estadiamento da doença.....   | 31 |
| FIGURA 5 - Curvas de sobrevida por tipo de tratamento .....      | 33 |
| FIGURA 6 - Curvas comparativas por estágio da doença.....        | 42 |
| FIGURA 7 - Gráficos da função de sobrevida por estágio da doença | 44 |

## **RESUMO**

Monografia de Especialização  
Curso de Pós-Graduação em Estatística e Modelagem Quantitativa  
Universidade Federal de Santa Maria, RS, Brasil

### **APLICAÇÃO DO MODELO DE REGRESSÃO DE COX**

Autora: Maria Helena Rigão

Orientadora: Anaelena Bragança de Moraes Ethur

Co-Orientadora: Roselaine Ruviaro Zanini

Santa Maria, 4 de junho de 2004.

O principal propósito, desta pesquisa, foi estudar a análise de sobrevida, ou técnicas estatísticas para dados censurados. Nessa análise, a variável resposta é o tempo até a ocorrência de um evento, denominado tempo de falha. Neste trabalho, realizou-se o estudo do modelo de regressão de Cox e a aplicação do mesmo em dados da área médica. Para essa análise, utilizou-se dados obtidos no Hospital Universitário de Santa Maria, referentes a 69 pacientes atendidos e acompanhados no período de 1991 a 1999. Após diagnóstico confirmado de câncer no aparelho digestivo, foram acompanhados até a morte, ou o fim do estudo, ou a perda do seguimento. As covariáveis avaliadas foram: idade, tipo de cirurgia, estágio da doença, tipo de tratamento cirúrgico e data de óbito, ou de censura. A função de sobrevida foi estimada através do método de Kaplan-Meier, e os teste Log-rank e qui-quadrado foram utilizados para a comparação das curvas de sobrevida. Na modelagem dos dados, utilizou-se o modelo de regressão de Cox. A idade média dos pacientes foi de 60 anos, sendo que 71% deles foi a óbito. O tempo mediano de sobrevida foi de 12 meses. A análise de sobrevida, através do método Kaplan-Meier e teste Log-rank, mostrou que as curvas de sobrevida, por tipo de cirurgia, não apresentaram diferença significativa ( $p > 0,05$ ), sendo que as curvas de sobrevida, por estágio de doença e por tipo de tratamento, foram significativamente diferentes ( $p < 0,05$ ). A regressão de Cox mostrou que o modelo ajustado, para os dados, incluiu apenas a covariável estágio ( $p = 0,001$ ), verificado pelo teste da razão de verossimilhança parcial e teste de Wald. Concluiu-se, pelo modelo encontrado, que o risco de óbito aumenta com o avanço do estágio da doença.

## **ABSTRACT**

Monografia de Especialização  
Curso de Pós-Graduação em Estatística e Modelagem Quantitativa  
Universidade Federal de Santa Maria, RS, Brasil

### **APLICAÇÃO DO MODELO DE REGRESSÃO DE COX** (The Cox Regression Model Application)

Author: Maria Helena Rigão

Adviser: Anaelena Bragança de Moraes Ethur

Co-Adviser: Roselaine Ruviaro Zanini

Santa Maria, June, 4<sup>th</sup>, 2004.

The main purpose of this research was study the survival analyses by statistics techniques to censored data. In this case the outcome variable is the time until the event occur, called failure time. In this research we studied the Cox Proportional Hazard Model with application in medical data. For the analyses we used data from Hospital Universitário de Santa Maria using a follow-up of 69 patients enrolled in the study from 1991 to 1999. After a confirmed diagnosis of system digestive cancer, members were followed until death or until the end of the study or until the subject was lost to follow-up. The covariates analyzed were: age, type of surgery, disease stage, type of treatment and date of the death or censored. The survival function was estimated by Kaplan-Meier method and Log-rank and Chi-square tests were used to compare the survival curves. The Cox Proportional Hazard Model was used to model the survival time data. The mean age was 60 years and 71% died because the cancer. The median survival time was 12 month. The Kaplan-Meier and Log-rank analysis showed that the survival curves by surgery type were not significantly different ( $p>0.05$ ) and the survival curves by disease stage and treatment type was significantly different ( $p<0.05$ ). The Cox regression model obtained included just the covariate disease stage ( $p=0.001$ ), tested by partial likelihood ratio and Wald tests. The conclusion was that the death risk increase with the increased of the disease stage and this is expected.

## 1 INTRODUÇÃO

A análise de sobrevivência é o conjunto de técnicas e modelos estatísticos utilizados na análise do comportamento de variáveis aleatórias positivas. É uma das áreas da estatística que mais tem crescido nos últimos anos. Nesse tipo de análise, a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Esse tempo é denominado tempo de falha, podendo ser o tempo até a falha de um equipamento, a morte de um paciente, bem como até a cura ou recidiva de uma doença. A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta (Colosimo, 2001).

Em engenharia, são comuns os estudos onde produtos, ou componentes, são colocados sob teste para se estimar características relacionadas aos seus tempos de vida. Os engenheiros denominam essa área de confiabilidade. O mesmo acontece em ciências sociais, onde várias situações de interesse são estudadas envolvendo dados censurados. Estudiosos do trabalho concentram-se em mudanças de empregos, promoções e aposentadoria. Demógrafos pesquisam nascimentos, mortes, casamentos, divórcios e migrações. Educadores estudam o tempo de permanência de alunos na escola. Na área da saúde, deseja-se identificar fatores de prognóstico, ou de risco, para

uma certa doença, ou comparar tratamentos de interesse em um estudo clínico controlado (Colosimo, 2001).

Assim como em qualquer análise estatística, o passo inicial da análise de sobrevida consiste na descrição dos dados. Técnicas estatísticas não-paramétricas são, freqüentemente, utilizadas para esse fim. Entre elas, o estimador de Kaplan-Meier para a função de sobrevida, e o teste Log-rank para comparar funções de sobrevida.

A análise estatística final de dados de sobrevida é, geralmente, realizada utilizando-se modelos de regressão. Existem duas classes de modelos de regressão: os modelos paramétricos, denominados modelos de tempo de vida acelerada, e os modelos semi-paramétricos, chamados modelos de riscos proporcionais, ou modelo de regressão de Cox, que é o assunto do presente trabalho.

O modelo de regressão de Cox, extensivamente usado em estudos médicos, é composto pelo produto de dois componentes, um não-paramétrico e outro paramétrico. Segundo Kleinbaum (1996), a popularidade do modelo justifica-se pela presença do componente não paramétrico, que dá maior flexibilidade ao modelo. Sua importância, na pesquisa clínica, pode ser evidenciada pela quantidade de citações feitas em publicações especializadas da área.

Pelo citado acima, justifica-se o interesse no estudo do modelo de regressão de Cox, bem como uma aplicação do mesmo na área médica, contribuindo com um texto que sirva como um subsídio teórico para futuras pesquisas.

## **1.1 Objetivos**

### **1.1.1 Objetivo geral**

Realizar um estudo sobre o modelo de regressão de Cox, bem como uma aplicação utilizando dados da área médica.

### **1.1.2 Objetivos específicos**

- Estudar o modelo de regressão de Cox;
- Construir um material teórico, para subsídio de outros pesquisadores;
- Aplicar o método em dados reais da área médica;
- Identificar as questões relacionadas a bancos de dados e suas implicações na utilização de programas computacionais estatísticos.

## **1.2 Estrutura do trabalho**

O presente trabalho é estruturado da seguinte forma:

- No Capítulo 1 tem-se a introdução, objetivos e estrutura do trabalho;
- No Capítulo 2 apresenta-se a revisão de literatura, constando da teoria relacionada ao assunto;
- No Capítulo 3 descreve-se o método utilizado no trabalho;
- No Capítulo 4 apresenta-se a aplicação e a discussão dos resultados;
- No capítulo 5 apresenta-se as conclusões do estudo.

As referências bibliográficas, e o apêndice, encerram o presente trabalho.

## 2 REVISÃO DA LITERATURA

A análise de sobrevivência, ou de sobrevida, é uma área da estatística que se ocupa do estudo de técnicas e modelos utilizados no estudo do tempo de sobrevivência, ou seja, do tempo até a ocorrência de um evento de interesse, sendo o tempo a variável resposta (Colosimo, 2001).

A diferença entre a análise de sobrevida, e as outras análises estatísticas, conforme Kleinbaun (1996), é a presença de dados censurados, para os quais o tempo de sobrevida não é determinado. A análise de sobrevida, então, fornece métodos capazes de incorporar as informações que são proporcionadas por dados censurados. Os métodos mais utilizados são: Função de sobrevida empírica; Tabelas de vida; Estimador de Kaplan-Meier; Modelo de regressão de Cox.

A seguir, serão definidos alguns conceitos importantes em análise de sobrevida para, posteriormente, desenvolver-se a teoria do modelo de regressão de Cox, que é o objeto deste trabalho.

### 2.1 Tempo de sobrevida, $t$

É o tempo observado desde a entrada do indivíduo, no estudo, até a ocorrência de um evento de interesse. Pode-se observar três elementos que constituem o tempo de sobrevida:

- Tempo inicial

Refere-se ao tempo (data) de início do estudo. Durante o período pré-determinado, para o estudo, pode-se observar diferentes indivíduos com diferentes datas de inclusão.

- Escala de medida

Refere-se ao tempo real de observação, ou seja, pode-se definir estudos onde o tempo observado é o número de anos que o indivíduo levou até a ocorrência do evento de interesse, ou, ainda, o número de semestres, número de meses, etc.

- Evento de interesse ou falha

Refere-se ao evento de interesse que, geralmente, é indesejável, como, por exemplo, a morte de um paciente, a recidiva de uma doença, etc.

## **2.2 Censura**

Ao analisar dados de sobrevida, muitas vezes ocorrem informações incompletas sobre o tempo de sobrevida, caracterizando, assim, a presença de dados censurados.

São, geralmente, três as razões pelas quais a censura pode ocorrer:

- Não ocorre o evento de interesse antes do fim do estudo;
- O indivíduo é “perdido” durante o período de estudo;
- O indivíduo retirou-se do estudo porque morreu (se a morte não é o evento de interesse), ou alguma outra razão.

É importante observar que todos os dados devem ser utilizados na análise, pois mesmo que as informações sejam incompletas, ou

censuradas, elas fornecem informações sobre o tempo de vida dos indivíduos, e a omissão desses dados, nos cálculos estatísticos, levaria a conclusões viesadas.

Os tipos de censuras existentes são:

- Censura Tipo I

É a que ocorre quando o estudo termina após um período pré-estabelecido de tempo. Quando uma amostra de  $n$  indivíduos entra junto no experimento, possuindo, assim, o mesmo tempo técnico de observação, o tipo de censura é dita simples; quando os indivíduos entram no experimento em tempos diferentes, tem-se a censura do Tipo I múltipla.

- Censura Tipo II

Ocorre quando uma amostra de  $n$  indivíduos é observada até que  $r$  ( $r < n$ ) falhas ocorram. Assim, colocam-se  $n$  unidades em teste e termina-se o experimento após a ocorrência da  $r$ -ésima falha. O número de falhas  $r$  é fixado antes da realização do experimento e o tempo, para o término do experimento, é aleatório.

Os  $n-r$  indivíduos restantes são as censuras do Tipo II. Ela é dita censura simples quando todos os indivíduos são colocados ao mesmo tempo no experimento, e múltipla quando os indivíduos entram em tempos distintos no experimento.

- Censura do Tipo aleatória

É a que mais ocorre na prática médica, quando o paciente é retirado no decorrer do estudo sem ter ocorrido a falha, ou quando o paciente morre por uma razão diferente da estudada. Os indivíduos entram no estudo de modo aleatório de acordo com o tempo de estudo.

Na censura aleatória podem ocorrer os seguintes casos:

- Observações perdidas: o indivíduo muda-se para outra localidade desconhecida, ou abandona o tratamento;
- Retirado do experimento: o indivíduo não se adapta ao tratamento e é necessário retirá-lo;
- Término do estudo: o tempo pré-determinado para o estudo termina e o indivíduo sobreviveu a esse tempo.

### 2.3 Função de sobrevida, $S(t)$

A função de sobrevida determina as probabilidades de sobrevida para diferentes valores de  $t$ , ou seja,  $S(t)$  é a probabilidade de um indivíduo sobreviver durante algum tempo especificado  $t$  (Colosimo, 2001).

Se  $T$  é uma variável aleatória que representa o tempo de sobrevida até o evento falha, a função de sobrevida é:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad (1)$$

sendo  $F(t)$  a função acumulada de probabilidade, que representa a probabilidade de um indivíduo sobreviver, a partir do tempo de origem, mais de  $t$  unidades de tempo (Colosimo, 2001).

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad (2)$$

As principais características da função de sobrevida são:

- $S(t)$  é uma função decrescente;

- No tempo  $t = 0$ ,  $S(t) = 1$ , isto é, no começo do estudo, uma vez que não possa ter ocorrido ainda nenhum evento, a probabilidade de sobreviver é 1.
- $\lim_{t \rightarrow \infty} S(t) = 0$ , isto é, a medida que o período de estudo cresce, diminui a probabilidade de sobrevivência, o que faz com que a curva caia para zero (Colosimo, 2001).

A representação gráfica da função de sobrevivência  $S(t)$  é mostrada na Figura 1.

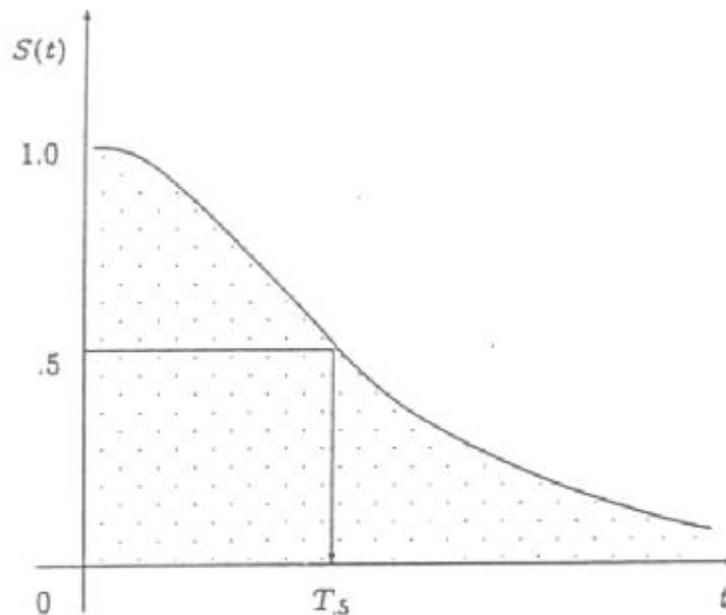


FIGURA 1 – Curva de sobrevivência

#### 2.4 Função taxa de falha ou taxa de risco, $h(t)$

A probabilidade de um indivíduo falhar no intervalo  $[t_1, t_2)$  é

$$P(t_1 \leq T < t_2) = S(t_1) - S(t_2) \quad (3)$$

pois:

$$\begin{aligned}
P(t_1 \leq T < t_2) &= P(T < t_2) - P(T \leq t_1) \\
&= [1 - S(t_2)] - [1 - S(t_1)] \\
&= S(t_1) - S(t_2)
\end{aligned}$$

A taxa de falha no intervalo  $[t_1, t_2)$  é a probabilidade de um indivíduo falhar nesse intervalo, dado que não tenha falhado antes de  $t_1$ , dividida pelo tamanho do intervalo, isto é:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} \quad (4)$$

pois:

$$\begin{aligned}
P(t_1 \leq T < t_2 / T \geq t_1) &= \frac{P([t_1 \leq T < t_2] \cap [T \geq t_1])}{\Delta t P(T \geq t_1)} \\
&= \frac{P(t_1 \leq T < t_2)}{\Delta t P(T \geq t_1)} \\
&= \frac{S(t_1) - S(t_2)}{\Delta t S(t_1)}
\end{aligned}$$

Redefinindo o intervalo  $[t_1, t_2)$  como  $[t, t + \Delta t)$ , a taxa de falha, nesse intervalo passa a ser:

$$P(t_1 \leq T < t_2 / T \geq t_1) = \frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

Fazendo  $\Delta t$  bem pequeno, isto é,  $\Delta t \rightarrow 0$ , tem-se a função taxa de falha, definida como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

que representa a taxa de falha instantânea, condicional à sobrevivência do indivíduo até o tempo  $t$  (Colosimo, 2001).

A função taxa de falha  $h(t)$  é útil para descrever o tempo de vida dos indivíduos, ou seja, descreve a forma em que a taxa instantânea de falha muda com o tempo, podendo ser constante, crescente ou decrescente.

Desenvolvendo  $h(t)$ , tem-se:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

Substituindo  $S(t)$  por  $1 - F(t)$ , onde  $F(t)$  é a função acumulada de probabilidade de  $T$ , tem-se:

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{[1 - F(t)] - [1 - F(t + \Delta t)]}{\Delta t S(t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{S(t)} \end{aligned}$$

então

$$h(t) = \frac{f(t)}{S(t)}, \text{ onde } f(t) \text{ é a função da distribuição de } T.$$

Sabe-se que

$$f(t) = -\frac{\partial}{\partial t} S(t) = -S'(t)$$

assim:

$$h(t) = \frac{-\frac{\partial}{\partial t} S(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

Integrando a função taxa de falha  $h(t)$ , tem-se a função acumulada de taxa de falha, indicada por  $H(t)$ , definida por:

$$H(t) = \int_0^t \frac{-S'(t)}{S(t)}$$

$$H(t) = -\ln S(t)$$

Pode-se, então, escrever a função de sobrevivência  $S(t)$  em função da taxa de falha acumulada  $H(t)$ .

$$S(t) = e^{-H(t)}$$

## 2.5 Estimação de $S(t)$ e $H(t)$

A estimação da função de sobrevivência  $S(t)$ , e da função acumulada da taxa de falha  $H(t)$ , realizada a partir dos tempos de sobrevivência observados, pode ser feita utilizando-se procedimentos paramétricos ou procedimentos não-paramétricos. Os procedimentos paramétricos, são aqueles onde a forma da distribuição de probabilidade de  $T$  é conhecida. As distribuições mais utilizadas, para descrever a variável  $T$ , nesses casos, são: distribuição exponencial, Weibull, Gamma e lognormal. Quando a estimação é feita utilizando-se procedimentos não-paramétricos, a forma da distribuição de probabilidade de  $T$  não precisa ser especificada. Dentre esses procedimentos, destacam-se: Função de sobrevivência empírica; Tabelas de vida; Estimador de Kaplan-Meier e o Modelo de regressão de Cox.

### 2.5.1 Função de sobrevivência empírica para dados sem censura

É a função utilizada na ausência de censura, definida como:

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ de observações no estudo}} \quad (5)$$

$\hat{S}(t)$  é uma função escada com degraus nos tempos observados de falha de tamanho  $1/n$ , onde  $n$  é o tamanho da amostra. Se existirem empates em um certo tempo  $t$ , o tamanho do degrau fica multiplicado pelo número de empates.

### 2.5.2 Tabelas de vida

São procedimentos de estimação que mostram a estrutura do tempo de sobrevivência para grupos homogêneos de indivíduos, quando esses tempos são submetidos à censura. Quase sempre, em tabelas de vida, os intervalos de tempo são de mesmo comprimento, o qual deve ser tão pequeno quanto possível, de acordo com a precisão dos valores observados, para permitir melhores estimadores.

As tabelas de vida são instrumentos estatísticos antigos, sendo usadas pelas companhias de seguro desde o século XVII (Soares & Bartman, 1983).

A construção de uma tabela de vida consiste em dividir o eixo do tempo em um certo número de intervalos de tempo. Supondo que o eixo do tempo foi dividido em  $k+1$  intervalos, com os seguintes pontos de corte:  $t_1, t_2, \dots, t_k$ . Assim:

$$I_i = [t_{i-1}, t_i); i = 1, \dots, k+1$$

sendo:

$$t_0 = 0;$$

$$t_{k+1} = +\infty.$$

Então, o estimador da tabela de vida é dado por:

$$S(t) = (1 - q_1)(1 - q_2) \dots (1 - q_i), \text{ com } t \in [t_{i-1}, t_i), \quad (6)$$

sendo:

$q_i$  = probabilidade de um indivíduo falhar no intervalo  $[t_{i-1}, t_i)$  sabendo-se que não falhou até  $t_{i-1}$ , e  $t_0 = 0$ .

Para estimar  $q_i$ , utiliza-se:

$$\hat{q}_i = \frac{\text{n}^\circ \text{ de falhas no intervalo } [t_{i-1}, t_i)}{\text{n}^\circ \text{ sob risco em } t_{i-1} - \frac{\text{n}^\circ \text{ de censuras em } [t_{i-1}, t_i)}{2}} \quad (7)$$

sendo que o segundo termo do denominador da equação (7) indica que as observações censuradas, no intervalo  $[t_{i-1}, t_i)$ , são tratadas como se estivessem sob risco durante a metade do intervalo considerado. Uma estimativa gráfica, para a função de sobrevida, é uma função escada, com valor constante em cada intervalo de tempo (Colosimo, 2001).

### 2.5.3 Estimador de Kaplan-Meier

Este método, proposto por Kaplan-Meier (1958), é utilizado para a estimação não-paramétrica da função de sobrevida, e também é conhecido por método do limite-produto.

O estimador de Kaplan-Meier, na sua construção, considera o número de intervalos de tempo equivalentes ao número de falhas distintas. As probabilidades de falha são calculadas somente no instante em que elas ocorrerem.

Suponha-se que  $n$  indivíduos sejam observados e, além disso, que o tempo de falha é observado para um dos indivíduos em  $k$  tempos distintos  $t_1 < t_2 < \dots < t_k$ .

Seja também,  $d_j$  o número de falhas,  $w_j$  o número de observações censuradas e  $n_j$  o número de indivíduos sob risco de falhas no tempo  $t_j$ .

Se uma censura e uma falha ocorrerem ao mesmo tempo, considera-se a falha antes da censura.

A estimativa da curva de sobrevivência, usando o estimador de Kaplan-Meier é dada por:

$$\hat{S}(t) = \prod_{j/t_j < t} 1 - \frac{d_j}{n_j}, \quad (8)$$

sendo:

$$n_j = n - d_{j-1} - w_{j-1}$$

$n$  = número de indivíduos observados;

$n_j$  = números de indivíduos sob risco, isto é, que não falharam e não foram censurados até o tempo  $t_j$ , exclusive;

$d_{j-1}$  = número de falhas no tempo  $t_{j-1}$ ;

$w_{j-1}$  = número de observações censuradas no tempo  $t_{j-1}$ .

#### 2.5.4 Modelo de regressão de Cox

O modelo de regressão de Cox, ou modelo de riscos proporcionais (Cox, 1972), é um modelo de regressão semi-paramétrico, que permite a análise de dados provenientes de estudos,

onde a resposta medida é o tempo até a ocorrência de um evento de interesse, ajustado por covariáveis ( $x_1, \dots, x_p$ ).

No caso geral o modelo assume a forma

$$h(t) = h_0(t) g(X' \beta) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_p x_p) \quad (9)$$

sendo:

$g$  = função a ser especificada.

Esse modelo é composto pelo produto de dois componentes, um não-paramétrico e outro paramétrico (Colosimo, 2001).

O componente não-paramétrico,  $h_0(t)$  não é especificado e é uma função não-negativa do tempo, usualmente chamada de função de risco base ou baseline. É a probabilidade de ocorrer o evento até o tempo ( $t$ ), quando não existirem covariáveis no modelo, ou seja, todos os  $\beta$  serão iguais a zero. Em estudos onde o objetivo é estimar valores prognósticos, não há interesse em defini-la, pois ela é comum a todos os indivíduos (Hosmer & Lemeshow, 1999).

O componente paramétrico  $g(X' \beta)$ , é uma combinação linear das covariáveis, e assumirá um valor diferente para cada uma delas, sendo  $\beta$  o vetor de parâmetros associado as covariáveis.

Esse modelo é chamado de modelo de riscos proporcionais, porque a razão das taxas de falhas (HR) entre dois indivíduos distintos,  $i$  e  $j$ , é constante no tempo, pois:

$$HR_{(i,j)} = \frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(X'_i \beta)}{h_0(t) \exp(X'_j \beta)} = \exp(X'_i \beta - X'_j \beta) \quad (10)$$

$$HR_{(t,i,j)} = \exp\beta(X'_i - X'_j) \quad (11)$$

Se, no início do estudo, o indivíduo  $i$  tem um risco de falha igual a duas vezes o risco do indivíduo  $j$ , então esta razão será a mesma para todo o período de acompanhamento, ou seja, não dependerá do tempo, sendo, esta, uma das pressuposições do modelo a ser testada.

#### 2.5.4.1 Estimação dos coeficientes $\beta$ do modelo de Cox

O modelo de regressão de Cox é caracterizado pelos coeficientes  $\beta$ , que medem o efeito das covariáveis sobre a função de taxa de falha. Essas quantidades devem ser estimadas a partir das observações amostrais, para que o modelo fique determinado. O método de máxima verossimilhança torna-se inapropriado para a estimação dos coeficientes  $\beta$ , devido à presença da componente não-paramétrica  $h_0(t)$ . Uma solução razoável consiste em condicionar a verossimilhança, para eliminar esta função de perturbação, sugerida por Cox, denominado de método de máxima verossimilhança parcial, ou condicional (Cox, 1972,1975).

Uma forma simples de entender a verossimilhança parcial considera o seguinte argumento condicional: a probabilidade condicional da  $i$ -ésima observação vir a falhar no tempo  $t_i$ , conhecendo quais observações estão sob risco em  $t_i$  é:

$$\frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_i)} = \frac{h_0(t) \exp(X'_i \beta)}{\sum_{j \in R(t_i)} h_0(t) \exp(X'_j \beta)} = \frac{\exp(X'_i \beta)}{\sum_{j \in R(t_i)} \exp(X'_j \beta)} \quad (12)$$

sendo:

$R(t_i)$  = conjunto dos índices das observações sob risco no tempo  $t_i$ .

Assim, ao condicionar falhas e censuras até o tempo  $t_i$ , o componente não-paramétrico  $h_0(t)$  desaparece em (12).

A função de verossimilhança a ser usada, para fazer inferências no modelo, é então formada pelo produto de todos os termos representados por (12) associados aos distintos tempos de falha

$$L(\beta) = \prod_{i=1}^k \frac{\exp(X'_i \beta)}{\sum_{j \in R(t_i)} \exp(X'_j \beta)} = \left[ \prod_{i=1}^n \frac{\exp(X'_i \beta)}{\sum_{j \in R(t_i)} \exp(X'_j \beta)} \right]^{\delta_i} \quad (13)$$

sendo:

$\delta_i$  = indicador de falha.

Os valores de  $\beta$  que maximizam  $L(\beta)$ , a função de verossimilhança parcial, são obtidos resolvendo o sistema de equações definido por  $U(\beta)=0$ , onde  $U(\beta)$  é o vetor escore de primeiras derivadas da função  $\ln(L(\beta))$ . Isto é:

$$U(\beta) = \sum_{i=1}^n \left[ \delta_i X_i - \frac{\sum_{j \in R(t_i)} X_j \exp(X'_j \beta)}{\sum_{j \in R(t_i)} \exp(X'_j \beta)} \right] = 0 \quad (14)$$

A função de verossimilhança parcial, ou condicional (13), assume que os tempos de sobrevivência são contínuos e, conseqüentemente, não pressupõe a possibilidade de empates nos valores observados. Na prática, podem ocorrer empates nos tempos de falhas, ou censuras, devido a escala de medida. Quando ocorrem

empates entre falhas e censuras, usa-se a convenção de que a censura ocorreu após a falha, o que define as observações a serem incluídas no conjunto de risco em cada tempo de falha (Colosimo, 2001).

Quando existem observações empatadas, a verossimilhança parcial (14) deve ser modificada. Breslow e Peto *apud* Colosimo (2001) propõem a seguinte aproximação.

$$L(\beta) = \prod_{i=1}^k \frac{\exp(s'_i \beta)}{[\sum_{j \in R(t_i)} \exp(s'_j \beta)]^{d_i}} \quad (15)$$

sendo:

$s_i$  = vetor formado pela soma das correspondentes  $p$  covariáveis para os indivíduos que falham no tempo  $t_i = 1, \dots, k$ ;

$d_i$  = número de falhas nesse mesmo tempo.

Essa aproximação é adequada quando o número de observações empatadas, em qualquer tempo, não é grande.

#### 2.5.4.2 Testes de significância dos coeficientes

Após a estimação dos coeficientes  $\beta$ , sua significância deve ser testada, utilizando-se o teste de Wald, ou o teste da razão de verossimilhança parcial.

- Teste de Wald

O teste de Wald é obtido da comparação de estimativas de verossimilhança dos parâmetros  $\beta$  com a estimativa do seu erro padrão. A razão resultante, sob a hipótese de que  $\beta=0$ , terá uma distribuição normal padrão (Hosmer & Lemeshow, 1999).

$$W = \frac{\hat{\beta}_i}{EP(\hat{\beta}_i)} \sim N(0,1) \quad (16)$$

sendo:

$\hat{\beta}_i$  = coeficiente;

$EP(\hat{\beta}_i)$  = erro padrão do coeficiente.

Para o teste de Wald também pode-se utilizar

$$W^2 = \left( \frac{\hat{\beta}_i}{EP(\hat{\beta}_i)} \right)^2 \sim \chi_1^2 \quad (17)$$

ou seja,  $W^2$  segue uma distribuição qui-quadrado com 1 grau de liberdade.

- Teste da razão da verossimilhança parcial

Esse teste é dado por:

$$\chi^2 = 2 [\ln L(0) - \ln L(\beta)], \quad (18)$$

sendo:

$\ln L(0)$  = valor do logaritmo da verossimilhança do modelo básico (sem covariáveis);

$\ln L(\beta)$  = valor do logaritmo da verossimilhança do modelo a ser testado (com covariável).

Sob a hipótese nula, de que os coeficientes são iguais a zero, a estatística do teste segue uma distribuição qui-quadrado, com um grau

de liberdade, para uma covariável contínua, e n-1 graus de liberdade se a covariável for categórica, sendo n a quantidade de categorias, ou faixas em que essa covariável é dividida (Hosmer & Lemeshow, 1999).

#### 2.5.4.3 Teste para a comparação de modelos

Para a comparação dos modelos utiliza-se o teste da razão da verossimilhança parcial. A estatística do teste é:

$$\chi^2 = 2 [\ln L(\text{modelo maior}) - \ln L(\text{modelo menor})] \quad (19)$$

que segue uma distribuição de  $\chi^2$  com o número de graus de liberdade igual a diferença do número de covariáveis entre os dois modelos, onde  $\ln L$  é o valor do logaritmo da verossimilhança de cada modelo ajustado (Fisher & Van Belle, 1995).

#### 2.5.4.4 Estimação de funções relacionadas com $h_0(t)$

A função taxa de falha acumulada é

$$H_0(t) = \int_0^t h_0(u) du \quad (20)$$

sendo:

$u = f(t)$  e a correspondente função de sobrevivência é:

$$S_0(t) = \exp(- H_0(t)) \quad (21)$$

A função de sobrevida

$$S(t) = S_0(t)^{\exp(X\beta)} \quad (22)$$

também é útil quando se deseja concluir a análise, em termos de percentis, associados a grupos de indivíduos.

Se  $h_0(t)$  fosse estimada parametricamente, poderia ser estimada por verossimilhança, entretanto, na verossimilhança parcial o argumento condicional elimina completamente essa função, da verossimilhança. Assim, os estimadores, para essas quantidades, são de natureza não-paramétrica.

Desse modo, a estimativa simples, proposta por Breslow, *apud* Colosimo (2001), para  $H_0(t)$ , é uma função escada com saltos nos tempos distintos de falhas dada por:

$$\hat{H}_0(t_i) = \sum_{j \in R_j} \frac{\delta_i}{\sum_{l \in R_j} \exp(X_l' \hat{\beta})} \quad (23)$$

Então as funções  $S_0(t)$  e  $S(t)$  podem ser estimadas a partir de (21) e (22), respectivamente.

#### 2.5.4.5 Técnicas para verificar a proporcionalidade de riscos

O modelo de regressão de Cox é bastante flexível devido à presença do componente não-paramétrico. No entanto, existe uma suposição básica a ser testada, que é a de riscos proporcionais, sendo que a violação dessa suposição pode acarretar sérios vícios na estimação dos coeficientes do modelo (Colosimo, 2001).

Serão apresentadas, agora, duas técnicas para a verificação da suposição de riscos proporcionais do modelo de regressão de Cox.

- Técnica gráfica

Inicialmente dividi-se os dados em  $m$  faixas, usualmente de acordo com alguma covariável. Em seguida, estima-se  $\hat{H}_0(t)$  para cada faixa, utilizando-se a expressão (23), e constrói-se o gráfico  $\hat{H}_0(t)$  versus  $t$ ,  $j = 1, \dots, m$ . Se a suposição de riscos proporcionais for válida, estas curvas devem ser aproximadamente paralelas no tempo, enquanto que curvas não paralelas indicam desvios da suposição.

- Teste da covariável dependente do tempo

Cox, *apud* Colosimo (2001), propôs um teste para examinar a suposição de riscos proporcionais. O teste consiste em acrescentar ao modelo uma covariável dependente do tempo. O interesse é verificar se a razão das taxas de falhas é a mesma em qualquer tempo  $t$ . O modelo de Cox, para esta situação, é

$$h(t) = h_0(t)\exp(X_1\beta_1)$$

onde  $X_1$  é uma covariável. O risco relativo de falha, em qualquer tempo  $t$  da covariável  $X_1$  com relação ao outro, é  $\exp(\beta_1)$ , se o modelo for adequado para os dados.

Uma outra covariável,  $X_2 = t$ , pode ser adicionada ao modelo

$$h(t) = h_0(t)\exp(X_1\beta_1 + t\beta_2)$$

de modo que o risco relativo é, agora,  $\exp(\beta_1 + t\beta_2)$  e, portanto, não é mais constante no tempo, e nem o modelo é mais de riscos proporcionais. Em particular, se  $\beta_2 < 0$ , o risco relativo decresce com o tempo. Se  $\beta_2 > 0$ , o risco relativo de falha aumenta com o tempo. No caso em que  $\beta_2 = 0$ , o risco relativo é constante e igual a  $\exp(\beta_1)$ . A regressão não é significativa, mostrando que essa hipótese corresponde à suposição de riscos proporcionais (Colosimo, 2001).

## 2.6 Comparação de curvas de sobrevida

Para a comparação entre duas curvas de sobrevida, ou mais, pode-se utilizar o teste Log-rank, ou um teste aproximado, que é o teste clássico do qui-quadrado.

O teste Log-rank é bastante utilizado em análise de sobrevida, e particularmente apropriado quando a razão das funções de risco, dos grupos a serem comparados é, aproximadamente, constante. A estatística desse teste é a diferença entre o número observado de falhas em cada grupo e o correspondente número esperado de falhas sob a hipótese nula, com as seguintes hipóteses:

$H_0$  = as curvas não apresentam diferença significativa;

$H_1$  = as curvas apresentam diferença significativa.

Sejam duas funções de sobrevivência  $S_1(t)$  e  $S_2(t)$  e  $t_1 < t_2 < \dots < t_k$  os tempos de falha distintos na amostra formada pela combinação das duas amostras individuais. Suponha que no tempo  $t_j$  aconteçam  $d_j$  falhas e  $n_j$  indivíduos estejam sob risco em um tempo imediatamente inferior a  $t_j$ . Em cada tempo de falha  $t_j$ , os dados poderão ser dispostos

em uma tabela de contingência 2x2 com  $d_{ij}$  falhas e  $n_{ij} - d_{ij}$  censuras na coluna  $i$ , de acordo com a Tabela 1.

TABELA 1 – Tabela de contingência gerada no tempo  $t_j$ .

|           | Grupos            |                   | Total       |
|-----------|-------------------|-------------------|-------------|
|           | 1                 | 2                 |             |
| Falha     | $d_{1j}$          | $d_{2j}$          | $d_j$       |
| Não-falha | $n_{1j} - d_{1j}$ | $n_{2j} - d_{2j}$ | $n_j - d_j$ |
| Total     | $n_{1j}$          | $n_{2j}$          | $n_j$       |

Condicionalo à experiência de falha e censura até o tempo  $t_j$  (fixando as marginais de coluna) e ao número de falhas no tempo  $t_j$  (fixando as marginais de linha),  $d_{1j}$  segue a distribuição hipergeométrica

$$P(X = d_{1j}) = \frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}} \quad (26)$$

Sendo que a média e a variância de  $d_{1j}$  são, respectivamente:

$$e_{1j} = \frac{n_{1j} d_j}{n_j} \quad V_j = \sum_j \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Então, a estatística  $(d_{1j} - e_{1j})$  tem média zero e variância  $V_j$ . Se as  $k$  tabelas de contingência forem independentes, um teste aproximado, para a igualdade das duas funções de sobrevivência, pode ser baseado na estatística:

$$T^2 = \frac{\left( \sum_{j=1}^k d_{1j} - \sum_{j=1}^k e_{1j} \right)^2}{\sum_{j=1}^k V_j} \sim \chi_{(1)}^2 \quad (27)$$

que é a estatística do teste Log-rank, tendo distribuição qui-quadrado, com 1 grau de liberdade para grandes amostras (Armitage & Berry, 1994).

O teste Log-rank também pode ser utilizado para comparar três ou mais curvas de sobrevida. A hipótese nula, para esta situação mais geral, é a de que todas as curvas de sobrevida são iguais (Armitage & Berry, 1994).

### 3 METODOLOGIA

Para a aplicação da teoria do método de Kaplan-Meier, e do modelo de regressão de Cox, utilizou-se uma amostra formada por 69 pacientes atendidos no Hospital Universitário de Santa Maria - HUSM. Esses pacientes apresentavam câncer no aparelho digestivo e foram tratados com procedimento cirúrgico, conforme o caso.

O período de admissão dos pacientes, no estudo, foi de outubro de 1991 a abril de 1999, e o seguimento dos mesmos encerrou-se em junho de 1999. O tempo de seguimento máximo, observado, foi de, aproximadamente, 7,5 anos, sendo que o tempo mediano de sobrevivência, desses pacientes, foi de 1 ano. O banco de dados encontra-se no Apêndice.

O momento de entrada do paciente no grupo de estudo foi a data da cirurgia e, como tempo final, foi considerada a data do óbito do paciente, ou a data do término do estudo.

O evento de interesse do estudo foi a morte desses indivíduos por complicações relacionadas à doença. A morte não causada pela doença foi considerada perda do acompanhamento, sendo este indivíduo censurado.

As variáveis avaliadas foram: idade, tipo de cirurgia, estágio da doença, tipo de tratamento cirúrgico e data de óbito ou de censura. Os

tipos de cirurgias do estudo foram: gastrectomia subtotal, gastrectomia total e irressecável. Quanto ao estadiamento, os indivíduos foram classificados conforme os estágios I, II, III ou IV, dependendo do estágio da doença. O tipo de tratamento cirúrgico foi classificado em: curativo ou paliativo.

O trabalho desenvolveu-se nas seguintes etapas: descrição das variáveis; utilização do estimador de Kaplan-Meier, para o cálculo e construção das curvas de sobrevida das variáveis; realização do teste para a comparação de curvas (Log-rank e qui-quadrado); construção da função de sobrevida, considerando-se todas as covariáveis iguais a zero; realização da análise univariada de todas as covariáveis; e, análise múltipla das covariáveis consideradas significativas. Após a escolha do modelo adequado, foi verificada a pressuposição de riscos proporcionais.

Para os cálculos estatísticos foram utilizados os programas computacionais: SPSS 10.0, Statistica 5.0 e SAS 8.0.

## 4 RESULTADOS

Neste capítulo são apresentados os resultados obtidos da análise dos dados da amostra.

### 4.1 Descrição da amostra

Nessa amostra, de 69 pacientes, a idade mínima foi de 23 anos e a máxima de 80 anos. A idade média foi de 60 anos, e o desvio padrão de 12,3 anos.

Dos pacientes estudados, 49 (71%) foram a óbito devido ao câncer no aparelho digestivo, até o final do estudo, sendo que 20 (29%) foram censurados. Esses 20 pacientes foram censurados porque sobreviveram a doença até a data do término do estudo, não retornaram para avaliação, ou morreram por outra causa, que não a doença estudada.

A Tabela 2 mostra a distribuição dos pacientes de acordo com as covariáveis de interesse: tipo de cirurgia (gastrectomia total, gastrectomia subtotal ou irressecável), estágio da doença (I, II, III ou IV) e tipo de tratamento (cirurgia curativa ou cirurgia paliativa).

TABELA 2 - Distribuição dos pacientes

|                       | Número de óbitos (%) | Número de sobreviventes (%) | Número de pacientes (%) |
|-----------------------|----------------------|-----------------------------|-------------------------|
| <b>Cirurgia</b>       |                      |                             |                         |
| Gastrectomia total    | 16 (80,0)            | 4 (20,0)                    | 20 (29,0)               |
| Gastrectomia subtotal | 23 (63,9)            | 13 (36,1)                   | 36 (52,2)               |
| Irressecável          | 10 (76,9)            | 3 (23,1)                    | 13 (18,8)               |
| Total                 | 49 (71,0)            | 20 (29,0)                   | 69 (100,0)              |
| <b>Estágio</b>        |                      |                             |                         |
| I                     | 2 (18,2)             | 9 (81,8)                    | 11 (15,9)               |
| II                    | 4 (57,1)             | 3 (42,9)                    | 7 (10,1)                |
| III                   | 20 (83,3)            | 4 (16,7)                    | 24 (34,9)               |
| IV                    | 13 (100,0)           | 0 (0,0)                     | 13 (18,8)               |
| Sem classificação     | 10 (71,4)            | 4 (28,6)                    | 14 (20,3)               |
| Total                 | 49 (71,0)            | 20 (29,0)                   | 69 (100,0)              |
| <b>Tratamento</b>     |                      |                             |                         |
| Curativo              | 26 (61,9)            | 16 (38,1)                   | 42 (60,9)               |
| Paliativo             | 13 (100,0)           | 0 (0,0)                     | 13 (18,8)               |
| Sem classificação     | 10 (71,4)            | 4 (28,6)                    | 14 (20,3)               |
| Total                 | 49 (71,0)            | 20 (29,0)                   | 69 (100,0)              |

Curativo = I+ II+III; Paliativo = IV

#### 4.2 Análise de sobrevida utilizando o estimador de Kaplan-Meier

A função de sobrevida acumulada foi estimada pelo método de Kaplan-Meier, e a curva de sobrevida dos 69 pacientes pode ser observada na Figura 2.

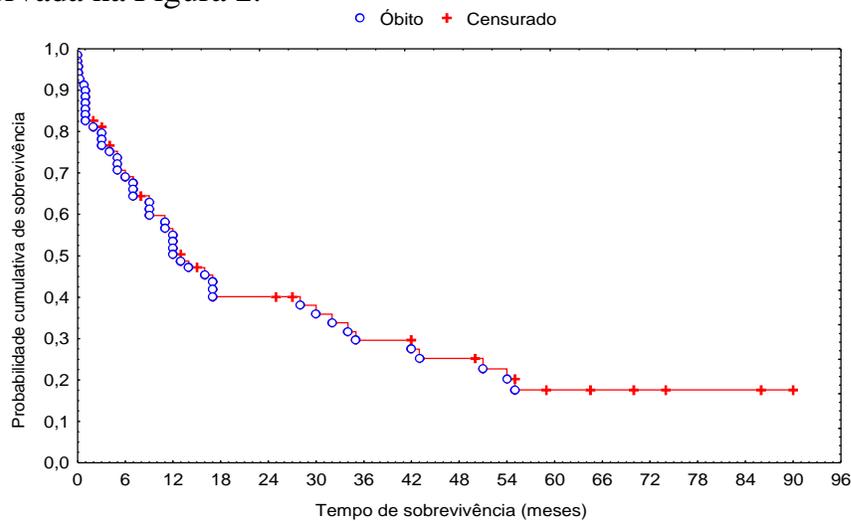


FIGURA 2 - Curva de sobrevida global

A Figura 2 mostra que, dos 69 pacientes estudados, a curva de sobrevida estimada decai para 17,6% a partir do tempo de sobrevivência de 55 meses, permanecendo esse mesmo valor até 90 meses (7,5 anos) pois, nesse período, não ocorreu mais nenhum óbito.

O tempo mediano de sobrevivência foi de 12 meses. Isso significa que 50% dos pacientes sobreviveu até esse tempo, sendo que os outros 50%, sobreviveram mais de 12 meses (Chiesa, 2002).

#### 4.2.1 Análise de sobrevida por tipo de cirurgia

A Figura 3 mostra as curvas de sobrevida para os tipos de cirurgia: gastrectomia subtotal, gastrectomia total e irreseccável.

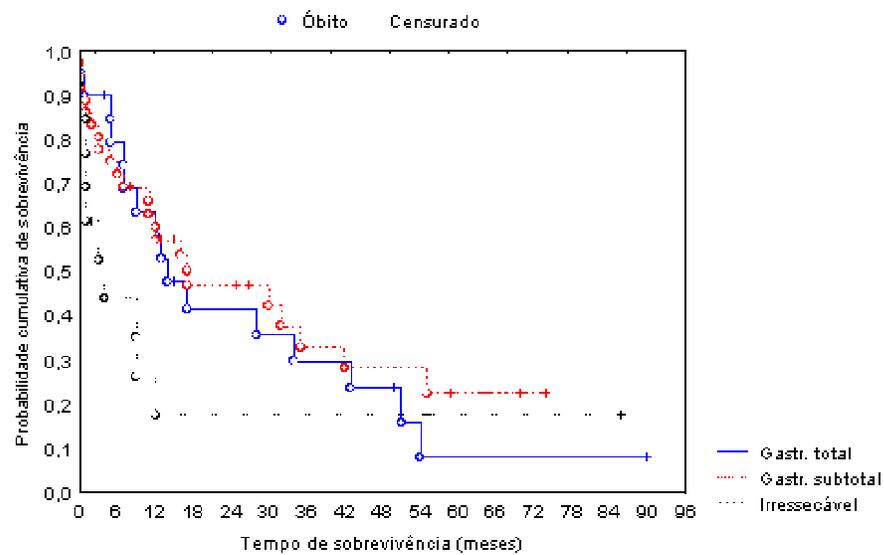


FIGURA 3 - Curvas de sobrevida por tipo de cirurgia

Os pacientes submetidos a gastrectomia subtotal apresentam a probabilidade de sobreviver mais de 12 meses, igual a 57,1%, com tempo mediano de sobrevivência de 17 meses.

Observa-se que a probabilidade de um indivíduo, que foi submetido a gastrectomia total, sobreviver até 12 meses é de 58,2%, e o tempo mediano de sobrevivência é de, aproximadamente, 13 meses.

Os pacientes submetidos à cirurgia irresssecável têm 17,6% de sobrevida em 12 meses, e apresentaram um tempo mediano de sobrevivência inferior aos outros dois grupos. Nesse grupo, 50% dos indivíduos sobreviveram somente até 3 meses, enquanto que, nos outros tipos de cirurgia, o tempo mediano foi maior.

O resultado do teste do qui-quadrado, para a comparação das curvas, não foi significativo, pois  $\chi_c^2=4,74$ , para 2 gl e  $p=0,0935$ , concluindo-se que não existe diferença significativa entre as curvas de sobrevivência, por tipo de cirurgia (Chiesa, 2002).

#### 4.2.2 Análise de sobrevida por estágio da doença

A Figura 4 mostra as curvas de sobrevida para os pacientes por estágio da doença.

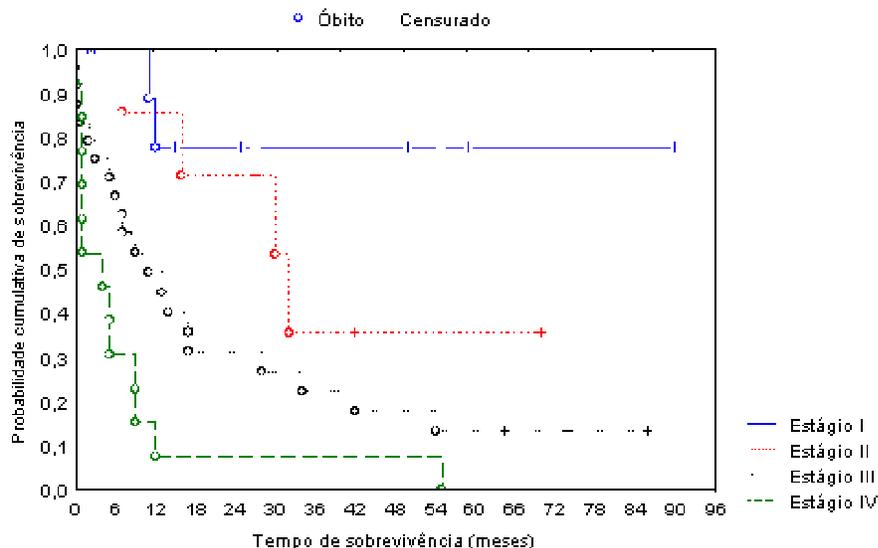


FIGURA 4 - Curvas de sobrevida por estágio da doença

Observa-se que, para os pacientes classificados no estágio I da doença, a probabilidade de óbito é pequena, ou seja, é grande a probabilidade de sobrevivência, sendo a probabilidade de sobrevida em 1 ano, de 77,8%. No estágio II, a probabilidade de sobrevida em 16 meses é de 71,4%. Já, no estágio III, a probabilidade de sobrevida em 13 meses é de 44,9%. E, os pacientes classificados no estágio IV, a probabilidade de sobrevida em 12 meses é pequena, apenas 7,7%. Esse resultado já era esperado, pois esse é o estágio mais avançado da doença.

O resultado do teste do qui-quadrado, para a comparação das curvas por estágio da doença, foi significativo, pois  $\chi^2=19,03$ , para 3 gl e  $p=0,0003$ , concluindo-se que existe diferença significativa entre as curvas de sobrevivência, por estágio da doença.

Utilizou-se o teste Log-rank para identificar quais as curvas que diferiam entre si, mostrando que a curva de sobrevida do estágio I difere da curva do estágio III ( $p=0,0029$ ) e da curva do estágio IV ( $p=0,0007$ ) e, a curva do estágio II, difere da curva do estágio IV ( $p=0,0035$ ), sendo as demais comparações, não-significativas, ao nível de 5% (Chiesa, 2002).

#### 4.2.3 Análise de sobrevida por tipo de tratamento cirúrgico

A Figura 5 mostra as curvas de sobrevida por tipo de cirurgia. Para o grupo de pacientes, que realizaram tratamento cirúrgico curativo, a probabilidade de sobrevida em 12 meses foi de 62,6%. Os pacientes que realizaram tratamento cirúrgico paliativo tiveram probabilidade de sobrevida de 7,7%, em 12 meses.

O teste qui-quadrado, utilizado para a comparação das curvas, mostrou resultado significativo, pois  $\chi^2=13,50$ , para 1 gl e  $p=0,00024$ , concluindo-se que existe diferença significativa entre as curvas de sobrevivência por tipo de tratamento cirúrgico (Chiesa, 2002).

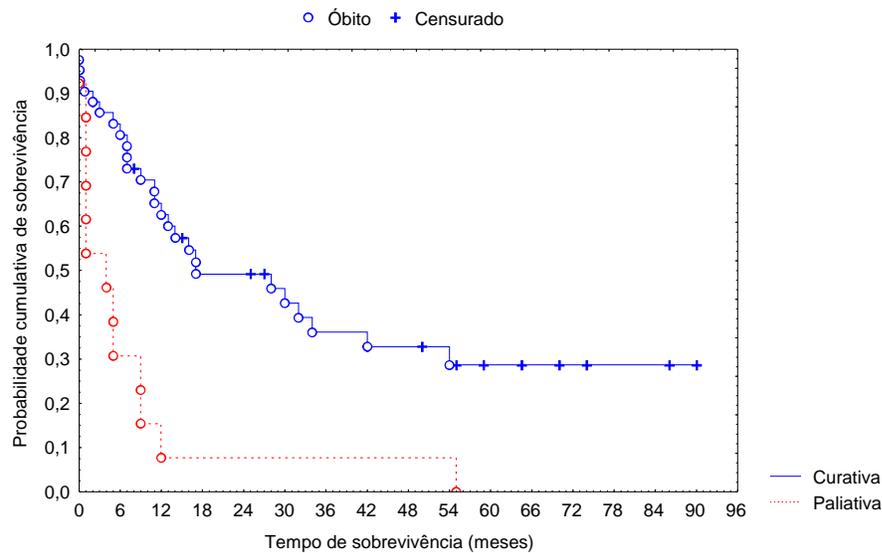


FIGURA 5 - Curvas de sobrevida por tipo de tratamento

Essa avaliação preliminar indica que as variáveis: tipo de cirurgia e estágio da doença, por terem apresentado diferença significativa entre as suas curvas, são importantes na estimação da função de sobrevida, e que, possivelmente, serão importantes para a construção do modelo de regressão de Cox, que será realizado a seguir.

#### 4.3 Modelo de regressão de Cox

Os modelos avaliados, a seguir, serão numerados apenas para efeito didático, não havendo nenhum grau de importância ou prioridade na seqüência escolhida. Optou-se por realizar uma análise

univariada de todas as covariáveis, utilizando-se um grau de significância de 20%, sugerido por Hosmer & Lemeshow (1999).

#### 4.3.1 Modelo 1 - Com a covariável cirurgia

Cirurgia é uma covariável categórica, e foi subdividida em três categorias, de acordo com o tipo de cirurgia a que foram submetidos os pacientes. Os tipos de cirurgia são: gastrectomia subtotal, gastrectomia total e irressecável.

TABELA 3 - Distribuição por cirurgia

| Dados avaliados | Evento | Censura | Tempo=0 | Missings | Total |
|-----------------|--------|---------|---------|----------|-------|
| Nº de pacientes | 47     | 20      | 2       | 0        | 69    |

Na Tabela 3 são apresentados os resultados em relação à ocorrência de tempo zero, censura e dados faltantes (*missings*).

TABELA 4 - Modelo 1 de regressão de Cox - covariável cirurgia

| Cirurgia                  | $\beta$ | SE   | $W^2$ | gl | p     | $\exp\beta$ |
|---------------------------|---------|------|-------|----|-------|-------------|
| Gastrectomia subtotal     |         |      | 2,010 | 2  | 0,366 |             |
| Gastrectomia total        | 0,234   | 0,33 | 0,504 | 1  | 0,478 | 1,264       |
| Gastrectomia irressecável | 0,559   | 0,40 | 1,951 | 1  | 0,163 | 1,749       |

sendo:

$\beta$  = coeficiente estimado;

SE = erro padrão do coeficiente;

$W^2$  = estatística de Wald (como foi definida na equação 17);

gl = número de graus de liberdade;

p = probabilidade ou p-valor;

$\exp\beta$  = HR = razão de riscos.

Os resultados da Tabela 4 mostram, como referência, a primeira categoria dos pacientes que foram submetidos à gastrectomia subtotal, não calculando, portanto, o valor do parâmetro  $\beta$ . Os pacientes localizados na segunda faixa, isto é, os que realizaram gastrectomia total, possuem uma razão de riscos (HR) de óbito 1,264 maior do que a dos pacientes que se submeteram à gastrectomia subtotal. Os pacientes submetidos à cirurgia irressecável têm uma razão de risco de óbito 1,749 maior do que os pacientes que se submeteram à gastrectomia subtotal.

Pelo teste de Wald, essa covariável é significativa, pois, para a cirurgia irressecável obteve-se  $p=0,163$ , e, segundo Hosmer & Lemeshow (1999), nessa fase preliminar da análise, deve-se utilizar o nível de significância de 20%.

A Tabela 5 irá mostrar os resultados do teste da razão da verossimilhança para o modelo 1.

TABELA 5 - Teste da razão de verossimilhança do modelo 1

| $-2 \ln L(0)$ | $-2 \ln (\beta)$ | $\chi^2$ | gl | p     |
|---------------|------------------|----------|----|-------|
| 334,093       | 332,185          | 1,907    | 2  | 0,385 |

O teste da razão de verossimilhança, mostrado na Tabela 5, indica que a inclusão da covariável cirurgia no Modelo 1 não é significativa, pois  $p=0,385$ .

#### 4.3.2 Modelo 2 - Com a covariável estágio

Estágio é uma covariável dividida em quatro categorias: I, II, III e IV, de acordo com a fase em que se encontra a doença.

A seguir, apresenta-se a distribuição dessa covariável na amostra.

TABELA 6 - Distribuição por estágio da doença

| Dados avaliados | Evento | Censura | Tempo=0 | Missings | Total |
|-----------------|--------|---------|---------|----------|-------|
| Nº de pacientes | 37     | 16      | 2       | 14       | 69    |

Na Tabela 6 são apresentados os resultados em relação à ocorrência de tempo zero, censura e dados faltantes (*missings*).

TABELA 7 - Modelo 2 de regressão de Cox - covariável estágio

| Estágio | $\beta$ | SE    | $W^2$  | gl | p     | $\exp\beta$ |
|---------|---------|-------|--------|----|-------|-------------|
| I       |         |       | 15,465 | 3  | 0,001 |             |
| II      | 1,036   | 0,867 | 1,428  | 1  | 0,232 | 2,818       |
| III     | 1,790   | 0,744 | 5,788  | 1  | 0,016 | 5,987       |
| IV      | 2,600   | 0,771 | 11,366 | 1  | 0,001 | 13,463      |

A Tabela 7 mostra os resultados do modelo de Cox, com a inclusão da covariável estágio da doença.

A inclusão da covariável estágio, no modelo 2, é significativa pelo teste de Wald, pois no estágio III  $p=0,016$ ; e no estágio IV  $p=0,001$ , sendo o resultado global  $p=0,001$ .

TABELA 8 - Teste da razão de verossimilhança do modelo 2

| $-2 \ln L(0)$ | $-2 \ln(\beta)$ | $\chi^2$ | gl | p     |
|---------------|-----------------|----------|----|-------|
| 249,739       | 230,092         | 19,647   | 3  | 0,000 |

Pelo teste da razão de verossimilhança ( $p=0,000$ ), obtém-se a mesma informação.

#### 4.3.3 Modelo 3 - Com a covariável idade

Na Tabela 9 tem-se os resultados da análise da covariável contínua, idade.

TABELA 9 - Distribuição por idade

| Dados avaliados | Evento | Censura | Tempo=0 | Missings | Total |
|-----------------|--------|---------|---------|----------|-------|
| Nº de pacientes | 33     | 11      | 1       | 24       | 69    |

Observa-se a ocorrência de 24 dados faltantes, e 1 tempo de sobrevida igual a zero.

TABELA 10 - Modelo 3 de regressão de Cox - covariável idade

|       | $\beta$ | SE    | $W^2$ | gl | p     | $\exp\beta$ |
|-------|---------|-------|-------|----|-------|-------------|
| Idade | -0,011  | 0,017 | 0,450 | 1  | 0,502 | 0,989       |

Pelo teste de Wald, a covariável idade não é significativa no modelo de Cox, pois  $p=0,502$ .

Na Tabela 11 apresentam-se os resultados do teste da razão de verossimilhança para o modelo de Cox, com a covariável idade.

TABELA 11 - Teste da razão de verossimilhança do modelo 3

| $-2 \ln L(0)$ | $-2 \ln (\beta)$ | $\chi^2$ | gl | p     |
|---------------|------------------|----------|----|-------|
| 206,627       | 206,183          | 0,444    | 1  | 0,505 |

A inclusão da covariável idade, no modelo 3, não é significativa pelo teste da razão da verossimilhança ( $p=0,505$ ), confirmando o teste de Wald.

#### 4.3.4 Modelo 4 - Com a covariável tratamento

A covariável tratamento é categórica, sendo que os tipos de tratamentos cirúrgicos utilizados foram: cirurgia paliativa, ou curativa. Os pacientes submetidos à cirurgia curativa são os mesmos classificados nos estágios I, II e III da doença, enquanto que os

pacientes submetidos à cirurgia paliativa são aqueles classificados no estágio IV. A seguir, apresenta-se a distribuição dessa covariável na amostra.

TABELA 12 - Distribuição por tratamento

| Dados avaliados | Evento | Censura | Tempo=0 | Missings | Total |
|-----------------|--------|---------|---------|----------|-------|
| Nº de pacientes | 37     | 16      | 2       | 14       | 69    |

Na Tabela 12 observa-se 14 dados faltantes e 2 tempos de sobrevida iguais a zero.

TABELA 13 - Modelo 4 de regressão de Cox - covariável tratamento

| Tratamento | $\beta$ | SE    | $W^2$  | gl | p     | $\exp\beta$ |
|------------|---------|-------|--------|----|-------|-------------|
| Curativo   |         |       |        |    |       |             |
| Paliativo  | 1,254   | 0,366 | 11,772 | 1  | 0,001 | 3,505       |

A Tabela 13 mostra os resultados do modelo de Cox para a covariável tratamento. O tratamento curativo não tem o valor do parâmetro calculado, pois esse tratamento é considerado como referência. Segundo esse modelo, a razão de risco (HR) de óbito de um paciente, submetido a um tratamento paliativo, seria 3,505 vezes àquela de um paciente submetido a um tratamento curativo.

TABELA 14 - Teste da razão de verossimilhança do modelo 4

| $-2 \ln L(0)$ | $-2 \ln(\beta)$ | $\chi^2$ | gl | p     |
|---------------|-----------------|----------|----|-------|
| 249,739       | 239,710         | 10,029   | 1  | 0,002 |

A inclusão da covariável tratamento no modelo é significativa, pelo teste de Wald ( $p=0,001$ ) e pelo teste da razão de verossimilhança ( $p=0,002$ ), mostrado nas Tabelas 13 e 14, respectivamente.

#### 4.3.5 Modelo 5 - Com todas as covariáveis significativas

A análise univariada mostrou que apenas a covariável idade não foi significativa, indicando que ela não contribui, estatisticamente, para o modelo. Assim, considerou-se o modelo múltiplo com as covariáveis significativas.

TABELA 15 – Distribuição por cirurgia, estágio e tratamento

| Dados avaliados | Evento | Censura | Tempo=0 | Missings | Total |
|-----------------|--------|---------|---------|----------|-------|
| Nº de pacientes | 37     | 16      | 2       | 14       | 69    |

A Tabela 15 mostra que os dados, para a análise dessas covariáveis, possuem 14 dados faltantes e 2 tempos iguais a zero. Na Tabela 16 tem-se os resultados do modelo de Cox para as covariáveis cirurgia, estágio e tratamento.

TABELA 16 - Modelo 5 de regressão de Cox - análise múltipla das covariáveis

| Covariável            | $\beta$ | SE    | $W^2$  | gl | p     | $\exp\beta$ |
|-----------------------|---------|-------|--------|----|-------|-------------|
| Gastrectomia subtotal |         |       | 0,507  | 2  | 0,776 |             |
| Gastrectomia total    | 0,303   | 0,434 | 0,487  | 1  | 0,485 | 1,354       |
| Irressecável          | 0,036   | 0,485 | 0,006  | 1  | 0,941 | 1,037       |
| Estágio I             |         |       | 14,332 | 3  | 0,002 |             |
| Estágio II            | 1,165   | 0,896 | 1,689  | 1  | 0,194 | 3,205       |
| Estágio III           | 1,763   | 0,746 | 5,589  | 1  | 0,018 | 5,829       |
| Estágio IV            | 2,695   | 0,795 | 11,502 | 1  | 0,001 | 14,804      |
| Tratamento curativo   |         |       |        |    |       |             |
| Tratamento paliativo  |         |       |        | *  |       |             |

\* O programa SPSS 10.0 indica que não é possível realizar a análise da covariável tratamento, juntamente com a covariável estágio, pois elas são covariáveis linearmente dependentes.

Na Tabela 17 são mostrados os resultados do teste da razão de verossimilhança para o modelo 5.

TABELA 17 - Teste da razão de verossimilhança do modelo 5

| $-2 \ln L(0)$ | $-2 \ln (\beta)$ | $\chi^2$ | gl | p     |
|---------------|------------------|----------|----|-------|
| 249,739       | 229,588          | 20,151   | 5  | 0,001 |

Os resultados mostram que a covariável tratamento deve ser retirada, pois ela não contribui no modelo, quando colocada junto com a covariável estágio, uma vez que existe uma relação linear entre elas.

Observa-se que a covariável cirurgia, quando modelada individualmente (modelo 1), é significativa ( $p=0,163$ ), mas quando considerada em conjunto, com as outras covariáveis (modelo 5), mostra-se não-significativa. Isso será investigado, novamente, eliminando-se a covariável tratamento, no próximo passo.

#### 4.3.6 Modelo 6 - Com as covariáveis cirurgia e estágio

A seguir, a Tabela 18 mostra a distribuição dos dados para as covariáveis cirurgia e estágio da doença.

TABELA 18 - Distribuição por cirurgia e estágio

| Dados avaliados | Evento | Censura | Tempo=0 | <i>Missings</i> | Total |
|-----------------|--------|---------|---------|-----------------|-------|
| Nº de pacientes | 37     | 16      | 2       | 14              | 69    |

Observa-se que, para essas covariáveis, existem 14 dados faltantes, 2 tempos iguais a zero e 16 dados censurados .

Na Tabela 19, a seguir, são mostrados os resultados obtidos do modelo de regressão de Cox, para a análise múltipla das covariáveis cirurgia e estágio.

TABELA 19 - Modelo 6 de regressão de Cox - covariáveis cirurgia e estágio

| Covariável            | $\beta$ | SE    | $W^2$  | gl | p     | $\exp\beta$ |
|-----------------------|---------|-------|--------|----|-------|-------------|
| Gastrectomia subtotal |         |       | 0,507  | 2  | 0,776 |             |
| Gastrectomia total    | 0,303   | 0,434 | 0,487  | 1  | 0,485 | 1,354       |
| Irressecável          | 0,036   | 0,485 | 0,006  | 1  | 0,941 | 1,037       |
| Estágio I             |         |       | 14,332 | 3  | 0,002 |             |
| Estágio II            | 1,165   | 0,896 | 1,689  | 1  | 0,194 | 3,205       |
| Estágio III           | 1,763   | 0,746 | 5,589  | 1  | 0,018 | 5,829       |
| Estágio IV            | 2,695   | 0,795 | 11,502 | 1  | 0,001 | 14,804      |

Observa-se que a inclusão da covariável cirurgia, no modelo 6, não é significativa ( $p > 0,05$ ), portanto será retirada.

TABELA 20 - Teste da razão de verossimilhança do modelo 6

| $-2 \ln L(0)$ | $-2 \ln (\beta)$ | $\chi^2$ | gl | p     |
|---------------|------------------|----------|----|-------|
| 249,739       | 229,588          | 20,151   | 5  | 0,001 |

Na Tabela 20 tem-se os resultados do teste da razão de verossimilhança, o qual mostra-se significativo para o modelo 6. Porém, pelo teste de Wald, a covariável cirurgia não é significativa, indicando que ela não contribui para o modelo.

#### 4.3.7 Modelo final

De acordo com as análises obtidas anteriormente, conclui-se que o modelo que melhor se ajusta aos dados é o modelo 2, isto é, a análise múltipla aponta a covariável estágio como sendo a que oferece a melhor contribuição para a modelagem dos dados em estudo. A Tabela 21 mostra o modelo final, a um nível de significância de 5%.

TABELA 21 - Modelo de regressão de Cox final, com a covariável estágio

| Estágio       | $\beta$ | SE    | $W^2$  | gl | p     | $\exp\beta$ | IC 95%        |
|---------------|---------|-------|--------|----|-------|-------------|---------------|
| I ( $X_1$ )   |         |       | 15,465 | 3  | 0,001 |             |               |
| II ( $X_2$ )  | 1,036   | 0,867 | 1,428  | 1  | 0,232 | 2,818       | 0,515; 15,418 |
| III ( $X_3$ ) | 1,790   | 0,744 | 5,788  | 1  | 0,016 | 5,987       | 1,393; 25,732 |
| IV ( $X_4$ )  | 2,600   | 0,771 | 11,366 | 1  | 0,001 | 13,463      | 2,970; 61,038 |

Nesta etapa da análise é necessário testar se os riscos de óbito do modelo encontrado são proporcionais.

#### 4.3.8 Teste gráfico para a proporcionalidade

Na Figura 6 são apresentadas as curvas, por categoria, da covariável estágio. Devido ao pequeno tamanho da amostra, a conclusão gráfica, sobre a proporcionalidade dos riscos relativos, fica prejudicada, mas pode-se dizer que os riscos são, aproximadamente, proporcionais pela análise da Figura 6.

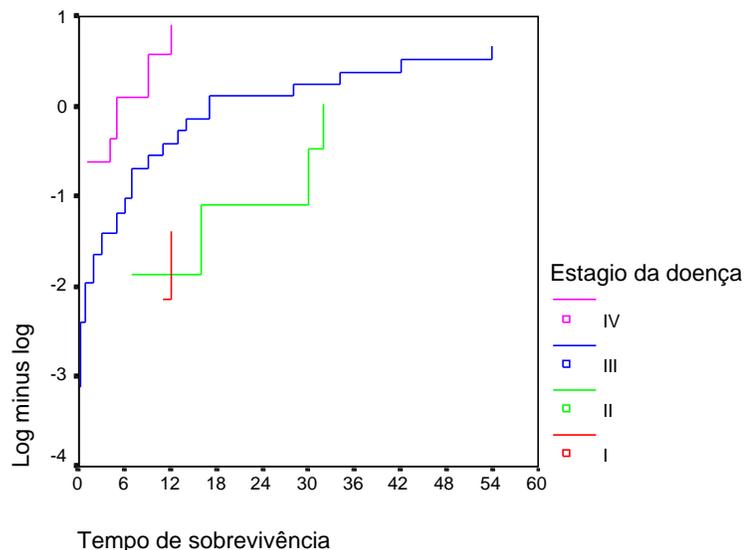


FIGURA 6 - Curvas comparativas por estágio da doença

No entanto, pode-se proceder ao teste de dependência no tempo para verificar a pressuposição de riscos proporcionais com maior segurança.

#### 4.3.9 Teste de covariável dependente do tempo.

Para realizar o teste de proporcionalidade foi criada uma nova covariável, dependente do tempo definida como:

$$\text{deptemp} = X (\ln t - 2,17),$$

sendo:

X = covariável estágio;

ln t = logaritmo neperiano do tempo de sobrevivida;

2,17 = média do ln t.

Os resultados são mostrados a seguir:

TABELA 22 - Modelo com a covariável dependente do tempo

| Covariável | $\beta$ | SE    | $W^2$ | gl | p     | $\exp\beta$ |
|------------|---------|-------|-------|----|-------|-------------|
| Estágio    | 0,787   | 0,196 | 16,10 | 1  | 0,000 | 2,198       |
| deptemp    | -0,075  | 0,124 | 0,366 | 1  | 0,545 | 0,928       |

A Tabela 22 mostra que a variável deptemp não é significativa (p=0,545), não contribuindo para o modelo. Os riscos se mantêm constantes no tempo, atendendo à pressuposição da metodologia da regressão de Cox, com  $\beta_2 = -0,075$ , ou seja,  $\beta_2 \cong 0$ , indicando que o modelo é um modelo de riscos proporcionais.

Assim, pode-se afirmar que, entre os modelos avaliados, o modelo 2 é o de melhor ajuste dos dados. A função de sobrevivida é:

$$\hat{S}(t) = \hat{S}_0(t) \exp(1,036X_2 + 1,790X_3 + 2,600X_4)$$

e a função razão de riscos relativos do modelo é:

$$HR = \exp(1,036X_2 + 1,790X_3 + 2,600X_4)$$

A razão de riscos relativos mostra o estágio I ( $X_1$ ) como referência. Os pacientes classificados no segundo estágio ( $X_2$ ) têm 2,8 vezes o risco de óbito, em relação aos pacientes do estágio I. Os pacientes do estágio III ( $X_3$ ) têm 6,0 vezes o risco de óbito em relação aos pacientes do estágio I. E, os pacientes do estágio IV ( $X_4$ ), tem 13,5 vezes o risco de óbito em relação aos pacientes que se encontram no estágio I da doença, conforme os valores de  $\exp\beta$  da Tabela 21.

Na Figura 7 tem-se as curvas de sobrevida.

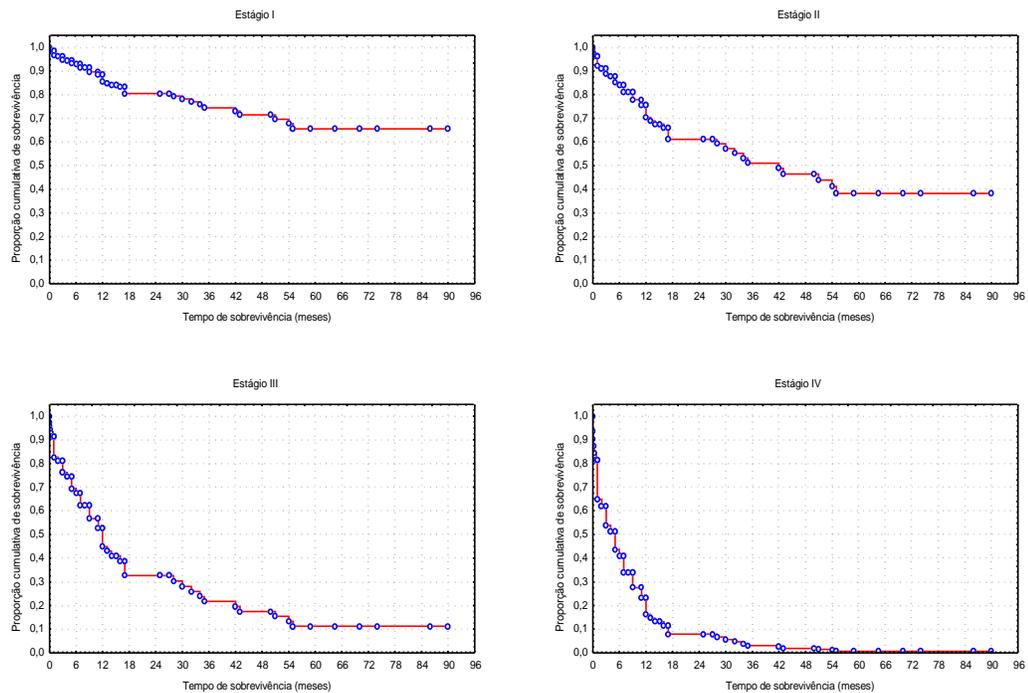


FIGURA 7 - Gráficos da função de sobrevida por estágio da doença

A Figura 7 mostra como se comportam as curvas de sobrevida nos quatro estágios, podendo-se observar uma queda mais rápida da sobrevida em função do maior grau do estágio da doença do paciente.

Observando-se as Tabelas 3, 6, 9, 12, 15 e 18 destaca-se em cada modelo, a quantidade de dados faltantes (*missings*) e a quantidade de dados onde o tempo é considerado zero, isto é, o paciente entrou no grupo e morreu imediatamente.

Esses dados são tratados de forma diferenciada, de acordo com o programa computacional utilizado. Conseqüentemente, os valores dos logaritmos das verossimilhanças, para os três softwares utilizados, foram diferentes, não afetando, entretanto, as conclusões finais.

Neste trabalho, para o cálculo dos logaritmos das verossimilhanças, foi utilizado o programa SPSS, o qual não leva em conta os tempos iguais a zero, nem os dados faltantes.

No caso do modelo 2, apresentado na Tabela 6, por exemplo, o programa considera, para os cálculos, apenas 53 observações, e não as dos 69 pacientes do banco de dados, já que, para essa covariável, existem 2 tempos iguais a zero e 14 dados faltantes.

Fato semelhante aconteceu nos outros modelos, onde ocorreram dados faltantes.

Para o teste de covariável dependente do tempo, e para a construção dos gráficos, utilizou-se o programa Statistica.

Embora os cálculos tenham sido realizados nos três softwares citados, optou-se pela análise do SPSS, por conveniência no manuseio do mesmo.

## 5 CONCLUSÃO

A análise de sobrevida, realizada pelo estimador de Kaplan-Meier, levou às seguintes conclusões:

A covariável tipo de cirurgia não apresentou diferença significativa ( $p > 0,05$ ) na sobrevida dos pacientes, quanto aos diferentes procedimentos cirúrgicos a que estes foram submetidos: gastrectomia subtotal, gastrectomia total e irresseccável.

Quanto à covariável estágio da doença, houve diferença significativa ( $p < 0,05$ ) em relação a sobrevida dos pacientes, sendo que, quanto mais avançado o estágio da doença menor a probabilidade de sobrevivência, ou seja, maior o risco de óbito do paciente. Observou-se, também, que a diferença significativa ocorreu entre os estágios I e III, I e IV e, também, entre os estágios II e IV, sendo o estágio IV o mais avançado da doença, e onde todos os pacientes foram à óbito até o final do estudo.

Quanto ao tipo de tratamento cirúrgico, curativo ou paliativo, sabe-se que há uma associação com o estágio da doença, pois os pacientes que estavam no estágio I, II e III receberam o tratamento curativo, já os pacientes que estavam no estágio IV receberam o tratamento paliativo, por estarem em uma fase muito avançada da doença. Na comparação das curvas, como já era esperada, pela relação

entre estas variáveis, a sobrevida dos pacientes foi, significativamente menor ( $p < 0,05$ ) no grupo que recebeu o tratamento curativo.

Quanto a modelagem dos dados, utilizando-se o modelo de regressão de Cox. A covariável estágio foi a única a apresentar resultado significativo no modelo ( $p = 0,001$ ), sendo, esta, uma covariável categórica, ou seja, avaliada em faixas de risco. Pode-se concluir que: quanto maior o estágio da doença, maior o risco de óbito, sendo este resultado bem claro, tanto na modelagem de Cox, quanto na estimação por Kaplan-Meier, ou seja, há uma clara concordância entre as duas análises, o que era esperado.

Embora algumas covariáveis apresentem diferenças significativas nas curvas de sobrevida pelo Kaplan-Meier, como é o caso do tipo de tratamento, quando essas variáveis são colocadas juntas, no modelo múltiplo, algumas perdem o poder de estimação devido a contribuições competitivas no modelo. Nesse caso, a relação entre tratamento e estágio era de perfeita dependência linear, não havendo sentido algum em manter as duas no modelo.

A covariável idade não se mostrou significativa em relação ao risco de óbito, possivelmente por tratar-se de um conjunto de pacientes com idade, em geral, mais avançada, cuja média foi de 60 anos.

Ao desenvolver as análises, utilizando os três programas computacionais (SAS, SPSS, STATISTICA), pode-se observar algumas diferenças nos resultados, em função dos critérios utilizados nos programas quanto a ocorrência de empates, ocorrência de valores de tempo de sobrevivência zero e a ocorrência de *missings*, ou dados

faltantes. Neste estudo, todos esses problemas ocorreram, e, além disso, a amostra era pequena ( $n=69$ ), portanto houve um cuidado especial na análise, principalmente quanto à comparação das funções de verossimilhança, pois não é possível compará-las quando os dois modelos apresentam *missings* diferentes.

Embora, nesse caso, não tenha ficado dúvidas quanto a inclusão de apenas a covariável estágio no modelo, se houvesse a necessidade de realizar testes de comparação entre modelos isso seria dificultado pela ocorrência de dados incompletos, pois estes geram valores distintos do logaritmo da verossimilhança para cada bloco, ou conjunto de dados. Só é possível realizar o teste da razão da verossimilhança, para comparar modelos, se esses modelos possuírem o mesmo valor do logaritmo da verossimilhança, para seu modelo base ou modelo sem covariáveis,  $\ln L(0)$ . Para que isso ocorra, é necessário um banco de dados completo.

Em função do pequeno tamanho da amostra, os intervalos de confiança ficaram grandes, ou seja, com pequena precisão, sendo que o modelo encontrado é apenas para fins didáticos, não sendo útil para produzir estimativas de risco de óbito reais.

Portanto, sugere-se que novos trabalhos sejam desenvolvidos nesta área, utilizando esta metodologia e que este texto possa ser útil como referência.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- ARMITAGE, P. & BERRY, G. **Statistical Methods in Medical Research**. 3 ed. Blackwell Science Ltda. USA, 1994.
- CHIESA, E M. F. **Análise de Sobrevida através de um método não paramétrico**. Santa Maria, 2002. 61 f. Monografia (Trabalho de conclusão apresentado à UFSM para a obtenção do título de Especialista em Estatística e Modelagem Quantitativa) - Departamento de Estatística, Universidade Federal de Santa Maria, 2002.
- COLOSIMO, E. A. **Análise de Sobrevida Aplicada**. REUNIÃO ANUAL DA RBRAS, 46ª, SEAGRO, 9º, Piracicaba, SP. **Anais**. ESALQ/USP, 2001.
- COX, D. R. Regression Models and Life Tables. **Journal of the Royal Statistical Society B**, 34, 187-220. 1972.
- COX, D. R. Partial likelihood. **Biometrika**, 62, 269-276, 1975.
- FISCHER, L.D. & BELLE, G.V. **Biostatistics – A Methodology for the Health Sciences**. New York : John Wiley & Sons, Inc, 1995.
- HOSMER, DAVID W., Jr & LEMESHOW, S. **Applied Survival Analysis - Regression Modeling of Time to Event Data**. New York: John Wiley & Sons, Inc, 1999.
- KAPLAN, E. L. & MEIER, P. Nonparametric estimation for incomplete observations. **Journal of the American Statistical Association**, v. 53, n.282, p.457-481, 1958.
- KLEINBAUM, D. G. **Survival analysis – A self learning text**. New York: Verlag New York, Inc, 1996.
- SOARES, J. F. & BARTMANN, F. C. **Métodos Estatísticos na Pesquisa Clínica**. Rio de Janeiro : IMPA, 1983.

## 7 APÊNDICE

TABELA 23 - Banco de dados

|    | Tempo(meses) | Censura | Cirurgia | Estágio | Idade (anos) | Tratamento |
|----|--------------|---------|----------|---------|--------------|------------|
| 1  | 64,53        | 1       | 2        | 3       | 55           | 1          |
| 2  | 12           | 0       | 1        |         |              |            |
| 3  | 7            | 0       | 1        | 3       | 70           | 1          |
| 4  | 13           | 0       | 1        | 3       |              | 1          |
| 5  | 9            | 0       | 3        | 4       | 37           | 2          |
| 6  | 51           | 0       | 1        |         |              |            |
| 7  | 3            | 0       | 3        |         | 73           |            |
| 8  | -            | 0       | 2        | 3       | 79           | 1          |
| 9  | 1            | 0       | 3        | 4       |              | 2          |
| 10 | 42           | 0       | 2        | 3       | 55           | 1          |
| 11 | 8            | 1       | 2        | 3       |              | 1          |
| 12 | 17           | 0       | 2        |         |              |            |
| 13 | 30           | 0       | 2        | 2       | 59           | 1          |
| 14 | 14           | 0       | 1        | 3       | 75           | 1          |
| 15 | 42           | 1       | 2        | 2       | 55           | 1          |
| 16 | 17           | 0       | 1        | 3       | 60           | 1          |
| 17 | 28           | 0       | 1        | 3       | 69           | 1          |
| 18 | 0,13         | 0       | 2        | 3       | 57           | 1          |
| 19 | 4            | 1       | 1        |         | 23           |            |
| 20 | 15           | 1       | 2        |         |              |            |
| 21 | 0,07         | 0       | 2        | 3       | 71           | 1          |
| 22 | 25           | 1       | 2        | 1       | 40           | 1          |
| 23 | 3            | 0       | 2        |         | 80           |            |
| 24 | 12           | 0       | 2        | 1       | 69           | 1          |
| 25 | 16           | 0       | 2        | 2       | 62           | 1          |
| 26 | 70           | 1       | 2        | 2       | 64           | 1          |
| 27 | 74           | 1       | 2        | 3       |              | 1          |
| 28 | 3            | 0       | 2        | 3       |              | 1          |
| 29 | 0,77         | 0       | 1        | 3       | 61           | 1          |
| 30 | 2            | 1       | 3        | 1       |              | 1          |
| 31 | 1            | 0       | 2        | 4       |              | 2          |
| 32 | 54           | 0       | 1        | 3       | 56           | 1          |
| 33 | 12           | 0       | 3        | 4       | 44           | 2          |
| 34 | 1            | 0       | 3        | 4       |              | 2          |
| 35 | 9            | 0       | 1        | 3       | 39           | 1          |

Continuação ...

|    |      |   |   |   |    |   |
|----|------|---|---|---|----|---|
| 36 | 35   | 0 | 2 |   | 50 |   |
| 37 | 13   | 1 | 2 |   |    |   |
| 38 | 32   | 0 | 2 | 2 | 53 | 1 |
| 39 | 1    | 0 | 3 | 4 | 64 | 2 |
| 40 | 55   | 0 | 2 | 4 | 66 | 2 |
| 41 | 11   | 0 | 2 | 3 | 64 | 1 |
| 42 | 5    | 0 | 2 | 4 | 63 | 2 |
| 43 | 25   | 1 | 2 |   | 62 |   |
| 44 | 90   | 1 | 1 | 1 |    | 1 |
| 45 | 43   | 0 | 1 |   | 57 |   |
| 46 | 59   | 1 | 2 | 1 |    | 1 |
| 47 | 7    | 0 | 1 | 3 | 55 | 1 |
| 48 | 6    | 0 | 2 | 3 | 60 | 1 |
| 49 | 5    | 0 | 1 | 4 |    | 2 |
| 50 | 3    | 1 | 2 | 1 |    | 1 |
| 51 | 27   | 1 | 2 | 2 | 77 | 1 |
| 52 | 5    | 0 | 1 | 3 |    | 1 |
| 53 | 7    | 0 | 2 | 2 |    | 1 |
| 54 | 1    | 0 | 3 | 4 | 61 | 2 |
| 55 | 50   | 1 | 1 | 1 | 74 | 1 |
| 56 | 12   | 0 | 2 |   |    |   |
| 57 | -    | 0 | 3 | 4 |    | 2 |
| 58 | 15   | 1 | 1 | 1 |    | 1 |
| 59 | 11   | 0 | 2 | 1 | 44 | 1 |
| 60 | 1    | 0 | 2 |   |    |   |
| 61 | 55   | 1 | 3 | 1 | 72 | 1 |
| 62 | 0,27 | 0 | 1 |   |    |   |
| 63 | 17   | 0 | 2 | 3 | 56 | 1 |
| 64 | 27   | 1 | 2 | 1 | 74 | 1 |
| 65 | 9    | 0 | 3 | 4 | 53 | 2 |
| 66 | 86   | 1 | 3 | 3 | 62 | 1 |
| 67 | 2    | 0 | 2 | 3 | 39 | 1 |
| 68 | 34   | 0 | 1 | 3 | 74 | 1 |
| 69 | 4    | 0 | 3 | 4 | 62 | 2 |

Censura: 1 = paciente censurado; 0 = paciente com óbito;

Tipo de cirurgia: 1 = Gastrectomia total; 2 = Gastrectomia subtotal; 3 = Irressecável;

Estadiamento: 1 = estágio I; 2 = estágio II; 3 = estágio III; 4 = estágio IV;

Tipo de tratamento: 1 = curativo; 2 = paliativo