


MODELOS DE ANÁLISE DE SOBREVIVÊNCIA PARA  
EXPERIMENTOS DOSE-RESPOSTA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra. SUELY RUIZ GIOLO e aprovada pela Comissão Julgadora.

Campinas, 29 de abril de 1994.

  
Profa. Dra. Círcia Yuko Wada.

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

UNIVERSIDADE ESTADUAL DE CAMPINAS  
DEPARTAMENTO DE ESTATÍSTICA - IMECC

**MODELOS DE ANÁLISE DE SOBREVIVÊNCIA  
PARA EXPERIMENTOS DOSE - RESPOSTA**

Suely Ruiz Giolo / 436

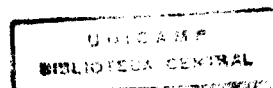
Orientação

Profa. Dra. Cicilia Yuko Wada †

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação da  
Universidade Estadual de Campinas para obtenção do título de Mestre em Estatística.

CAMPINAS - S. P.

1994



Aos meus Pais  
e irmãs .

## AGRADECIMENTOS

Escrever uma dissertação de mestrado é um trabalho árduo que requer a colaboração de muitos. Pela impossibilidade de nomear a todos, gostaria de destacar :

- a Profa. Dra. Cicilia Yuko Wada pela orientação;
- a CAPES pelo apoio financeiro;
- o amigo Damião Nóbrega da Silva pela amizade e ajuda na elaboração dos gráficos;
- os amigos que conquistei, em especial a Dione M. Valença, Silvia E. Shimakura e Ysela D. Agüero;
- os funcionários da biblioteca, da limpeza e do laboratório de estatística da Unicamp, que no cumprimento de seu trabalho me propiciaram condições para a realização desta dissertação.

A todos, meus sinceros agradecimentos.

# SUMÁRIO

INTRODUÇÃO	1
1 - CONCEITOS FUNDAMENTAIS EM ANÁLISE DE SOBREVIVÊNCIA	3
1.1 - INTRODUÇÃO	3
1.2 - CONCEITOS BÁSICOS E NOTAÇÕES	7
1.2.1 - Censuras	8
1.2.2 - Função de densidade de probabilidade	10
1.2.3 - Função de sobrevivência	12
1.2.4 - Função Risco	13
1.2.5 - Relações entre as funções	15
1.3 - MÉTODOS NÃO PARAMÉTRICOS PARA ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA E DA FUNÇÃO RISCO	15
1.3.1 - Método não paramétrico de Kaplan-Meier	15
1.3.2 - Método atuarial ou tabela de vida	16
1.4 - MÉTODO PARAMÉTRICO PARA ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA	17
1.4.1 - Função de verossimilhança	18
1.4.2 - Método iterativo de Newton-Raphson	20
1.5 - VARIÁVEIS CONCOMITANTES	22
2 - MODELOS DE REGRESSÃO EM EXPERIMENTOS DE DOSE-RESPOSTA EM UM CONTEXTO DE UM ÚNICO RISCO	23
2.1 - INTRODUÇÃO	23
2.2 - MODELOS UTILIZANDO OS TEMPOS COMO RESPOSTA	25
2.3 - MODELOS DE REGRESSÃO	30
2.3.1 - Modelo de Locação-escala	31
2.3.2 - Modelo de Riscos Proporcionais	36
2.3.3 - Modelo de Riscos Proporcionais de Cox	40

2.4 - INVESTIGAÇÃO DE UM MODELO	42
2.5 - ANÁLISE DE RESÍDUOS	44
2.5.1 - Caso dos modelos paramétricos	44
2.5.2 - Caso do modelo semi-paramétrico	48
2.6 - ESTIMAÇÃO DOS COEFICIENTES $\beta$ POR MÁXIMA VEROSSIMILHANÇA	51
2.7 - TESTES E INTERVALOS ESTIMADOS PARA OS COEFICIENTES $\beta$	54
2.7.1 - Para os modelos paramétricos	54
2.7.2 - Para o modelo semi-paramétrico	55
3 - MODELOS DE REGRESSÃO EM EXPERIMENTOS DE DOSE-RESPOSTA EM UM CONTEXTO DE RISCOS COMPETITIVOS	56
3.1 - INTRODUÇÃO	56
3.2 - RISCOS COMPETITIVOS - CONCEITOS BÁSICOS E NOTAÇÕES	57
3.2.1 - Abordagem de Prentice para o problema de riscos competitivos	58
3.2.2 - O problema de riscos competitivos usando tempos de falha latentes	60
3.2.3 - O problema da identificabilidade	64
3.3 - MODELOS DE REGRESSÃO EM UM CONTEXTO DE RISCOS COMPETITIVOS	66
3.3.1 - Modelo log-linear	67
3.3.2 - Modelo de Riscos Proporcionais	68
3.3.3 - Modelo de Riscos Proporcionais de Cox	69
3.4 - INVESTIGAÇÃO DE UM MODELO	69
3.5 - ANÁLISE DE RESÍDUOS	70
3.5.1 - Caso dos modelos paramétricos	70
3.5.2 - Caso do modelo semi-paramétrico	72
3.6 - ESTIMAÇÃO DOS ELEMENTOS DE $\beta$ POR MÁXIMA VEROSSIMILHANÇA	73
3.6.1 - Modelos de regressão paramétricos	73
3.6.2 - Modelo semi-paramétrico	77
3.7 - TESTES DE HIPÓTESES	78
4 - APLICAÇÕES NUMÉRICAS	80
4.1 - INTRODUÇÃO	80
4.2 - APLICAÇÃO NUMÉRICA EM UM CONTEXTO DE UM ÚNICO RISCO DE FALHA	82

4.2.1 - Descrição dos dados gerados	82
4.2.2 - Modelos ajustados e estimação	82
4.2.3 - Testes dos parâmetros e coeficientes de regressão estimados	83
4.2.4 - Análise Residual	84
4.2.5 - Funções de Sobrevivência e Risco estimadas, DVS e $t_{mR}$	89
4.3 - APLICAÇÃO NUMÉRICA EM UM CONTEXTO DE RISCOS COMPETITIVOS	94
4.3.1 - Descrição dos dados gerados	94
4.3.2 - Modelos ajustados e estimação	95
4.3.3 - Testes dos parâmetros e coeficientes de regressão estimados	96
4.3.4 - Análise Residual dos modelos ajustados	97
4.3.5 - Funções de Sobrevivência e Risco estimadas, DVS e $t_{mR}$	101
4.3.6 - Testes do efeito da dose no tipo e tempo de falha	104
CONCLUSÕES	106
APÊNDICE A - Programas utilizados nas aplicações numéricas	109
APÊNDICE B - Resultados obtidos nas aplicações numéricas	126
REFERÊNCIAS BIBLIOGRÁFICAS	144

## INTRODUÇÃO

A importância da análise estatística nas áreas aplicadas tem sido amplamente ressaltada e documentada nas últimas décadas.

Particularmente na área de pesquisas biomédicas, a análise de dados de sobrevivência tem merecido grande atenção e reconhecimento. Os dados de sobrevivência são caracterizados, em geral, por uma medida da sobrevivência do paciente, uma resposta a um dado tratamento e por características do paciente e do tratamento que o envolve. Uma dessas medidas frequentemente usada é o "tempo de sobrevivência", o qual refere-se ao tempo observado desde o início de um dado tratamento até a ocorrência de uma resposta a qual seja de interesse. É importante salientar que a resposta não necessariamente é a falha (morte) do paciente. Ela pode ser, por exemplo, a remissão de uma determinada doença, a reação após aplicação de uma droga, dentre outras.

Um aspecto importante a ser destacado nos dados de sobrevivência é o fato de poder ter-se ao final do estudo indivíduos cuja resposta de interesse não ocorreu, indivíduos que antes do término do estudo tiveram de ser removidos do tratamento ou que simplesmente optaram em abandonar o mesmo. Os "tempos exatos" desses indivíduos são portanto desconhecidos e desse modo, registra-se para os mesmos, os períodos em que permaneceram sob observação. Estes denominam-se "tempos censurados".



A presente dissertação tem por objetivo fazer uma revisão de alguns modelos de regressão possíveis de serem utilizados na análise de dados de sobrevivência e o seu uso em experimentos de dose-resposta cuja medida associada a resposta seja o tempo de sobrevivência dos indivíduos; bem como o desenvolvimento de programas computacionais dos métodos quando não disponíveis nos usuais pacotes estatísticos.

Os modelos serão apresentados sob dois contextos: o primeiro deles em situações nas quais apenas uma causa de falha é considerada, e o segundo em situações nas quais  $k$  ( $k > 1$ ) causas de falhas são levadas em consideração. Situações deste tipo são denominadas "riscos competitivos".

No capítulo 1 são apresentados conceitos fundamentais em análise de sobrevivência. O objetivo deste capítulo é o de apresentar alguns conceitos, funções de interesse e notações necessárias ao entendimento da teoria contida nos capítulos posteriores.

No capítulo 2 são apresentados os modelos de regressão locação-escala (em particular o log-linear), de riscos proporcionais e de riscos proporcionais de Cox em um contexto de uma única causa de falha. Para cada um destes modelos é tratado, em detalhes, o ajuste dos mesmos por meio de análise residual gráfica, a estimação por máxima verossimilhança dos parâmetros e coeficientes de regressão envolvidos, testes e intervalos estimados para os referidos coeficientes de regressão. Ainda, para análise de experimentos dose-resposta são estudados o tempo médio estimado para a ocorrência da resposta e o método DVS para obtenção de doses virtualmente seguras em regiões de doses baixas.

No capítulo 3 são apresentados os modelos abordados no capítulo 2 estendidos as situações que envolvem "riscos competitivos". Inicialmente é feito uma breve revisão dos conceitos envolvidos na teoria de riscos competitivos e em seguida é também apresentado em detalhes, os modelos de regressão em tais situações.

Finalmente, é apresentado no capítulo 4, duas aplicações numéricas. As mesmas têm por finalidade ilustrar os modelos apresentados sob os dois contextos em experimentos de dose-resposta.

# CAPÍTULO 1

## CONCEITOS FUNDAMENTAIS EM ANÁLISE DE SOBREVIVÊNCIA

### 1.1 - INTRODUÇÃO

Os ensaios biológicos ou experimentos de dose-resposta são experimentos que têm por finalidade avaliar propriedades de uma droga observando sua atividade em um sistema biológico. Entende-se por sistema biológico um conjunto de unidades experimentais constituído por animais, ou plantas de uma mesma espécie, ou pessoas, etc.

Como a utilização de um sistema biológico em um experimento deste tipo provoca uma considerável variação nas medidas obtidas em repetições aparentemente idênticas do experimento, surge a necessidade da utilização de métodos estatísticos que permitam um melhor entendimento dos dados obtidos.

Para a aplicação de métodos estatísticos adequados faz-se necessário conhecer o tipo de ensaio biológico realizado. Com este intuito e de modo sucinto, é tratado a seguir à classificação dos ensaios biológicos.

Inicialmente, os ensaios biológicos podem ser classificados em :

– **ensaios biológicos qualitativos e ensaios biológicos quantitativos.**

Os ensaios **biológicos qualitativos** têm como objetivo identificar as substâncias contidas na droga de interesse através de reações características provocadas no sistema biológico. Já os **ensaios quantitativos** objetivam avaliar "propriedades" de uma droga e com tal finalidade são realizadas medidas das respostas de interesse nas unidades experimentais as quais receberam doses da referida droga.

Geralmente, um ensaio biológico é caracterizado por :

(a) ser um experimento feito com a droga teste e com a droga padrão, isto é, uma amostra de elementos recebe doses da droga teste e uma outra amostra de elementos recebe doses da droga padrão.

Obs: – as doses e o número de doses dependem do planejamento.

– é considerado que as amostras submetidas a cada uma das doses são independentes.

(b) possuir uma resposta para cada unidade amostral à dose a que foi submetida. Esta resposta pode ser uma medida de alguma propriedade ou apenas o registro da ocorrência ou não de um determinado fenômeno. Obviamente, a intensidade ou frequência da resposta deve depender da dose.

Dependendo da forma em que é feita a escolha das doses a serem aplicadas, os **ensaios biológicos quantitativos** podem ser classificados em :

– **diretos ou indiretos.**

Ainda, dependendo do tipo de resposta observada, os ensaios **indiretos** podem ser:

– **com resposta quantitativa ou com resposta quantal.**

Nos **ensaios quantitativos diretos**, aplicam-se doses crescentes de uma droga a cada uma das unidades do sistema biológico até que ocorra a resposta de interesse. A dose crítica que provoca esta resposta é definida como "tolerância" da unidade experimental à droga a qual foi submetida. Ensaios desse tipo são pouco utilizados em situações práticas devido a dificuldade de sua execução, bem como pela dificuldade na determinação exata da dose crítica.

Nos ensaios quantitativos indiretos com resposta quantitativa cada unidade experimental do sistema biológico recebe somente uma das doses as quais foram fixadas previamente no experimento. As respostas observadas neste tipo de ensaio biológico são medidas tais como: peso, tempo de sobrevivência, etc.

Para certas drogas, no entanto, não é possível obter-se uma medida quantitativa da resposta de interesse. Pode-se contudo, para cada unidade experimental submetida a uma dose determinada, observar se ocorre ou não um evento crítico. Ensaios deste tipo são denominados "ensaios biológicos quantitativos indiretos com resposta quantal".

Tais ensaios são amplamente utilizados em situações práticas e abordados na literatura. Neste tipo de ensaio são escolhidas  $K_T$  doses da droga teste e  $K_S$  doses da droga padrão e a cada dose é submetida uma amostra independente de unidades experimentais observando-se o número dessas unidades para as quais ocorreu o evento crítico (ou resposta positiva).

Os ensaios indiretos com resposta quantal têm semelhança lógica com os ensaios diretos. Em ensaios diretos define-se a "tolerância" de uma unidade experimental como a dose crítica que provoca a resposta de interesse e nos ensaios indiretos com resposta quantal, se a uma dose  $x$  uma proporção  $p$  de unidades experimentais respondem positivamente, então uma proporção  $p$  de indivíduos têm tolerância menor ou igual a  $x$  e uma proporção  $(1 - p)$  de indivíduos têm tolerância maior que  $x$ .

Em geral, em um ensaio quantal deseja-se estimar a potência relativa da droga teste, bem como determinar a DL50 (dose letal 50, isto é, a dose que provoca resposta positiva em 50% das unidades experimentais a ela submetida). É comum na análise de dados de ensaios quantal utilizar-se transformações apropriadas (por exemplo: probit, logit, angular) para linearizar a relação dose-resposta.

Um exemplo de ensaio biológico quantal foi apresentado por Finney (1978). O objetivo do ensaio era medir a potência de uma droga T em relação a insulina padrão S. Para isso foram aplicadas 9 doses de S e 5 doses de T a amostras independentes de camundongos e a cada dose foi registrado o número de camundongos que apresentaram sintomas de convulsão. Os dados obtidos foram:

DROGA S				DROGA T			
$d_i$	$n_i$	$r_i$	$P_i$	$d_i$	$n_i$	$r_i$	$P_i$
3.4	33	0	0.00	6.5	40	2	0.05
5.2	32	5	0.16	10.0	30	10	0.33
7.0	38	11	0.38	14.0	40	18	0.45
8.5	37	14	0.45	21.5	35	21	0.60
10.5	40	18	0.57	29.0	37	27	0.73
13.0	37	21	0.57				
18.0	31	23	0.74				
21.0	37	30	0.81				
28.0	30	27	0.90				

TABELA 1.1 - Conjunto de dados apresentados por Finney (1978) .

OBS: Na tabela 1.1 tem-se:  $d_i$  = níveis das doses,  $n_i$  = unidades amostrais,  $r_i$  = respostas positivas e  $p_i$  = proporção de respostas positivas.

Para a análise destes dados foi utilizado as transformações citadas anteriormente e as conclusões obtidas foram as mesmas, ou seja, o ensaio não apresentou evidências de invalidade.

O objetivo nesta dissertação não é o de analisar ensaios biológicos qualitativos, quantitativos diretos ou quantitativos indiretos com resposta quantal. O enfoque se restringirá aos "ensaios biológicos quantitativos indiretos com resposta quantitativa", mais especificamente àqueles cuja medida associada a resposta seja o tempo de sobrevida, onde os mesmos podem ser tempos observados desde a aplicação da dose até a ocorrência da resposta de interesse, ou tempos observados desde a aplicação da dose até o término pré-fixado do experimento (denominados tempos censurados).

Deseja-se portanto analisar dados do seguinte tipo: Suponha que 5 doses distintas de uma droga foram aplicadas a 5 grupos de ratos, cada qual recebendo uma dessas diferentes doses e que os tempos de sobrevida observados foram:

dose ( $d_j$ )	unidades amostrais ( $n_j$ )	tempos ( $t_j$ )
2.5	11	5 6 6 7 8 8 8 9* 10 11 12
3.4	13	4 7 7 7 9 9 10 12 13 15* 16 17* 18
5.3	15	8 9 10 10 10 14* 16 18 20* 21 21 22 23* 24 24
8.0	12	5 7 10 13 13 13 14* 15 15 15 16 17*
9.4	14	3 4 5 7 * 10 10 12 12 13* 14 14 15 15 16*

Tabela 1.2 - Conjunto de dados hipotético ( $t^*$  = tempo censurado).

Para a realização da análise de dados como os citados acima, utilizar-se-á métodos baseados na teoria de análise de sobrevivência. Tais métodos tratam do ajuste de modelos baseados nos tempos de sobrevivência os quais permitem também a incorporação dos tempos censurados ao modelo.

Por este fato, faz-se nas seções seguintes deste capítulo uma sucinta revisão da teoria básica e das notações utilizadas em análise de sobrevivência, uma vez que esta teoria será necessária na análise de dados do tipo citado na tabela 1.2.

Nos capítulos 2 e 3 serão tratados alguns modelos de regressão possíveis de serem ajustados à dados do tipo acima. Estes modelos obviamente, objetivam auxiliar na análise em que se deseja realizar.

## 1.2 - CONCEITOS BÁSICOS E NOTAÇÕES

Nesta seção é apresentada uma síntese dos principais conceitos básicos e notações utilizadas em análise de sobrevivência. O intuito é facilitar o entendimento dos desenvolvimentos que serão feitos posteriormente.

### 1.2.1 - Censuras

Em muitos experimentos, tais como os experimentos de dose-resposta, onde os tempos de vida são utilizados na análise, há a necessidade de que os mesmos se encerrem num determinado tempo fixo e deste modo nem todos os indivíduos falham até o término pré-fixado destes experimentos. Os indivíduos que sobrevivem a este tempo fixo são chamados de censuras e seu tempo corresponde a um tempo censurado.

Os métodos desenvolvidos em análise de sobrevivência incorporam tais dados censurados, o que não é feito, em geral, em outros tipos de análise. Os tipos de censuras existentes são:

- 1- Censura tipo I -  $\begin{cases} \text{simples} \\ \text{múltipla.} \end{cases}$
- 2- Censura tipo II -  $\begin{cases} \text{simples,} \\ \text{múltipla e progressiva.} \end{cases}$
- 3- Censura Aleatória.

**I - Censura tipo I.** Suponha que  $n$  indivíduos estão em estudo e que a cada indivíduo tem-se associado um tempo de vida  $T_i$ . Seja  $T$  o tempo de seguimento pré-fixado do experimento. Se  $T_i \leq T$  ocorre a falha e se  $T_i > T$  ocorre a censura do tipo I. A censura tipo I pode ocorrer em dois tipos de situações:

**Situação 1** - Os  $n$  indivíduos entram **juntos** no experimento de modo que todos os  $n$  indivíduos têm o mesmo tempo teórico de observação, digamos  $L_i$  ( $i = 1, 2, \dots, n$ ), uma vez que o tempo para o término do experimento é fixo. Os indivíduos sobreviventes são as censuras do tipo I simples.

**Situação 2** - Os  $n$  indivíduos entram no experimento em tempos **diferentes** de modo que possuem tempos teóricos de observação  $L_i$  ( $i = 1, 2, \dots, n$ ) distintos. Uma vez fixado o término do experimento, os indivíduos sobreviventes são as censuras do tipo I múltipla. Portanto, se  $T_i > L_i$  onde  $L_i$  ( $i = 1, 2, \dots, n$ ) são os períodos de observação dos  $n$  indivíduos, censuras múltiplas do tipo I ocorrem. Assim, os indivíduos são representados por  $(t_i, \delta_i)$  onde:

$$t_i = \min (T_i, L_i) \quad e$$

$$\delta_i = \begin{cases} 1 & \text{se } T_i \leq L_i \\ 0 & \text{se } T_i > L_i . \end{cases}$$

II - Censura tipo II. A censura tipo II ocorre quando uma amostra de  $n$  indivíduos é observada até que  $r$  ( $r < n$ ) falhas ocorram. Os  $n - r$  indivíduos restantes são as censuras do tipo II. Nestes tipos de experimentos, tem-se então:

- tempo para o término do experimento aleatório e
- número de falhas fixo.

Note-se que censura tipo II difere da censura tipo I em que se verificam :

- tempo para o término do experimento fixo e
- número de falhas aleatório.

Quando todos os indivíduos são colocados ao mesmo tempo no experimento, ocorrem as censuras simples do Tipo II e quando os indivíduos entram em tempos distintos no experimento, ocorrem as censuras múltiplas do tipo II.

Uma generalização da censura tipo II é a **censura tipo II progressiva**. Neste caso,  $r_1$  falhas na amostra de  $n$  indivíduos são observadas e  $n_1$  dos  $n - r_1$  indivíduos não falhos são removidos do experimento. Dos  $n - r_1 - n_1$  indivíduos que ficaram no experimento,  $r_2$  falhas são observadas e  $n_2$  indivíduos não falhos são retirados do experimento. O experimento termina após uma série pré-estabelecida de repetições.

### III - Censura aleatória

As censuras aleatórias ocorrem geralmente em ensaios clínicos ou em aplicações médicas em estudos com animais. Em ensaios clínicos, por exemplo, os pacientes entram no estudo de modo aleatório de acordo com o tempo de diagnóstico. As censuras aleatórias podem ocorrer de uma das seguintes maneiras:



- 1 - **Perdidos de observação ( loss to follow-up )** : o paciente muda-se para outra localidade desconhecida ou abandona o tratamento.
- 2 - **Retirado do experimento ( drop-out )** : o paciente não se adapta ao tratamento e é necessário retirá-lo.
- 3 - **Término de estudo** : O tempo pré-determinado para o estudo termina e o paciente sobreviveu a esse tempo.

**OBS:** Pode-se então observar que a censura tipo I é um caso especial da censura aleatória.

### 1.2.2 - Função de densidade de probabilidade

Seja  $T$  a variável aleatória tempo de vida, então  $T$  é não negativa e usualmente contínua. A função de densidade de probabilidade ( f.d.p.) da variável aleatória  $T$  é definida como :

$$f_T(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Pr[ um indivíduo falhar no intervalo } [t, t + \Delta t] ]}{\Delta t} . \quad (1.1)$$

A função de densidade tem as seguintes propriedades:

a)  $f_T(t)$  é uma função não negativa, isto é,  $f_T(t) = \begin{cases} \geq 0 & \text{para } t \geq 0 \\ = 0 & \text{para } t < 0. \end{cases}$

b)  $\int_0^{\infty} f_T(t) dt = 1.$

Em consequência, a função de distribuição de probabilidade é dada por:

$$F_T(t) = \int_0^t f_T(u) du = \text{Pr} [ T \leq t]. \quad (1.2)$$

Se o tempo de vida  $T$  é tratado como uma variável aleatória discreta onde  $T$  toma os valores  $t_1, t_2, \dots, t_n$  com  $0 \leq t_1 < t_2 < \dots < t_n$ , então a função de probabilidade pode ser estimada por:

$$p(t_i) = \text{Pr}[T = t_i] = \frac{\# \text{ pacientes falhos no tempo } t_i}{\# \text{ total de pacientes}} . \quad (1.3)$$

Quando T é considerada uma variável aleatória contínua, os modelos paramétricos mais comuns assumidos para T ( ou para uma função de T) e suas respectivas f.d.p. são:

(1) T com distribuição Exponencial ( $\alpha$ ),

$$f_T(t) = \frac{1}{\alpha} \exp\left(-\frac{t}{\alpha}\right), \quad t \geq 0 \text{ e } \alpha > 0. \quad (1.4)$$

(2) T com distribuição Weibull ( $\alpha, \delta$ ),

$$f_T(t) = \frac{\delta}{\alpha} \left(\frac{t}{\alpha}\right)^{\delta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\delta\right\}, \quad t \geq 0, \delta > 0 \text{ e } \alpha > 0. \quad (1.5)$$

(3) Y = log T com distribuição Valor-Extremo ( $\mu, \sigma$ ), onde  $T \sim$  Weibull ( $\alpha, \delta$ ),

$$f_Y(y) = \frac{1}{\sigma} \exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right], \quad -\infty < y < \infty, -\infty < \mu < \infty \text{ e } \sigma > 0, \quad (1.6)$$

onde:  $\sigma = \frac{1}{\delta}$  e  $\log \alpha = \mu$ .

(4) Y = log T com distribuição log - Normal ( $\mu, \sigma$ ),

$$f_Y(y) = \frac{1}{(2\pi)^{1/2} \sigma \exp\{y\}} \exp\left[-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty, -\infty < \mu < \infty \text{ e } \sigma > 0. \quad (1.7)$$

(5) T com distribuição Gama Generalizada ( $\alpha, \delta, k$ ),

$$f_T(t) = \frac{\delta t^{\delta k - 1}}{G(k) \alpha^{\delta k}} \exp\left\{-\left(\frac{t}{\alpha}\right)^\delta\right\}, \quad \alpha, \delta, k > 0 \text{ e } t \geq 0, \quad (1.8)$$

onde G(k) é a função gama, isto é,  $G(k) = \int_0^\infty x^{k-1} \exp(-x) dx$ .

Observe que : i) se  $\delta = k = 1 \Rightarrow T \sim$  Exponencial ( $\alpha$ ).

ii) se  $k = 1 \Rightarrow T \sim$  Weibull ( $\alpha, \delta$ ).

iii) se  $\delta = 1 \Rightarrow T \sim$  Gama ( $\alpha, k$ ).

iv) se  $k \rightarrow \infty \Rightarrow T \sim$  log - Normal.

### 1.2.3 - Função de Sobrevivência

A função de sobrevivência, denotada por  $S_T(t)$ , é definida como a probabilidade de um indivíduo sobreviver ao tempo  $t$ , isto é :

$$S_T(t) = \Pr[T > t] = 1 - \Pr[T \leq t] = 1 - F_T(t). \quad (1.9)$$

Ainda,  $S_T(t) = 1$  para  $t = 0$  e  $S_T(t) = 0$  para  $t = \infty$ .

No caso em que  $T$  é uma variável aleatória discreta,

$$S_T(t) = \Pr[T > t] = \sum_{i: t_i > t} p(t_i),$$

e no caso em que  $T$  é uma variável aleatória contínua,

$$S_T(t) = \int_t^{\infty} f_T(u) du.$$

As respectivas funções de sobrevivência  $S_T(t)$  correspondentes as funções de densidade de probabilidade citadas na seção 1.2.2 são:

(1) Se  $T \sim \text{Exp}(\alpha)$ ,

$$S_T(t) = \exp\left(-\frac{t}{\alpha}\right), \quad t \geq 0 \text{ e } \alpha > 0. \quad (1.10)$$

(2) Se  $T \sim \text{Weibull}(\alpha, \delta)$ ,

$$S_T(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^\delta\right], \quad t \geq 0 \text{ e } \delta, \alpha > 0. \quad (1.11)$$

(3) Se  $Y = \log T \sim \text{Valor - Extremo}(\mu, \sigma)$ ,

$$S_Y(y) = \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right], \quad -\infty < y < \infty, -\infty < \mu < \infty \text{ e } \sigma > 0. \quad (1.12)$$

(4) Se  $Y = \log T \sim \text{Log-Normal}(\mu, \sigma)$ ,

$$S_T(y) = 1 - \Phi\left(\frac{y - \mu}{\sigma}\right), \quad -\infty < y < \infty, \quad -\infty < \mu < \infty \text{ e } \sigma > 0, \quad (1.13)$$

onde:  $\Phi(x) = \int_{-\infty}^x \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}u^2\right) du$ .

(5) Se  $T \sim \text{Gama Generalizada}(\alpha, \delta, k)$ ,

$$S_T(t) = 1 - F_T(t) = 1 - \frac{G_w(k)}{G(k)}, \quad (1.14)$$

onde:  $G_w(k)$  é a função gama incompleta, isto é,  $G_w(k) = \int_0^w x^{k-1} \exp\{-x\} dx$

e  $G(k)$  é a função gama, isto é,  $G(k) = \int_0^\infty x^{k-1} \exp(-x) dx$  e  $w = \left(\frac{x}{\alpha}\right)^\delta$ .

#### 1.2.4 - Função Risco

A função risco do tempo de vida  $T$ , também conhecida como taxa de falha instantânea, denotada por  $\lambda_T(t)$ , é definida por:

$$\begin{aligned} \lambda_T(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr[t < T < t + \Delta t \mid T > t]}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\Pr[T \in (t, t + \Delta t)]}{\Pr[T > t]} = \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Pr[T < t + \Delta t] - \Pr[T < t]}{\Delta t \Pr[T > t]} = \lim_{\Delta t \rightarrow 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t} \frac{1}{S_T(t)} = \\ &= (\text{por definição de derivada}) = \frac{\partial F_T(t)}{\partial t} \frac{1}{S_T(t)} = \frac{f_T(t)}{S_T(t)} = \frac{f_T(t)}{1 - F_T(t)}. \quad (1.15) \end{aligned}$$

Portanto,  $\lambda_T(t)$  pode ser escrita em termos de  $F_T(t)$  e  $f_T(t)$  como mostrado em (1.15).

A função  $\lambda_T(t)$  pode não ter uma forma definida, podendo assim crescer, decrescer, manter-se constante ou indicar um processo mais complicado dependendo da distribuição a que ela pertence.

OBS:  $\lambda_T(t) \geq 0$ .

Para as funções consideradas anteriormente, as funções risco correspondentes são:

(1) Se  $T \sim \text{Exp}(\alpha)$ ,

$$\lambda_T(t) = \frac{1}{\alpha}, \quad t \geq 0 \text{ e } \alpha > 0. \quad (1.16)$$

(2) Se  $T \sim \text{Weibull}(\alpha, \delta)$ ,

$$\lambda_T(t) = \frac{\delta}{\alpha} \left(\frac{t}{\alpha}\right)^{\delta-1}, \quad t \geq 0 \text{ e } \alpha, \delta > 0. \quad (1.17)$$

(3) Se  $Y = \log T \sim \text{Valor-Extremo}(\mu, \sigma)$ ,

$$\lambda_Y(y) = \frac{1}{\sigma} \exp\left[\frac{y-\mu}{\sigma}\right], \quad -\infty < y < \infty, \quad -\infty < \mu < \infty \text{ e } \sigma > 0. \quad (1.18)$$

(4) Se  $Y = \log T \sim \text{log-Normal}(\mu, \sigma)$ ,

$$\lambda_Y(y) = \frac{\frac{1}{(2\pi)^{1/2} \sigma \exp\{y\}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}}{1 - \Phi\left(\frac{y-\mu}{\sigma}\right)}, \quad -\infty < y < \infty, \quad -\infty < \mu < \infty \text{ e } \sigma > 0. \quad (1.19)$$

(5) Se  $T \sim \text{Gama Generalizada}(\alpha, \delta, k)$ ,

$$\lambda_T(t) = \frac{\frac{\delta t^{\delta k - 1}}{G(k) \alpha^{\delta k}} \exp\left\{-\left(\frac{t}{\alpha}\right)^\delta\right\}}{1 - \frac{G_w(k)}{G(k)}}, \quad t \geq 0 \text{ e } \alpha, \delta, k > 0. \quad (1.20)$$

Uma outra função a qual também é bastante utilizada, é a função risco acumulada e esta é definida por :

$$H(t) = \int_0^t \lambda_T(u) du. \quad (1.21)$$

### 1.2.5 - Relações entre as funções

Sejam as funções :

$f_T(t)$  - função de densidade de probabilidade de T,

$S_T(t)$  - função de sobrevivência de T,

$\lambda_T(t)$  - função risco de T.

Para o caso destas funções serem contínuas, tem-se válida as seguintes relações :

$$(1) \lambda_T(t) = \frac{f_T(t)}{S_T(t)} .$$

$$(2) f_T(t) = - \frac{\partial}{\partial t} S_T(t) = - S'_T(t) .$$

$$(3) \lambda_T(t) = \frac{\left( - \frac{\partial}{\partial t} S_T(t) \right)}{S_T(t)} = \frac{- S'_T(t)}{S_T(t)} = - \frac{\partial}{\partial t} \ln S_T(t) .$$

$$(4) S_T(t) = \exp \left\{ - \int_0^t \lambda_T(u) du \right\} .$$

$$(5) f_T(t) = \lambda_T(t) S_T(t) = \lambda_T(t) \exp \left\{ - \int_0^t \lambda_T(u) du \right\} .$$

### 1.3 - MÉTODOS NÃO PARAMÉTRICOS PARA ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA E DA FUNÇÃO RISCO

A escolha errônea da distribuição de probabilidade associada a variável aleatória T pode induzir a conclusões não corretas quando da realização de inferências sobre a população amostrada. Para evitar este problema propõe-se estimar a própria distribuição desconhecida através de um método não paramétrico.

#### 1.3.1 - Método não paramétrico de Kaplan-Meier

Considere uma amostra de tamanho n de uma população com função de distribuição acumulada  $F_T(t)$ . Quando censuras estão presentes, para estimar-se a função de sobrevivência  $S_T(t) = 1 - F_T(t)$ , o que equivale a estimar  $F_T(t)$ , pode-se utilizar a estimativa Produto-Limite (P.L.) derivada por

Kaplan-Meier (1958). Os  $n$  tempos  $t_1, t_2, \dots, t_n$  são ordenados de forma crescente, isto é,  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ . As censuras entram na ordenação e caso censura e falha ocorram ao mesmo tempo, convencionou-se ordenar censura depois da falha. A estimativa da função de sobrevivência é dada por :

$$\hat{S}_T(t) = \prod_{i: t_i \leq t} \left[ 1 - \frac{d_i}{n_i} \right], \quad (1.22)$$

onde :  $d_i = \#$  de indivíduos que falharam no tempo  $t_i$  ,  
 $n_i = \#$  de indivíduos em risco no tempo  $t_i$  ,  
 $d_i/n_i =$  proporção de falhas em  $t_i$  ,  
 $1 - d_i/n_i =$  proporção de sobrevivência a  $t_i$  , depois de ter sobrevivido a  $(i-1)$  tempos.

Conseqüentemente,

$$\hat{F}_T(t) = 1 - \hat{S}_T(t). \quad (1.23)$$

A variância da estimativa  $\hat{S}_T(t)$  é dada por :

$$\text{Var } \hat{S}_T(t) = [\hat{S}_T(t)]^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \quad (1.24)$$

A função risco acumulada pode ser estimada por  $\hat{H}(t_i) = -\ln \hat{S}_T(t_i)$  ou por  $\hat{H}(t_i) = \sum_{j=1}^i \frac{d_j}{n_j}$  para  $i = 1, 2, \dots, n$ , a qual nos fornece a função risco acumulada empírica.

### 1.3.2 - Método atuarial ou tabela de vida

Quando os tempos de sobrevivência estão dispostos em intervalos de classes  $[t_i, t_{i+1})$ , a estimativa da função de sobrevivência é dada pelo método atuarial ou tabela de vida. Esta função é então estimada ( Elisa Lee - 1980) por :

$$\hat{S}(t_i) = \prod_{j=1}^{i-1} \hat{p}_j \quad \Rightarrow \quad \begin{cases} \hat{S}(t_1) = 1 \\ \hat{S}(t_i) = \hat{p}_{i-1} \hat{S}(t_{i-1}), \quad i = 2, 3, \dots, s \end{cases} \quad (1.25)$$

onde :  $\hat{p}_i = 1 - \hat{q}_i \rightarrow$  proporção condicional de sobrevivência no intervalo  $i$ ,  
 $\hat{q}_i = \frac{d_i}{n'_i} \rightarrow$  proporção condicional de falha no intervalo  $i$ ,  
 $n'_i = n_i - \frac{1}{2}(l_i + w_i) \rightarrow$  # de indivíduos expostos ao risco no intervalo  $i$ ,  
 $n_i \rightarrow$  # de indivíduos que entraram no  $i$ -ésimo intervalo,  
 $d_i \rightarrow$  # de indivíduos que falharam no  $i$ -ésimo intervalo,  
 $l_i \rightarrow$  # de indivíduos perdidos de observação no  $i$ -ésimo intervalo,  
 $w_i \rightarrow$  # de indivíduos retirados vivos no  $i$ -ésimo intervalo.

A variância de  $\hat{S}(t_i)$  foi derivada por Greenwood (1926) e é aproximada por :

$$\text{Var}(\hat{S}(t_i)) \simeq [\hat{S}(t_i)]^2 \sum_{j=1}^{i-1} \frac{d_j}{n'_j (n'_j - d_j)}, \quad i = 2, 3, \dots, s. \quad (1.26)$$

A função risco é estimada por :

$$\hat{h}(t_{mi}) = \frac{2 \hat{q}_i}{b_i (1 + \hat{p}_i)}, \quad i = 1, 2, \dots, s-1, \quad (1.27)$$

onde :  $\begin{cases} t_{mi} = \text{ponto médio do intervalo } [t_i, t_{i+1}) \\ b_i = \text{tamanho do intervalo } [t_i, t_{i+1}). \end{cases}$

A variância da estimativa  $\hat{h}(t_{mi})$  é dada por :

$$\text{Var}(\hat{h}(t_{mi})) = \frac{[\hat{h}(t_{mi})]^2}{n'_i \hat{q}_i} \left\{ 1 - \left[ \frac{\hat{h}(t_{mi}) \cdot b_i}{2} \right]^2 \right\}. \quad (1.28)$$

#### 1.4 - MÉTODO PARAMÉTRICO PARA ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA

No caso paramétrico a função de sobrevivência é obtida através da estimação dos seus parâmetros usando os estimadores de máxima verossimilhança (e.m.v). Assim, tendo assumido uma família paramétrica aos tempos  $T$ , com função de distribuição  $F_T(t)$ , pode-se estimar a função de sobrevivência  $S_T(t) = 1 - F_T(t)$



através da estimação dos parâmetros de  $F_T(t)$ . Esses parâmetros são estimados pelo método da máxima verossimilhança.

Para obter-se tais estimadores é necessário estabelecer a função de verossimilhança considerando o tipo de censura atuando sobre os dados.

A seção 1.4.1 trata dessas funções de verossimilhança e a seção 1.4.2 trata do método iterativo de Newton-Raphson, o qual é normalmente utilizado para obtenção das estimativas dos parâmetros envolvidos.

#### 1.4.1 - Função de verossimilhança

A forma da função de verossimilhança para dados de sobrevivência pode ser estudada considerando-se os três tipos de censura.

Suponha que os dados consistem de  $n$  observações independentes  $t_1, t_2, \dots, t_n$  de uma variável aleatória cuja função de densidade de probabilidade envolva  $p$  parâmetros  $\theta_i$ , isto é,  $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p)'$ . A função de verossimilhança, denotada por  $L(\underline{\theta})$  será dada por:

##### (a) Considerando-se a existência de censuras tipo I

Sejam os tempos  $t_1, t_2, \dots, t_n$  os quais são tempos de falha ou tempos de censura do tipo I, isto é, os tempos de censura são consequência do término do experimento. Considere o conjunto  $F$  dos indivíduos que falharam e o conjunto  $C$  dos indivíduos censurados. A função de verossimilhança pode ser particionada nos conjuntos  $F$  e  $C$  e é expressa por:

$$L(\underline{\theta}) = \prod_F f_T(t_i, \underline{\theta}) \prod_C S_T(t_i, \underline{\theta}), \quad (1.29)$$

onde:  $f_T(t_i, \underline{\theta})$  é a função de densidade de probabilidade de  $T$  e

$S_T(t_i, \underline{\theta})$  é a função de sobrevivência de  $T$ .

**(b) Considerando-se a existência de censuras tipo II**

Seja  $t_1, t_2, \dots, t_n$  os tempos observados de  $n$  indivíduos nos quais  $r$  deles são tempos de falha e  $n - r$  deles são tempos censurados e tais censuras são do tipo II. Então, somente as  $r$  menores observações da amostra de tamanho  $n$  são observadas ( $r$  é fixo,  $1 \leq r \leq n$ ). Seja  $t_{(1)} < t_{(2)} \dots < t_{(r)}$  as  $r$  menores observações, a função de verossimilhança neste caso é :

$$L(\underline{\ell}) = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r f_T(t_{(i)}; \underline{\ell}) \right] [S_T(t_{(r)}; \underline{\ell})]^{n-r}, \quad (1.30)$$

onde :  $f_T(t_{(i)}; \underline{\ell})$  é a função de densidade de probabilidade de  $T$ ,  
 $S_T(t_{(r)}; \underline{\ell})$  é a função de sobrevivência de  $T$ ,  
 $t_{(r)}$  é o tempo de vida correspondente a  $r$ -ésima falha.

**(c) Considerando-se a existência de censura aleatória**

Ao assumir um processo aleatório para censura considera-se:  $T$  o tempo de falha e  $L$  o tempo de censura. Para cada indivíduo observa-se  $(t_i, \delta_i)$ ,  $i = 1, 2, \dots, n$  onde :

$$t_i = \min(T_i, L_i) \text{ e } \delta_i = \begin{cases} 1 & \text{se } T_i \leq L_i \\ 0 & \text{se } T_i > L_i \end{cases}$$

Considerando-se os pares  $(T_i, L_i)$  independentes e ainda :

$$\begin{cases} f_T(t; \underline{\ell}) & \text{a f.d.p. de } T \\ g_L(t) & \text{a f.d.p. de } L \end{cases} \text{ e } \begin{cases} S_T(t; \underline{\ell}) & \text{a f. sobrevivência de } T \\ G_L(t) & \text{a f. de sobrevivência de } L \end{cases}$$

e como, a probabilidade de um indivíduo ter um tempo de sobrevivência  $t$  e ser censurado neste tempo é dada por :

$$\begin{aligned} \Pr[t_i=t; \delta_i=0] &= \Pr[L_i=t; T_i > L_i] = \Pr[L_i=t; T_i > t] = \\ &= (\text{independência}) = \Pr[L_i=t] \Pr[T_i > t] = g_L(t) S_T(t; \underline{\ell}), \end{aligned}$$

e a probabilidade de um indivíduo ter um tempo de sobrevivência  $t$  e não ser censurado neste tempo é :

$$\begin{aligned} \Pr[ t_i = t ; \delta_i = 1 ] &= \Pr[ T_i = t ; T_i \leq L_i ] = \Pr[ T_i = t ] \Pr[ L_i \geq T ] = \\ &= f_T ( t ; \varrho ) G_L(t), \end{aligned}$$

a função de verossimilhança de  $(t_i, \delta_i)$ ,  $i = 1, 2, \dots, n$  pode ser escrita por :

$$\begin{aligned} L(\varrho) &= \prod_F f_T( t_i ; \varrho ) G_L(t_i) \prod_C g_L(t_i) S_T ( t_i ; \varrho ) = \\ &= \prod_{i=1}^n [ f_T(t_i ; \varrho) G_L(t_i) ]^{\delta_i} [ g_L(t_i) S_T( t_i ; \varrho ) ]^{1-\delta_i} = \\ &= \prod_{i=1}^n f_T ( t_i ; \varrho )^{\delta_i} S_T ( t_i ; \varrho )^{1-\delta_i} g_L(t_i)^{1-\delta_i} G_L(t_i)^{\delta_i}. \end{aligned}$$

Em geral, o parâmetro de interesse é aquele associado ao tempo de sobrevivência e pode-se observar que  $G_L(t)$  e  $g_L(t)$  não envolvem qualquer parâmetro de interesse. Assim, a função de verossimilhança acima pode ser simplificada para :

$$L(\varrho) \cong \prod_{i=1}^n f_T( t_i ; \varrho )^{\delta_i} S_T ( t_i ; \varrho )^{1-\delta_i}. \quad (1.31)$$

O método da máxima verossimilhança para estimação dos parâmetros nos modelos paramétricos requer a maximização das funções de verossimilhança ( 1.29), (1.30) e (1.31), ou seja, requer a solução das equações :

$$\frac{\partial \ln L(\varrho)}{\partial \theta_i} = 0 \quad \text{para } i = 1, 2, \dots, p.$$

Geralmente, os estimadores não são encontrados analiticamente. Assim métodos iterativos de maximização são utilizados, entre os quais, o método iterativo de Newton - Raphson é o mais usado.

#### 1.4.2 - Método iterativo de Newton - Raphson

O método iterativo de Newton- Raphson é um procedimento iterativo numérico que pode ser usado para resolver equações não-lineares. Essa técnica é baseada na expansão da série de Taylor de primeira ordem.

Seja  $U(\underline{\theta}) = \frac{\partial \ln L(\underline{\theta})}{\partial \underline{\theta}}$ , o qual é chamado vetor de escores. Dado um valor inicial  $\underline{\theta}_o$ , a estatística escore  $U(\hat{\underline{\theta}})$  pode ser escrita como :

$$U(\hat{\underline{\theta}}) = U(\underline{\theta}_o) - i(\underline{\theta}^*)(\hat{\underline{\theta}} - \underline{\theta}_o), \quad (1.32)$$

onde  $\underline{\theta}^*$  está "entre"  $\underline{\theta}_o$  e  $\hat{\underline{\theta}}$  e  $i(\underline{\theta}^*) = \left[ - \frac{\partial^2 \ln L(\underline{\theta})}{\partial^2 \underline{\theta}} \right] |_{\underline{\theta} = \underline{\theta}^*}$ .

Para  $\underline{\theta}_o$  na vizinhança de  $\hat{\underline{\theta}}$ ,  $i(\underline{\theta}^*)$  é aproximadamente igual a  $i(\underline{\theta}_o)$ , tal que fazendo  $U(\hat{\underline{\theta}}) = 0$  e resolvendo-se (1.32) para o primeiro passo, obtem-se :

$$\hat{\underline{\theta}}_1 = \underline{\theta}_o + i(\underline{\theta}_o)^{-1} U(\underline{\theta}_o). \quad (1.33)$$

Os passos subsequentes são dados por :

$$\hat{\underline{\theta}}_k = \hat{\underline{\theta}}_{k-1} + i(\hat{\underline{\theta}}_{k-1})^{-1} U(\hat{\underline{\theta}}_{k-1}) \quad \text{para } k = 2, 3, \dots \quad (1.34)$$

A expressão (1.34) fornece um novo valor para  $\hat{\underline{\theta}}_k$ , de modo que o processo é repetido até que as sucessivas estimativas  $\hat{\underline{\theta}}_k$  aproxime a um determinado valor, isto é, o processo será repetido até que

$$\frac{\|\hat{\underline{\theta}}_k\|}{\|\hat{\underline{\theta}}_{k-1}\|} < \epsilon.$$

Alguns comentários sobre o método devem ser feitos aqui:

- o método trabalha bem se  $i(\underline{\theta})$  é razoavelmente estável sobre um intervalo de valores próximo a  $\hat{\underline{\theta}}$ , isto é, se a função de verossimilhança é próxima de uma curva normal.
- A teoria da verossimilhança assintótica garante a curva Normal para amostras "grandes", sendo que nestes casos o método é eficiente.
- o método tipicamente falha se a verossimilhança é multimodal.

## 1.5 - VARIÁVEIS CONCOMITANTES

Além da variável aleatória tempo  $T$ , outras informações podem também ser avaliáveis, como é o caso de variáveis adicionais descrevendo diferentes condições nas quais o experimento foi realizado. No caso particular de experimentos de dose-resposta, uma variável adicional de interesse é a dose. Essas variáveis adicionais são chamadas variáveis concomitantes ou covariáveis.

Assim, em um experimento de dose-resposta, cada indivíduo possui um tempo de falha ou um tempo de censura e uma dose  $d$  a qual é uma variável concomitante. Em geral, nos experimentos de dose-resposta, deseja-se saber o efeito da dose  $d$  sobre o tempo de falha  $T$  ou ainda estimar o risco de um indivíduo num tempo  $t$  com uma fixada dose  $d$ .

Quando considera-se variáveis concomitantes, trabalha-se de forma análoga ao que foi visto nas seções anteriores, com a diferença de que as funções são agora funções condicionais da variável aleatória  $T$  dado  $a(s)$  variável(s) concomitante(s)  $\underline{x}$ , isto é:

$F_T(t|\underline{x})$  é a função de distribuição acumulada de  $T$  dado  $\underline{x}$ ,

$S_T(t|\underline{x})$  é a função de sobrevivência de  $T$  dado  $\underline{x}$ ,

$\lambda_T(t|\underline{x})$  é a função risco de  $T$  dado  $\underline{x}$ ,

$f_T(t|\underline{x})$  é a f.d.p. de  $T$  dado  $\underline{x}$ .

Por exemplo, se é assumido que os tempos  $T$ , dado as variáveis concomitantes  $\underline{x}$  tem distribuição Weibull  $(\alpha(\underline{x}), \delta)$  onde  $\alpha(\underline{x})$  é o parâmetro de escala o qual depende de  $\underline{x}$  e  $\delta$  é o parâmetro de forma, tem-se:

$$F_T(t|\underline{x}) = 1 - \exp\left\{-\left(\frac{t}{\alpha(\underline{x})}\right)^\delta\right\}, \quad t \geq 0,$$

$$f_T(t|\underline{x}) = \frac{\partial}{\partial t} F_T(t|\underline{x}) = \frac{\delta}{\alpha(\underline{x})} \left(\frac{t}{\alpha(\underline{x})}\right)^{\delta-1} \exp\left\{-\left(\frac{t}{\alpha(\underline{x})}\right)^\delta\right\}, \quad t \geq 0,$$

$$S_T(t|\underline{x}) = 1 - F_T(t|\underline{x}) = \exp\left\{-\left(\frac{t}{\alpha(\underline{x})}\right)^\delta\right\}, \quad t \geq 0,$$

$$\lambda_T(t|\underline{x}) = \frac{\delta}{\alpha(\underline{x})} \left(\frac{t}{\alpha(\underline{x})}\right)^{\delta-1}, \quad t \geq 0.$$

## CAPÍTULO 2

### MODELOS DE REGRESSÃO EM EXPERIMENTOS DE DOSE-RESPOSTA EM UM CONTEXTO DE UM ÚNICO RISCO

#### 2.1 - INTRODUÇÃO

Muitas situações na prática envolvem populações heterogêneas nas quais é desejado considerar a relação entre certas variáveis concomitantes e os tempos para a ocorrência de um evento de interesse. Estes tempos são frequentemente referidos como tempos de falha ou tempos de vida, isto é, o tempo de um indivíduo contado a partir de um início pré-estabelecido até que ocorra a resposta de interesse.

Uma destas situações é aquela que trata de experimentos de dose-resposta quantitativos indiretos com resposta quantitativa. Nestes experimentos, cada um dos diversos grupos experimentais recebe uma das diversas doses de uma determinada droga em teste e esses grupos são então acompanhados por um especificado período de exposição. Os tempos dos indivíduos que falham antes do período de exposição são os tempos de falha. Pode-se ter ao final deste período de exposição indivíduos sobreviventes, ou seja, indivíduos cuja resposta de interesse não ocorreu neste período. Nestes casos, registra-se os tempos decorrido desde a

aplicação da dose até o término do experimento e estes são referidos como tempos de vida censurados.

Considerações com respeito ao custo destes experimentos de dose-resposta conduzem, em geral, a estudos pequenos que envolvem um grupo controle e três a quatro grupos nos quais diferentes doses são aplicadas. Estes estudos, embora pequenos, fornecem valiosas informações à respeito da relação dose-resposta.

Portanto, nestes experimentos consideram-se os tempos  $T$  (os quais correspondem a uma falha ou uma censura) para a ocorrência da resposta de interesse e a variável concomitante dose  $d$ . Cabe lembrar que pode-se tomar outras variáveis concomitantes dos indivíduos amostrados além da dose, ou seja, pode-se considerar um vetor  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  de  $p$  variáveis concomitantes e este vetor pode incluir tanto variáveis quantitativas quanto variáveis qualitativas. O interesse é então modelar e determinar a relação entre os tempos e a dose.

Uma maneira de determinar essa relação é através de modelos de regressão, onde a dependência dos tempos de vida  $T$  sobre a variável concomitante dose  $d$  é explicitamente admitida. Os modelos de regressão a serem considerados para estes experimentos biológicos envolvem situações de análise de sobrevivência, uma vez que dados censurados são considerados.

Em geral, os experimentos biológicos têm como resposta considerada uma variável dicotômica, como por exemplo, ocorrência ou não de falha. Deste modo, os tempos até a ocorrência da falha ou da censura dos indivíduos não são levados em consideração. É importante portanto ressaltar que não é essa a abordagem tratada aqui. A abordagem considerada trata de modelos para experimentos biológicos onde a medida associada a resposta é a variável contínua **tempo** onde os tempos de falha bem como os tempos censurados são devidamente considerados. Em síntese, o objetivo é relacionar os **tempos de vida** com a **dose** utilizando modelos de regressão em situações de análise de sobrevivência.

Neste capítulo, serão considerados alguns modelos de regressão. Caso a distribuição de  $T | \mathbf{x}$  seja especificada, um modelo de regressão paramétrico é considerado e caso contrário, considera-se um modelo semi-paramétrico.

Na seção 2.2 destacam-se as principais funções de interesse na análise de sobrevivência, funções estas as quais utilizam o tempo como medida

associada a resposta de interesse. Na seção 2.3 são tratados detalhadamente os **modelos paramétricos: de locação-escala (em particular o log-linear) e o de riscos proporcionais**, bem como o **modelo semi-paramétrico de riscos proporcionais de Cox**. Na seção 2.4 são apresentadas algumas sugestões gráficas as quais auxiliam na escolha de um modelo inicial apropriado. Na seção 2.5 é discutida a análise de resíduos tanto para os modelos paramétricos quanto para o modelo semi-paramétrico. Os resíduos tratados em tal seção baseiam-se na definição geral de resíduos dada por Cox e Snell (1968). Finalmente, as seções 2.6 e 2.7 tratam, respectivamente, da estimação dos elementos de  $\beta$  (coeficientes de regressão) pelo método da máxima verossimilhança e de testes e intervalos estimados para estes coeficientes de regressão.

## 2.2 - MODELOS UTILIZANDO OS TEMPOS COMO RESPOSTA

O tempo  $T$ ,  $T > 0$ , é uma variável aleatória contínua e pode ser, portanto, definida por um modelo paramétrico. Seja então,

$$F(t | d) = \Pr( T \leq t | d ), \quad (2.1)$$

a função de distribuição acumulada de  $T$  na dose  $d$  (ou em uma função da dose), ou seja, a probabilidade de ocorrer a resposta de interesse no tempo  $t$  ou anterior a  $t$  na dose  $d$ .

Na análise de sobrevivência, as funções de maior interesse são: a função de sobrevivência  $S(t | d)$  e a função risco  $\lambda(t | d)$ , definidas a seguir.

a) A **função de sobrevivência** de  $T$  na dose  $d$ , dada por:

$$S(t | d) = 1 - F(t | d) = 1 - \Pr( T \leq t | d ), \quad (2.2)$$

é a probabilidade de não ocorrer a resposta de interesse no intervalo  $(0, t]$  e dose  $d$ . Em consequência, pode-se definir:

$$f(t | d) = - \frac{\partial}{\partial t} S(t | d) . \quad (2.3)$$



b) A função risco de T na dose d, definida por :

$$\lambda(t | d) = \frac{f(t | d)}{S(t | d)}, \quad \lambda(t | d) \geq 0 \quad (2.4)$$

é a taxa instantânea de ocorrer a resposta de interesse ao tempo t dado que a mesma não ocorreu até o tempo t na dose d.

c) Outra função muito útil na análise de sobrevivência é a função risco acumulada e esta é definida por:

$$H(t | d) = \int_0^t \lambda(u | d) du. \quad (2.5)$$

$$\begin{aligned} \text{Como: } H(t | d) &= \int_0^t \lambda(u | d) du = \int_0^t \frac{f(u | d)}{S(u | d)} du = \\ &= \int_0^t \left[ - \frac{\partial}{\partial u} S(u | d) / S(u | d) \right] du = \\ &= \int_0^t - \frac{\partial}{\partial u} [ \ln S(u | d) ] du = - \ln S(t | d). \end{aligned}$$

Segue, aplicando a exponenciação, que :

$$S(t | d) = \exp \{ - H(t | d) \} = \exp \left\{ - \int_0^t \lambda(u | d) du \right\}. \quad (2.6)$$

Assim, a função risco  $\lambda(t | d)$  caracteriza a distribuição de T. A função risco é portanto, a função de interesse a ser analisada e o estudo da sua variação de acordo com as variáveis concomitantes é o objeto dos modelos de regressão. Isto é, deseja-se estudar os efeitos das variáveis concomitantes na função risco. Em particular, no modelo de dose-resposta, deseja-se estudar o efeito da dose no risco de falha.

Um outro interesse nos experimentos de dose-resposta, é o de estabelecer níveis de tolerância ou de aceitabilidade do risco. Para isto, é necessário especificar alguma conveniente medida do risco a qual possa ser usada para esse propósito. Considere  $F(t | d)$ , a função de distribuição acumulada no tempo t dado uma dose d. Desde que  $F(t | d)$  para um tempo t fixo, especifica a probabilidade de uma resposta como função da dose d e desde que a exposição aos agentes do meio ambiente é baixa, a forma da curva de dose-resposta como função de

$d$  na região de doses baixas é de particular interesse. O método mais conhecido para a determinação da tolerância é o método das doses virtualmente seguras (D V S) o qual é definido por:

$$\Pi(t | d_0) = \pi, \quad (2.7)$$

onde:  $t$  : tempo fixado,

$d_0$  : dose a ser determinada e que deve assegurar o risco fixado como aceitável,

$\pi$  : quantidade a ser "reduzida" ou "adicionada" em  $F(t | 0)$  = probabilidade da resposta ocorrer em  $d=0$  e tempo  $t$ , para obtenção do risco fixado como aceitável.

Para a obtenção da dose virtualmente segura,  $d_0 = d_0(t)$ , utiliza-se a curva de dose-resposta  $\Pi(t | d)$  a qual é uma função da dose no tempo  $t$  fixo e é definida de acordo ao objetivo do experimento conforme descrito a seguir :

(1) Se o objetivo consiste em "reduzir" a probabilidade de falha dos indivíduos e para isso doses crescentes ( $d_1 < d_2 < \dots < d_k$ ) de uma droga são aplicadas para investigação dos seus efeitos positivos, ou seja, para investigar qual das doses analisadas produz uma maior redução na probabilidade de falha  $F(t | 0)$ , então  $\Pi(t | d)$  é denominada de risco reduzido sobre  $F(t | 0)$  e é definida por :

$$\Pi(t | d) = F(t | 0) - F(t | d) = S(t | d) - S(t | 0), \quad t \text{ fixo e } d > 0,$$

onde:  $F(t | 0) = \Pr [T \leq t | d=0]$  e  $S(t | 0) = 1 - F(t | 0)$ ,  $t$  fixo

$F(t | d) = \Pr [T \leq t | d]$  e  $S(t | d) = 1 - F(t | d)$ ,  $t$  fixo e  $d > 0$ .

Experimentos com este objetivo ocorrem comumente na área biomédica onde o pesquisador deseja conhecer a dose  $d_0$  a qual produz a fixada redução  $\pi$  ( $0 < \pi < F(t | 0)$ ) na probabilidade de falha  $F(t | 0)$ . Em geral, o mínimo valor aceitável para o risco ou equivalentemente a máxima redução possível em  $F(t | 0)$  é especificada pelo pesquisador com base no sua experiência e conhecimento do assunto.

As figuras 2.1 e 2.2 abaixo, ilustram uma situação onde três doses crescentes  $d_1 < d_2 < d_3$  estão sendo investigadas.

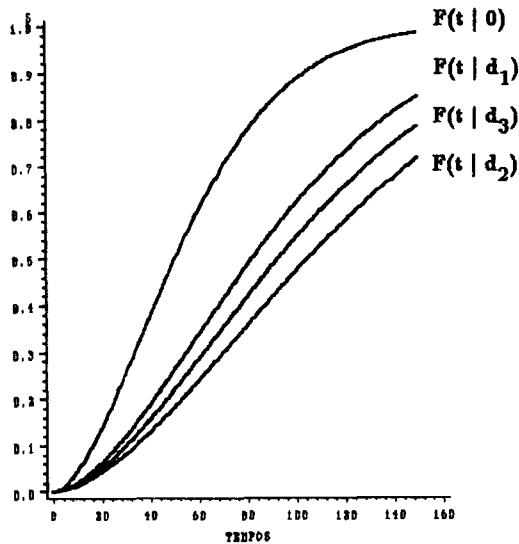


Figura 2.1 - Funções de distribuição acumulada.

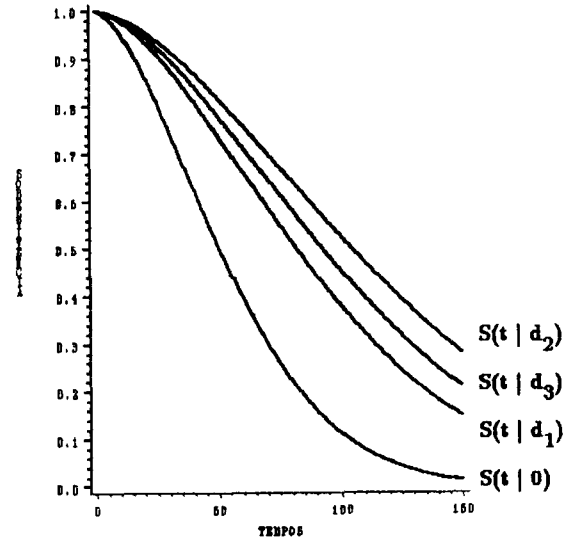


Figura 2.2 - Funções de Sobrevivência.

Pode-se observar que  $\Pi(t | 0) = F(t | 0) - F(t | d) = S(t | d) - S(t | 0)$  é maior para a dose  $d_2$ , ou seja, a dose  $d_2$  produz a maior redução sobre  $F(t | 0)$  dentre as três doses analisadas em um tempo fixo  $t$ .

(2) Se o objetivo é o oposto ao tratado até o momento, ou seja, é desejado "aumentar" a probabilidade de falha dos indivíduos e para isso doses crescentes de uma substância são aplicadas para a investigação dos seus efeitos negativos, então  $\Pi(t | d)$  denota o **risco adicional** sobre  $F(t | 0)$  e é definida por :

$$\Pi(t | d) = F(t | d) - F(t | 0) = S(t | 0) - S(t | d), \quad t \text{ fixo e } d > 0.$$

Experimentos com tal objetivo ocorrem frequentemente na área industrial onde o efeito de dosagens crescentes de uma determinada variável concomitante é estudada através de "testes acelerados". Por exemplo, se é desejado estudar o efeito negativo da voltagem em lâmpadas, então voltagens crescentes são aplicadas para um melhor entendimento deste efeito.

Na área médica, embora com menor frequência, experimentos são realizados com tal objetivo. Como exemplo, tem-se os experimentos relacionados ao Cancer, onde o interesse concentra-se no estudo do efeito negativo causado por doses crescentes de carcinogeno. Estes experimentos são geralmente realizados com animais.

O máximo risco adicional  $\pi$  ( $0 < \pi < 1 - F(t | 0)$ ), é em geral estabelecido pelo pesquisador de acordo aos seus objetivos e conhecimento do assunto.

As formas comumente encontradas em doses baixas para  $\Pi(t | d)$  são ilustradas na figura 2.3 abaixo. Nota-se que convexidade, linearidade e concavidade produzem progressivamente valores maiores de  $d_0$ .

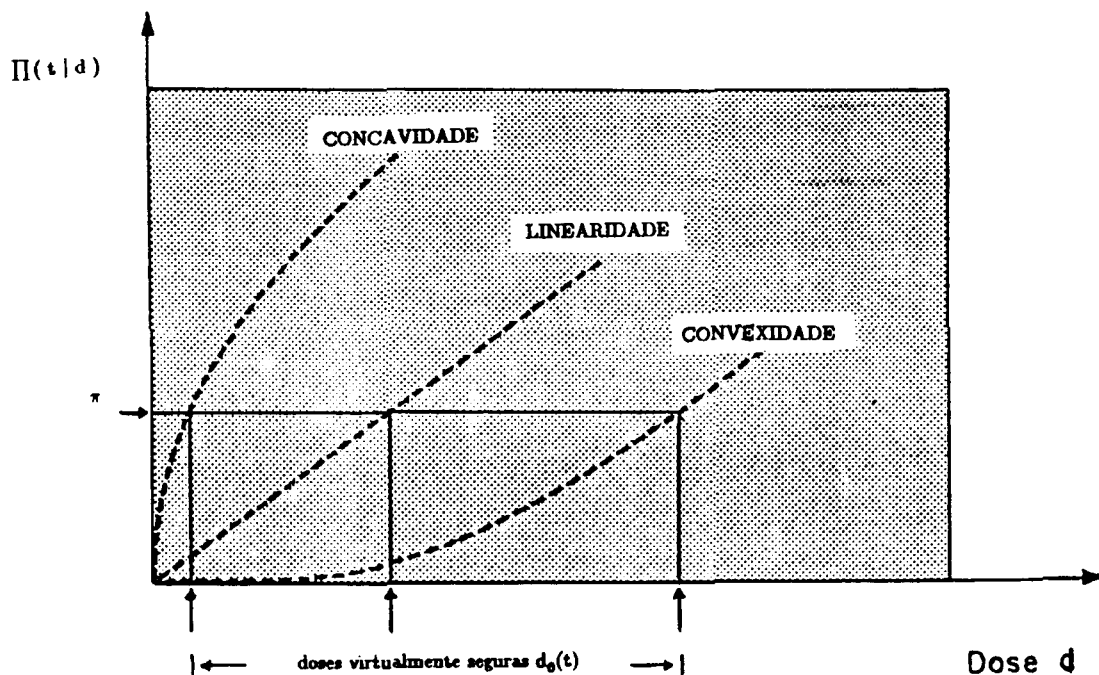


Figura 2.3 - Convexidade, linearidade e concavidade em doses baixas.

Retirado do artigo de Kalbfleish, J.D. et al (1983).

Outra medida de interesse em experimentos de dose-resposta é o "tempo médio para a resposta" até um especificado tempo  $t$ . Se  $T_o = \min\{T, t\}$ , com  $T$  = tempo observado e  $t$  algum tempo especificado, o tempo médio para a resposta na dose  $d$  pode ser expresso, segundo Kalbfleisch et al (1983), por :

$$t_{mR} = E(T_o | d) = \int_0^t (1 - F(u | d)) du = \int_0^t S(u | d) du, \quad (2.8)$$

onde:  $F(t | d)$  é a função de distribuição de  $T$  dado a dose  $d$ .

### 2.3 - MODELOS DE REGRESSÃO

Como já foi mencionado anteriormente, deseja-se relacionar os tempos de vida  $T$  com as variáveis concomitantes  $\underline{x}$  utilizando-se modelos de regressão. Em particular, a variável concomitante de interesse é a dose.

Os seguintes modelos de regressão serão considerados:

- 1- Modelos paramétricos : - de locação-escala  
- de riscos proporcionais
- 2- Modelo semi-paramétrico : - de riscos proporcionais de Cox.

Nos modelos de regressão paramétricos, é assumido uma distribuição de probabilidade  $f(t; \theta)$  associada aos tempos de vida  $T$  dado as variáveis concomitantes  $\underline{x}$ . O modelo de locação-escala pode ser escrito por:

$$E[\log T] = \mu(\underline{x}, \underline{\beta}) + K \quad (2.9)$$

ou

$$E[T] = \exp\{\mu(\underline{x}, \underline{\beta})\} \exp\{K\},$$

onde  $\mu$  é uma função que depende das variáveis concomitantes  $\underline{x}$  e dos parâmetros desconhecidos  $\underline{\beta}$  e  $K$  é uma constante. Assim, pode-se notar que no modelo de locação-escala, tem-se para  $Y = \log T$  um modelo aditivo e para  $T$  um modelo multiplicativo. O modelo de riscos proporcionais é definido por :

$$\lambda(t | \underline{x}) = \lambda_o(t) g(\underline{x}, \underline{\beta}), \quad (2.10)$$

onde :  $\lambda_o(t)$  é a função risco básica, a qual depende somente dos tempos.

No modelo de regressão semi-paramétrico de Cox, nenhuma família paramétrica é associada aos tempos de vida  $T$  dado as variáveis concomitantes  $\underline{x}$ . O modelo de regressão para este caso é definido por :

$$\lambda(t | \underline{x}) = \lambda_0(t) \exp \left\{ \sum_i x_i \beta_i \right\}, \quad (2.11)$$

onde:  $\lambda_0(t)$  é uma função risco básica arbitrária não especificada para  $T$  contínuo.

As seções a seguir tratam esses modelos em maiores detalhes.

### 2.3.1 - Modelos de locação-escala

Uma vez tendo assumido uma família paramétrica à distribuição dos tempos de vida  $T$  dado as variáveis concomitantes  $\underline{x}$ , é frequentemente útil trabalhar-se com o log desses tempos, isto é, com  $Y = \log(T | \underline{x})$ . Se  $Y$  possui distribuição com um parâmetro de locação  $\mu(\underline{x}, \underline{\beta})$  (onde  $\underline{\beta}$  é um vetor de parâmetros desconhecidos) e com um parâmetro de escala constante  $\sigma$ , o modelo de regressão para  $Y = \log(T | \underline{x})$  pode ser escrito por:

$$Y = \mu(\underline{x}, \underline{\beta}) + \sigma W, \quad (2.12)$$

onde  $\mu(\underline{x}, \underline{\beta})$  é o parâmetro de locação,  $\sigma > 0$  é o parâmetro de escala da distribuição de  $Y = \log(T | \underline{x})$  e  $W$  é a variável aleatória erro a qual tem distribuição conhecida e é independente de  $\underline{x}$ . Alternativamente, o modelo de regressão (2.12) pode ser escrito por:

$$E[Y] = \mu(\underline{x}, \underline{\beta}) + K, \quad (2.13)$$

onde  $K = \sigma E[W]$ . Diversas formas funcionais para  $\mu(\underline{x}, \underline{\beta})$  podem ser empregadas. Em particular, se :

$$\mu(\underline{x}, \underline{\beta}) = \underline{x} \underline{\beta}, \quad (2.14)$$

tem-se para (2.12) um "modelo log-linear". Tal forma funcional será usada no decorrer desta dissertação.

A tabela a seguir sumariza algumas distribuições dos erros que são frequentemente encontradas no modelo (2.12).

distribuição de T	distribuição de W
log - Normal	Normal - padrão
Weibull	Valor Extremo - padrão
Exponencial	Valor Extremo - padrão
Gama Generalizada	Log -gama

Tabela 2.1 - Distribuição dos erros W encontradas no modelo de regressão log-linear dado por  $Y = \underline{x} \underline{\beta} + \sigma W$ .

**TEOREMA:** Em termos da função risco para os tempos de vida T dado o vetor de variáveis concomitantes  $\underline{x}$ , tem-se que a função risco para T dado  $\underline{x}$  é dada por:

$$\lambda(t | \underline{x}) = \lambda_0 \left( t \exp \{ -\underline{x} \underline{\beta} \} \right) \exp \{ -\underline{x} \underline{\beta} \}. \tag{2.15}$$

*Demonstração:*

Primeiramente, sabe-se que o modelo de regressão para  $Y = \log(T | \underline{x})$  é  $Y = \mu(\underline{x}, \underline{\beta}) + \sigma W$ . Assim, o modelo de regressão para  $T = \exp Y$  é:

$$T = \exp \left\{ \mu(\underline{x}, \underline{\beta}) \right\} T', \tag{2.16}$$

onde:  $T' = \exp \{ \sigma W \}$  e conseqüentemente  $T' = T \exp[-\mu(\underline{x}, \underline{\beta})]$ . Sabe-se também que a distribuição de W (erros),  $F_W(w)$ , é conhecida e portanto a distribuição de  $T' = \exp\{\sigma W\}$  é facilmente encontrada ser:

$$F_W(\log t' / \sigma).$$

Portanto,  $S(t | \underline{x}) = 1 - F_T(t | \underline{x}) = 1 - F_W[(\log t - \underline{x} \underline{\beta}) / \sigma] = S_W[(\log t - \underline{x} \underline{\beta}) / \sigma] =$

$$= S_1 \left[ \left( \frac{t}{\exp \{ \underline{x} \underline{\beta} \}} \right)^\delta \right] = S_0 \left[ \frac{t}{\exp \{ \underline{x} \underline{\beta} \}} \right], \text{ onde } S_1(w) = S(\log w), \delta = 1/\sigma \text{ e } S_0(w) = S_1(w^\delta).$$

A função risco de  $T$  dado  $\underline{x}$  expressa por:  $\lambda(t | \underline{x}) = \frac{f(t | \underline{x})}{S(t | \underline{x})} = \left[ -\frac{\partial}{\partial t} S(t | \underline{x}) \right] / S(t | \underline{x})$ , pode ser reescrita como:

$$\lambda(t | \underline{x}) = \left[ -\frac{\partial}{\partial t} S_{\alpha} \left( \frac{t}{\exp(\underline{x} \underline{\beta})} \right) \right] / S_{\alpha} \left( \frac{t}{\exp(\underline{x} \underline{\beta})} \right) = \frac{f_{\alpha} (t e^{-\underline{x} \underline{\beta}}) e^{-\underline{x} \underline{\beta}}}{S_{\alpha} (t e^{-\underline{x} \underline{\beta}})}.$$

Assim,  $\lambda(t | \underline{x}) = \lambda_{\alpha} (t e^{-\underline{x} \underline{\beta}}) e^{-\underline{x} \underline{\beta}} = \lambda_{\alpha} (t') e^{-\underline{x} \underline{\beta}}$ , onde  $\lambda_{\alpha}(\cdot)$  é a função risco de  $T'$ ,

uma vez que  $t' = t e^{-\underline{x} \underline{\beta}}$ . Portanto a função risco de  $T$  dado  $\underline{x}$  é dada por (2.15).  $\square$

Pode-se portanto observar que o efeito das variáveis concomitantes  $\underline{x}$  em (2.15) é atuar multiplicativamente na função risco de  $T | \underline{x}$ , ou ainda aditivamente em  $Y = \log(T | \underline{x})$ .

**Exemplo:** Suponha que foi assumido aos tempos de vida  $T$  dado as variáveis concomitantes  $\underline{x}$ , a distribuição Weibull com parâmetro de escala  $\alpha(\underline{x}, \underline{\beta}) = \exp\{\underline{x} \underline{\beta}\}$  e parâmetro de forma  $\delta$ . A função de densidade de  $T$  dado  $\underline{x}$  é expressa por:

$$f(t | \underline{x}) = \frac{\delta}{\exp\{\underline{x} \underline{\beta}\}} \left( \frac{t}{\exp\{\underline{x} \underline{\beta}\}} \right)^{\delta - 1} \exp \left\{ - \left( \frac{t}{\exp\{\underline{x} \underline{\beta}\}} \right)^{\delta} \right\}, \quad t \geq 0. \quad (2.17)$$

**TEOREMA:** Aplicando  $\log$  aos tempos de vida  $T$  dado as variáveis concomitantes  $\underline{x}$ , a distribuição de  $Y = \log(T | \underline{x})$  é Valor-Extremo com parâmetro de locação  $\mu(\underline{x}, \underline{\beta})$  e parâmetro de escala  $\sigma$ , onde  $\sigma = \frac{1}{\delta}$  e  $\mu(\underline{x}, \underline{\beta}) = \log \alpha(\underline{x}, \underline{\beta}) = \underline{x} \underline{\beta}$ , como é demonstrado a seguir.

Como,  $F_Y(y) = P[Y \leq y] = P[\log T \leq y] = P[T \leq e^y] = F_T(e^y)$ , a função de densidade de  $Y = \log(T | \underline{x})$  é encontrada ser:



$$\begin{aligned}
 f_Y(y) &= \frac{\partial}{\partial y} F_T(e^y) = f_T(e^y) e^y = \frac{\delta}{\exp\{\underline{x}\beta\}} \left\{ \left( \frac{e^y}{\exp\{\underline{x}\beta\}} \right)^\delta - 1 \right\} \exp \left\{ - \left( \frac{e^y}{\exp\{\underline{x}\beta\}} \right)^\delta \right\} e^y \\
 &= \frac{\delta e^{\delta y}}{(\exp\{\underline{x}\beta\})^\delta} \exp \left\{ - \frac{e^{\delta y}}{(\exp\{\underline{x}\beta\})^\delta} \right\} = \\
 &= \delta \exp \left\{ \delta (y - \underline{x}\beta) - \exp[\delta(y - \underline{x}\beta)] \right\}. \quad \text{Fazendo } \delta = 1/\sigma \text{ obtém-se:}
 \end{aligned}$$

$$f_Y(y) = \frac{1}{\sigma} \exp \left\{ \left( \frac{y - \underline{x}\beta}{\sigma} \right) - \exp \left( \frac{y - \underline{x}\beta}{\sigma} \right) \right\} \quad \text{para } -\infty < y < \infty,$$

a qual é a densidade de uma variável aleatória Valor-Extremo com parâmetro de locação  $\mu(\underline{x}, \beta) = \underline{x}\beta$  e parâmetro de escala  $\sigma$ . Logo, o modelo de regressão para  $Y$  dado  $\underline{x}$ , pode ser escrito por:

$$Y = \underline{x}\beta + \sigma W, \quad (2.18)$$

desde que,  $W = \frac{y - \underline{x}\beta}{\sigma}$  tem distribuição Valor-Extremo padrão cuja função de densidade é  $f_W(w) = \exp \{ w - \exp(w) \}$  para  $-\infty < w < \infty$ . Pela expressão (2.18), verifica-se que  $\underline{x}$  atua aditivamente em  $Y = \log(T | \underline{x})$ . Alternativamente, o modelo (2.18) pode ser escrito por:

$$E[Y] = \underline{x}\beta + \sigma E[W], \quad (2.19)$$

onde  $E[W]$  é a esperança da variável aleatória  $W$  que tem distribuição Valor-Extremo padrão.

Para verificar que  $\underline{x}$  atua multiplicativamente na função risco de  $T | \underline{x}$ , tem-se em termos da função risco que:

$$\lambda(t | \underline{x}) = \lambda_0 \left( t e^{-\underline{x}\beta} \right) e^{-\underline{x}\beta}, \quad (2.20)$$

onde  $\lambda_0(\cdot)$  é a função risco de  $T' = \exp\{\sigma W\}$ , isto é,

$$\lambda_0(t') = \lambda_0 \left( t \exp\{-\underline{x}\beta\} \right) = \delta (t')^{\delta - 1}.$$

De fato, a função risco básica de  $T' = \exp\{\sigma W\}$  é dada por:

$$\lambda_o(t') = \frac{f_o(t')}{S_o(t')} = \left[ -\frac{\partial}{\partial t'} S_o(t') \right] / S_o(t').$$

Substituindo na expressão acima, a função de sobrevivência básica  $S_o(t') = 1 - F_o(t')$ , onde a função acumulada de  $T'$  é:

$$\begin{aligned} F_o(t') &= P[T' \leq t'] = P[e^{\sigma W} \leq t'] = P[W \leq \log t'/\sigma] = \\ &= 1 - \exp\left\{-\exp\left(\frac{\log t'}{\sigma}\right)\right\} = 1 - \exp\left\{-\left(\exp \log t'\right)^{1/\sigma}\right\} = \\ &= (\text{como } 1/\sigma = \delta) = 1 - \exp\left\{-\left(t'\right)^\delta\right\}, \end{aligned}$$

segue que:  $\lambda_o(t') = \delta (t')^{\delta-1}$  para  $t' \geq 0$  onde  $t' = t \exp\{-x \hat{\beta}\}$ . Portanto,  $x$  atua multiplicativamente na função risco de  $T|x$ .

Ainda, se  $x \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 d$ , doses virtualmente seguras  $d_o = d_o(t)$ , em um fixado tempo  $t$  podem ser encontradas por:

$$\Pi(t | d_o) = \pi,$$

onde, dependendo do objetivo do experimento (ver pg. 27), tem-se:

$$\begin{aligned} \Pi(t | d) = F(t | d) - F(t | 0) &= \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\} - \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}}\right)^{\hat{\delta}}\right\} \\ \text{com } 0 < \pi < \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\}, \end{aligned}$$

ou,

$$\begin{aligned} \Pi(t | d) = F(t | 0) - F(t | d) &= \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}}\right)^{\hat{\delta}}\right\} - \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\} \\ \text{com } 0 < \pi < 1 - \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\}. \end{aligned}$$

O tempo médio para a resposta até um especificado tempo  $t$  ou  $y = \log t$  em uma dose  $d$ , pode também ser de interesse. Estes são obtidos, respectivamente, para  $Y = \log T$  e para  $T$  por :

$$t_{mR} = E(Y_o | d) = \int_0^y \exp \left\{ - \exp \left( \frac{u - (\hat{\beta}_0 + \hat{\beta}_1 d)}{\hat{\sigma}} \right) \right\} du$$

$$e \quad t_{mR} = E(T_o | d) = \int_0^t \exp \left\{ - \left( \frac{u}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}} \right)^{\hat{\delta}} \right\} du,$$

onde  $Y_o = \min(Y, y)$  e  $T_o = \min(T, t)$ .

### 2.3.2 - Modelos de Riscos Proporcionais

Uma família de riscos proporcionais é uma classe de modelos com a propriedade de que diferentes indivíduos têm funções riscos proporcionais, isto é,

a razão  $\frac{\lambda(t | \underline{x}_1)}{\lambda(t | \underline{x}_2)}$  das funções risco para dois indivíduos com vetores de variáveis concomitantes  $\underline{x}_1$  e  $\underline{x}_2$  não varia com o tempo. Quando isto ocorre, a função risco de  $T$  dado  $\underline{x}$ , pode ser escrita na forma :

$$\lambda(t | \underline{x}) = \lambda_o(t) g(\underline{x}, \underline{\beta}), \quad (2.21)$$

onde  $\lambda_o(t)$  é uma função risco básica e é a função risco para um indivíduo quando  $g(\underline{x}, \underline{\beta}) = 1$ . Ainda,  $g(\underline{x}, \underline{\beta})$  é uma função que depende das variáveis concomitantes  $\underline{x}$  e dos parâmetros desconhecidos  $\underline{\beta}$ .

Obs: Em geral usa-se as variáveis concomitantes  $\underline{x}$  transformadas e centradas.

A escolha de  $g(\underline{x}, \underline{\beta})$  pode depender dos particulares dados sendo considerados e dentre as várias escolhas, a mais natural é :

$$g(\underline{x}, \underline{\beta}) = \exp \{ \underline{x} \underline{\beta} \}, \quad (2.22)$$

a qual possui a vantagem de tomar somente valores positivos e assim não exige que nenhuma restrição seja feita com relação aos valores de  $\underline{\beta}$  para garantir que  $g(\underline{x}, \underline{\beta}) > 0$  para todo  $\underline{x}$ . Sendo  $g(\underline{x}, \underline{\beta}) = \exp\{ \underline{x} \underline{\beta} \}$  sempre positiva,

$\lambda(t | \underline{x})$  é automaticamente não negativa para todo  $\underline{x}$  e  $\underline{\beta}$ . Obviamente, outras formas funcionais podem ser utilizadas para  $g(\underline{x}, \underline{\beta})$  e podem ser mais apropriadas em situações específicas.

Modelos de riscos proporcionais assumem portanto que as variáveis concomitantes têm um efeito multiplicativo na função risco.

Na família de modelos (2.21) a função de sobrevivência de T dado o vetor de variáveis concomitantes  $\underline{x}$  é expressa por:

$$\begin{aligned} S(t | \underline{x}) &= \exp \left\{ - \int_0^t \lambda(u | \underline{x}) du \right\} = \exp \left\{ - \int_0^t \lambda_o(u) g(\underline{x}, \underline{\beta}) du \right\} = \\ &= \exp \left\{ -g(\underline{x}, \underline{\beta}) \int_0^t \lambda_o(u) du \right\} = \left\{ \exp \left[ - \int_0^t \lambda_o(u) du \right] \right\}^{g(\underline{x}, \underline{\beta})}. \end{aligned}$$

Então:  $S(t | \underline{x}) = [S_o(t)]^{g(\underline{x}, \underline{\beta})}$ , onde  $S_o(t) = \exp \left[ - \int_0^t \lambda_o(u) du \right]$  é chamada função de sobrevivência básica para um indivíduo quando  $g(\underline{x}, \underline{\beta}) = 1$ .

Existe então, em uma família de riscos proporcionais, uma ordenação das funções de sobrevivência para dois indivíduos com vetores de variáveis concomitantes  $\underline{x}_1$  e  $\underline{x}_2$ , isto é, ou  $S(t | \underline{x}_1) \geq S(t | \underline{x}_2)$  para todo  $t$  ou vice-versa.

**Exemplo:** Considerando o mesmo modelo tomado anteriormente, ou seja, o modelo Weibull de dois parâmetros cuja função de densidade de probabilidade (f.d.p.) é dada por:

$$f(t | \underline{x}) = \frac{\delta}{\alpha(\underline{x}, \underline{\beta})} \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^{\delta - 1} \exp \left\{ - \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^{\delta} \right\}, \quad t \geq 0,$$

onde  $\alpha(\underline{x}, \underline{\beta})$  é claramente uma função das variáveis concomitantes  $\underline{x}$ , a sua função de sobrevivência é expressa por:

$$S(t | \underline{x}) = 1 - F(t | \underline{x}) = \exp \left\{ - \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^{\delta} \right\}, \quad t \geq 0.$$

Então, a função risco de T dado o vetor de variáveis concomitantes  $\underline{x}$  é:

$$\lambda(t | \underline{x}) = \frac{f(t | \underline{x})}{S(t | \underline{x})} = \frac{\delta}{\alpha(\underline{x}, \underline{\beta})} \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^{\delta - 1}.$$

Assim, a razão das funções risco para dois indivíduos com vetores de variáveis concomitantes  $\underline{x}_1$  e  $\underline{x}_2$  é:

$$\frac{\lambda(t | \underline{x}_1)}{\lambda(t | \underline{x}_2)} = \frac{\frac{\delta}{\alpha(\underline{x}_1, \underline{\beta})}}{\frac{\delta}{\alpha(\underline{x}_2, \underline{\beta})}} \frac{\left(\frac{t}{\alpha(\underline{x}_1, \underline{\beta})}\right)^{\delta-1}}{\left(\frac{t}{\alpha(\underline{x}_2, \underline{\beta})}\right)^{\delta-1}} = \left[\frac{\alpha(\underline{x}_2, \underline{\beta})}{\alpha(\underline{x}_1, \underline{\beta})}\right]^{\delta}$$

a qual não depende de t. Em outras palavras, indivíduos diferentes têm funções risco proporcionais. Consequentemente, o modelo Weibull é um modelo de riscos proporcionais.

Ainda, sendo o modelo Weibull um modelo de riscos proporcionais, a função risco de T dado o vetor de variáveis concomitantes  $\underline{x}$ , é da forma (2.21). De fato, como foi visto,

$$\begin{aligned} \lambda(t | \underline{x}) &= \frac{\delta}{\alpha(\underline{x}, \underline{\beta})} \left(\frac{t}{\alpha(\underline{x}, \underline{\beta})}\right)^{\delta-1} = \delta t^{\delta-1} \left[\frac{1}{\alpha(\underline{x}, \underline{\beta})}\right]^{\delta} = \\ &= \left(\text{tomando } \alpha(\underline{x}, \underline{\beta}) = \exp\{\underline{x}\underline{\beta}\}\right) = \\ &= \left(\delta t^{\delta-1}\right) \exp\{-\underline{x}\underline{\beta}\delta\} = \left(\text{fazendo } \underline{\beta}\delta = \underline{\beta}^*\right) = \\ &= \delta t^{\delta-1} \exp\{-\underline{x}\underline{\beta}^*\}, \end{aligned}$$

onde:  $\lambda_0(t) = \delta t^{\delta-1}$  e  $g(\underline{x}, \underline{\beta}) = \exp\{\underline{x}\underline{\beta}^*\}$ .

Se  $\underline{x}\hat{\underline{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 d$ , doses virtualmente seguras  $d_0 = d_0(t)$  em um fixado tempo t podem ser encontradas por:

$$\Pi(t | d_0) = \pi,$$

onde, dependendo do objetivo do experimento (ver pg. 27), tem-se:

$$\begin{aligned} \Pi(t | d) = F(t | d) - F(t | 0) &= \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\} - \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}}\right)^{\hat{\delta}}\right\} \\ &\text{com } 0 < \pi < \exp\left\{-\left(\frac{t}{\exp\{\hat{\beta}_0\}}\right)^{\hat{\delta}}\right\}, \end{aligned}$$

ou,

$$\Pi(t | d) = F(t | 0) - F(t | d) \quad \text{com} \quad 0 < \pi < F(t | 0).$$

Ainda, o tempo médio para a resposta até um especificado tempo  $t$  em uma dose  $d$ , pode ser obtido por:

$$t_{mR} = E(T_o | d) = \int_0^t \exp\left\{-\left(\frac{u}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}}\right)^{\hat{\delta}}\right\} du,$$

onde:  $T_o = \min(T, t)$ .

#### OBSERVAÇÕES:

1 - O modelo Weibull considerado aqui, tem como parâmetro de escala  $\alpha(\underline{x}, \underline{\beta})$  o qual depende das variáveis concomitantes  $\underline{x}$  e como parâmetro de forma  $\delta$  o qual não depende das variáveis concomitantes. É conveniente notar que se ambos os parâmetros dependessem das variáveis concomitantes  $\underline{x}$ , o modelo não seria de riscos proporcionais, mas continuaria sendo um modelo log-linear.

2 - Kalbfleish e Prentice (1980) mostraram que "o único" modelo que pertence tanto a classe de modelos log-linear quanto a classe de modelos de riscos proporcionais é o modelo Weibull de dois parâmetros considerado nos exemplos. Deve-se atentar ao fato de que o modelo Exponencial é um caso especial do modelo Weibull.

A família de modelos de riscos proporcionais (2.21) é portanto essencialmente distinta da família de modelos log-linear (2.12). Os modelos de regressão log-linear Weibull (e particularmente o Exponencial) são então os únicos modelos log-lineares que também são de riscos proporcionais.

É interessante notar o efeito das variáveis concomitantes  $\underline{x}$  nas duas famílias de modelos: no modelo log-linear  $Y = \underline{x}\underline{\beta} + \sigma W$ ,  $\underline{x}$  atua multiplicativamente em  $T$ , enquanto que no modelo de riscos proporcionais  $\lambda(t | \underline{x}) = \lambda_o(t) \exp\{\underline{x}\underline{\beta}\}$ ,  $\underline{x}$  atua multiplicativamente na função risco para  $T$ .

### 2.3.3 - Modelo de Riscos Proporcionais de Cox

Já foi visto na seção 2.3.2 que nos modelos de riscos proporcionais é assumido uma forma paramétrica aos tempos  $T$  dado as variáveis concomitantes  $\underline{x}$ . O modelo de riscos proporcionais de Cox não assume uma forma à distribuição dos tempos  $T$  dado as variáveis concomitantes  $\underline{x}$ . Assim, se  $\underline{x} = (x_1, x_2, \dots, x_j)$  denota as variáveis concomitantes (possivelmente transformadas e centradas) e  $\lambda_0(t)$  denota a função risco da desconhecida distribuição dos tempos em  $x_1 = x_2 = \dots = x_j = 0$ , o modelo de riscos proporcionais de Cox para a função risco nos possíveis valores das variáveis concomitantes  $x_1, x_2, \dots, x_j$  é dado por :

$$\lambda(t | \underline{x}) = \lambda_0(t) \exp \{ \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j \} = \lambda_0(t) \exp \{ \sum_i x_i \beta_i \}, \quad (2.23)$$

onde:  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_j)'$  é um vetor de coeficientes de regressão e

$\lambda_0(t)$  é uma função risco básica arbitrária não especificada para  $T$  contínuo.

Note que o modelo de Cox (2.23) possui uma similaridade com a relação log-linear dada por  $\log T = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j$  sendo que o modelo de Cox não possui o coeficiente intercepto, pois  $\lambda_0(t)$  ocupa em (2.23) o lugar de  $\beta_0$ .

O modelo de riscos proporcionais de Cox é amplamente utilizado em aplicações biomédicas, possivelmente pelo aspecto atrativo de não assumir uma forma à distribuição dos tempos dado as variáveis concomitantes. No caso particular de experimentos de dose-resposta, o modelo de riscos proporcionais de Cox pode ser usado para extrapolação da dose mas não do tempo, pelo fato do mesmo ser livre de distribuição (semi-paramétrico). Ele não pode ser usado para extrapolar a distribuição em tempos anteriores ao limite inferior ou em tempos posteriores ao limite superior dos tempos observados. Assim, o modelo de Cox é útil somente para estimar a distribuição de vida na extensão observada.

A função de densidade condicional de T dado o vetor de variáveis concomitantes  $\underline{x}$ , correspondente ao modelo (2.23) é:

$$\begin{aligned}
 f(t | \underline{x}) &= \lambda(t | \underline{x}) S(t | \underline{x}) = \\
 &= \lambda(t | \underline{x}) \exp \left\{ - \int_0^t \lambda(u | \underline{x}) du \right\} = \\
 &= \lambda_o(t) \exp\{ \underline{x} \underline{\beta} \} \exp \left\{ - \int_0^t \lambda_o(u) \exp(\underline{x} \underline{\beta}) du \right\} = \\
 &= \lambda_o(t) \exp\{ \underline{x} \underline{\beta} \} \left\{ \exp \left( - \int_0^t \lambda_o(u) du \right) \right\}^{\exp\{ \underline{x} \underline{\beta} \}}
 \end{aligned} \tag{2.24}$$

A função de sobrevivência de T dado o vetor  $\underline{x}$  é expressa por:

$$S(t | \underline{x}) = \left\{ \exp \left( - \int_0^t \lambda_o(u) du \right) \right\}^{\exp\{ \underline{x} \underline{\beta} \}} = [S_o(t)]^{\exp\{ \underline{x} \underline{\beta} \}} \tag{2.25}$$

onde:  $S_o(t) = \exp \left( - \int_0^t \lambda_o(u) du \right)$  é chamada função de sobrevivência básica.

Um exemplo do comportamento de  $S_o(t)$  e  $S(t | \underline{x})$  é ilustrado na figura abaixo.

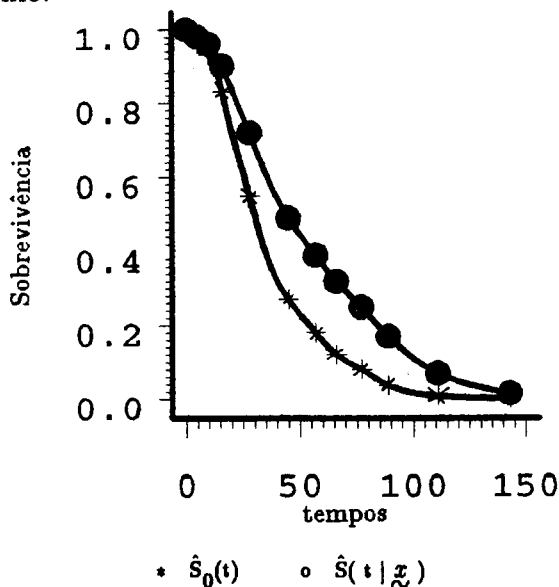


Figura 2.4 - Possível comportamento das funções  $S_o(t)$  e  $S(t | \underline{x})$ .



## 2.4 - INVESTIGAÇÃO DE UM MODELO

Um primeiro problema encontrado ao trabalhar-se com variáveis concomitantes, é aquele de estabelecer a natureza da relação entre os tempos de vida  $T$  e essas variáveis concomitantes  $x_j$ . É desejado saber qual o modelo que deve ser assumido: log-linear, de riscos proporcionais, de riscos proporcionais de Cox ou ainda um outro que não os considerados aqui.

Em situações simples, onde uma única variável concomitante está envolvida, como é o caso tratado aqui, onde a variável concomitante envolvida é a dose, gráficos do tempo  $t$  contra dose ou log dos tempos contra a dose fornecem uma boa visualização da relação existente entre os tempos e a dose. O conhecimento do particular experimento envolvido frequentemente pode sugerir tal relação.

Se um modelo log-linear é assumido, gráficos de probabilidade ou outros gráficos dos dados em  $x_j$  podem ser usados para examinar particulares modelos para  $W$ . Mais geralmente, os dados podem ser particionados de acordo com os níveis da dose e modelos ajustados para essas diferentes partes podem então ser examinados para a constância de  $\sigma$  e de outros parâmetros e também para específicas distribuições do erro.

A figura 2.5 abaixo ilustra uma situação onde é possível supor a existência de uma relação linear entre o log dos tempos e a dose.

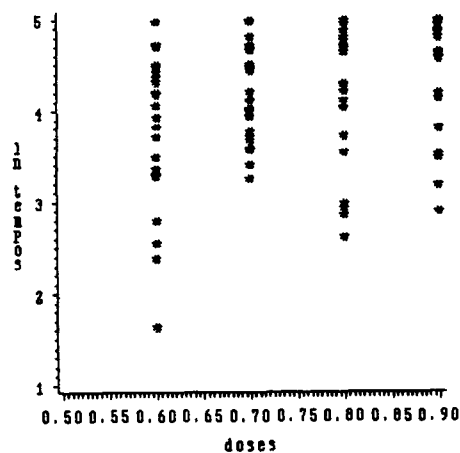


Figura 2.5 - Ln dos tempos versus dose.

Se um modelo de riscos proporcionais é assumido, uma verificação desta suposição pode ser feita calculando-se, através dos dados, as estimativas produto-limite da função de sobrevivência  $\hat{S}_i(t) = \hat{S}(t | x_i)$  em cada  $x_i$  e então gráficos do  $\log(-\log \hat{S}_i(t))$  contra  $t$  devem ser aproximadamente paralelos caso o modelo de riscos proporcionais seja apropriado. Ainda gráficos do  $\log(-\log \hat{S}_i(t))$  versus  $\log t$  devem ser aproximadamente lineares e paralelos se um modelo de riscos proporcionais Weibull é apropriado.

Se os dados são particionados de acordo com os  $K$  níveis da dose e se modelos de riscos proporcionais são supostos para essas partições dos dados, gráficos do  $\log(-\log \hat{S}_{o_j}(t))$  contra  $t$  ( $j = 1, 2, \dots, k$ ) devem ser aproximadamente paralelos para essas partições, onde  $\hat{S}_{o_j}(t) = \hat{S}(t | x_{o_j})$  é a estimativa produto-limite da função de sobrevivência para a partição  $j$ .

As figuras 2.6 e 2.7 ilustram situações nas quais não existem evidências de violação da suposição de que os riscos são proporcionais, uma vez que as curvas são aproximadamente paralelas.

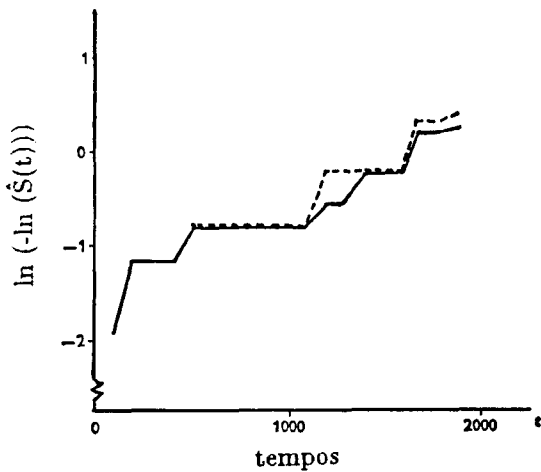


Figura 2.6 - tempo  $t$  versus  $\ln(-\ln \hat{S}(t))$   
para dois níveis de uma  
dada variável concomitante.

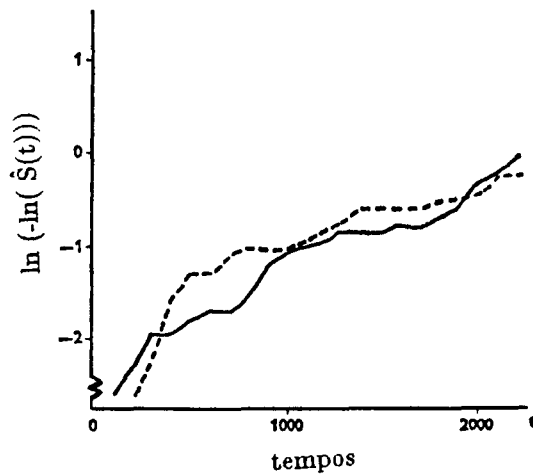


Figura 2.7 - tempo  $t$  versus  $\ln(-\ln \hat{S}(t))$   
para dois níveis de uma  
dada variável concomitante.

Ainda, a figura 2.8 ilustra uma situação na qual há evidências de que o modelo de riscos proporcionais Weibull é apropriado.

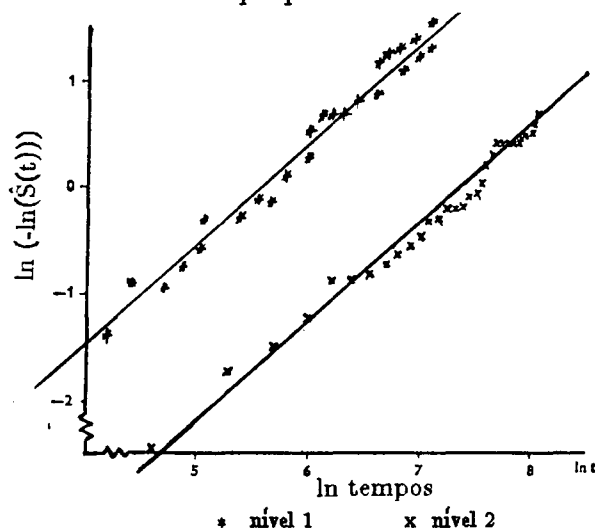


Figura 2.8 -  $\ln$  dos tempos  $t$  versus  $\ln(-\ln \hat{S}(t))$  para dois níveis da variável concomitante dose.

## 2.5 - ANÁLISE DE RESÍDUOS

Uma vez escolhido o modelo e realizado a análise baseado no modelo escolhido, o exame de resíduos, especialmente por técnicas gráficas, é uma importante maneira de verificar o ajuste do modelo. Os resíduos servem propósitos diversos:

- detectar a presença de outliers;
- relevância de um fator adicional, omitido do modelo, detectado pelo plot dos resíduos versus níveis do fator;
- não constância da variância, detectada pelo plot resíduos versus fatores pensados afetar a variância; entre outros.

### 2.5.1 - Caso dos modelos paramétricos

O conceito de resíduos para os modelos paramétricos de riscos proporcionais e log-linear é considerado aqui.

Suponha que  $Y_i = \log T_i$  é a variável resposta e  $\underline{x}_i$  é um vetor associado de variáveis concomitantes. Considere que o vetor observado da variável aleatória  $Y$  possa ser expresso por um modelo em termos de um vetor de

parâmetros desconhecidos  $\underline{\beta}$ , do vetor de variáveis concomitantes  $\underline{x}$  e de um vetor  $\underline{\epsilon}$  de variáveis aleatórias i.i.d. (independentes e identicamente distribuídas) não observadas. Mais particularmente, assuma que cada observação  $Y_i$  depende somente de um dos  $\epsilon_i$ ,  $i = 1, 2, \dots, n$  de modo que cada  $Y_i$  possa ser escrito por :

$$Y_i = h_i(\underline{\beta}, \underline{x}_i, \epsilon_i) \quad i = 1, 2, \dots, n. \quad (2.26)$$

Para definir-se resíduos para (2.26), considere  $\hat{\underline{\beta}}$  o estimador de máxima verossimilhança (e.m.v.) de  $\underline{\beta}$  e suponha que a equação  $Y_i = h_i(\hat{\underline{\beta}}, \underline{x}_i, \epsilon_i)$  tem uma única solução dada por :

$$\hat{\epsilon}_i = g_i(y_i, \hat{\underline{\beta}}, \underline{x}_i) \quad i = 1, 2, \dots, n. \quad (2.27)$$

Consequentemente,

$$\epsilon_i = g_i(y_i, \underline{\beta}, \underline{x}_i) \quad i = 1, 2, \dots, n. \quad (2.28)$$

Toma-se (2.27) como definição de resíduos para o modelo (2.26). Essa definição de resíduos é devida a Cox e Snell (1968) que em seu artigo mostraram que um resíduo  $\hat{\epsilon}_i$  difere do erro  $\epsilon_i$  por uma quantidade de ordem  $n^{-1/2}$  em probabilidade, bem como examinaram as propriedades estatísticas dos  $\hat{\epsilon}_i$  ( $i = 1, 2, \dots, n$ ) as quais diferem pouco das propriedades dos erros  $\epsilon_i$  ( $i = 1, 2, \dots, n$ ). Devido a isso, os resíduos  $\hat{\epsilon}_i$  ( $i = 1, 2, \dots, n$ ) são frequentemente considerados como comportando-se aproximadamente como uma amostra aleatória de tamanho  $n$  da distribuição dos  $\epsilon_i$  ( $i = 1, 2, \dots, n$ ), e embora eles não sejam independentes e em geral não são identicamente distribuídos, existe pequeno prejuízo em tratá-los como tal em análises gráficas.

Se o modelo ajustado é representado por um modelo log-linear, isto é, se  $Y = \log T | \underline{x}$ , pode ser escrito na forma :

$$Y = \mu(\underline{x}, \underline{\beta}) + \sigma \epsilon,$$

onde  $\sigma > 0$ ,  $\mu(\underline{x}, \underline{\beta}) = \underline{x} \underline{\beta}$  e  $\epsilon$  tem distribuição conhecida e independente de  $\underline{x}$ , os resíduos são naturalmente definidos por :

$$\hat{\epsilon}_i = g_i(y_i, x_i, \hat{\beta}) = \frac{(y_i - x_i \hat{\beta})}{\hat{\sigma}} \quad i = 1, 2, \dots, n, \quad (2.29)$$

onde  $\hat{\beta}$  e  $\hat{\sigma}$  são os estimadores de máxima verossimilhança (e.m.v) de  $\beta$  e  $\sigma$ , respectivamente.

Considerando :

$$T = \exp Y = \exp \{ x \beta + \sigma \epsilon \},$$

tem-se que:

$$\exp \{ \hat{\epsilon}_i \} = [ t_i \exp \{ - x_i \hat{\beta} \} ]^{1/\hat{\sigma}} \quad i = 1, 2, \dots, n. \quad (2.30)$$

Com  $\exp \{ \hat{\epsilon}_i \} = \hat{\epsilon}_i^*$  e  $1/\hat{\sigma} = \hat{\delta}$ , a expressão é reescrita como:

$$\hat{\epsilon}_i^* = [ t_i \exp \{ - x_i \hat{\beta} \} ]^{\hat{\delta}}, \quad i = 1, 2, \dots, n, \quad (2.31)$$

a qual define os resíduos para o modelo dado para  $T|x$ .

Claramente vê-se que os  $\hat{\epsilon}_i$  ( $i = 1, 2, \dots, n$ ) estão relacionados com os  $\hat{\epsilon}_i^*$  ( $i = 1, 2, \dots, n$ ) por:

$$\hat{\epsilon}_i = \log \hat{\epsilon}_i^*, \quad i = 1, 2, \dots, n. \quad (2.32)$$

Os resíduos  $\hat{\epsilon}_i$  são vistos como uma amostra aleatória proveniente da distribuição dos  $\epsilon_i$  e os resíduos  $\hat{\epsilon}_i^*$  como uma amostra aleatória proveniente da distribuição dos  $\exp\{\epsilon_i\}$ ,  $i = 1, 2, \dots, n$ .

**TEOREMA:** Nos casos particulares onde aos tempos  $T$  dado o vetor de variáveis concomitantes  $x$  assume-se a distribuição Weibull ou Exponencial,

as quantidades  $\hat{\epsilon}_i^* = [ t_i \exp \{ - x_i \hat{\beta} \} ]^{\hat{\delta}}$  são os riscos acumulados estimados, isto é,

$$\hat{\epsilon}_i^* = \hat{H}(t_i | x_i), \quad (2.33)$$

para  $i = 1, 2, \dots, n$ .

*Demonstração :*

Assumindo que  $T \sim Weibull(\alpha(\underline{x}, \underline{\beta}); \delta)$  então :

$$f(t | \underline{x}) = \frac{\delta}{\alpha(\underline{x}, \underline{\beta})} \left[ \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right]^{\delta - 1} \exp \left\{ - \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^\delta \right\}; \quad S(t | \underline{x}) = \exp \left\{ - \left( \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right)^\delta \right\}$$

•  $\lambda(t | \underline{x}) = \frac{\delta}{\alpha(\underline{x}, \underline{\beta})} \left[ \frac{t}{\alpha(\underline{x}, \underline{\beta})} \right]^{\delta - 1}$ . Tomando-se  $\alpha(\underline{x}, \underline{\beta}) = \exp \{ \underline{x} \underline{\beta} \}$  tem-se:

$$H(t | \underline{x}) = \int_0^t \delta u^{\delta - 1} \left( \exp \{ - \underline{x} \underline{\beta} \} \right)^\delta du = \left[ t \exp \{ - \underline{x} \underline{\beta} \} \right]^\delta. \quad \text{Logo,}$$

$$\hat{e}_i^* = \left[ t_i \exp \{ - \underline{x}_i \hat{\underline{\beta}} \} \right]^\delta = \hat{H}(t_i | \underline{x}_i), \quad i = 1, 2, \dots, n.$$

Como a Exponencial é um caso particular da Weibull quando  $\delta = 1$ , segue que

$$\hat{e}_i^* = t_i \exp \{ - \underline{x}_i \hat{\underline{\beta}} \} = \hat{H}(t_i | \underline{x}_i), \quad i = 1, 2, \dots, n,$$

definem os resíduos para  $T \sim \text{Exp}(\alpha(\underline{x}, \underline{\beta}))$ .  $\square$

**TEOREMA:** Ainda os  $e_i^* = H(t_i | \underline{x}_i)$ ,  $i = 1, 2, \dots, n$  são variáveis aleatórias i.i.d. (independentes e identicamente distribuídas) Exponencial Padrão (Lawless-1980/ pg.281). Essa afirmação é devidamente demonstrada abaixo.

*Demonstração:*

Suponha que  $H(t_i | \underline{x}_i)$  é a função risco acumulada de  $T_i$  dado  $\underline{x}_i$ . Sabe-se que  $S(t_i | \underline{x}_i) = \exp \{ - H(t_i | \underline{x}_i) \}$ ,  $i = 1, 2, \dots, n$  são variáveis aleatórias i. i. d. Uniforme em  $(0,1)$ , o que é decorrente do teorema que diz : " Se  $X$  é uma variável aleatória e  $F$  é uma função de distribuição absolutamente contínua  $\Rightarrow F(X)$  é função de distribuição de uma Uniforme em  $(0,1)$  ".

Seja  $Z = S(t_i | \underline{x}_i) = \exp \{ - H(t_i | \underline{x}_i) \}$ , então :

$$W = H(t_i | \underline{x}_i) = - \log S(t_i | \underline{x}_i) = - \log Z \quad e$$

$$F_W(w) = P[W \leq w] = P[- \log Z \leq w] = P[Z \geq e^{-w}] = \begin{cases} 1 - e^{-w} & \text{para } w \geq 0 \\ 0 & \text{caso contrário,} \end{cases}$$

a qual é a função de distribuição de uma Exponencial padrão. Como a função de distribuição determina a distribuição, segue que  $W = H(t_i | \mathcal{X}_i) \sim \text{Exp}(1)$ .  $\square$

Deste modo, os resíduos  $\hat{e}_i^*$ ,  $i = 1, 2, \dots, n$  para observações não censuradas, podem ser tratados como uma amostra aleatória proveniente da distribuição Exponencial padrão.

Quando existem dados censurados, modificações são necessárias. Em tal caso é sugerido por Lawless (1980) definir os resíduos como:

$$\hat{e}_i^* = H(t_i^* | \mathcal{X}_i) + 1 \quad \text{para } i \in C, \quad (2.34)$$

onde:  $t_i^*$  refere-se ao observado tempo censurado e

$C$  = conjunto formado pelos indivíduos cujos tempos observados são censurados.

$$\text{Então :} \quad \hat{e}_i^* = \hat{H}(t_i | \mathcal{X}_i) \quad \text{se } i \in F \quad (2.35)$$

$$\hat{e}_i^* = \hat{H}(t_i^* | \mathcal{X}_i) + 1 \quad \text{se } i \in C.$$

Essa definição de resíduos é particularmente útil com modelos de riscos proporcionais e como é sabido, os modelos Weibull e Exponencial são também modelos de riscos proporcionais.

Se resíduos são baseados em (2.35), os gráficos de  $-\log(\hat{S}_1(\hat{e}^*))$  versus  $\hat{e}^*$  devem dar aproximadamente uma reta com inclinação 1 se o modelo é adequado.  $\hat{S}_1(\hat{e}^*)$  é a estimativa produto-limite da função de sobrevivência para os  $\hat{e}_i^*$ , ( $i = 1, 2, \dots, n$ ).

### 2.5.2 - Caso do modelo semi-paramétrico

Aqui será tratado a análise de resíduos no modelo semi-paramétrico de riscos proporcionais de Cox onde nenhuma forma paramétrica é assumida para  $\lambda_0(t)$ .

Segundo Lawless (1980), procedimentos similares aos tratados na seção 2.5.1 podem ser usados na definição de resíduos para o modelo de riscos proporcionais de Cox. O resíduo correspondente a um tempo de vida não censurado é então definido por:

$$\hat{e}_i = \hat{H}(t_i | \mathcal{X}_i) = \hat{H}_o(t_i) \exp \{ \mathcal{X}_i \hat{\beta} \} = [ -\log \hat{S}_o(t_i) ] \exp \{ \mathcal{X}_i \hat{\beta} \}, \quad (2.36)$$

onde  $H_o(t)$  é a função risco básica acumulada e  $\hat{H}_o(t) = -\log \hat{S}_o(t)$  é um estimador de  $H_o(t)$ .

Para estimar-se  $H_o(t)$ , é preciso estimar  $S_o(t)$  e para isso procede-se da seguinte maneira:

- determina-se  $\hat{\beta}$  pelo método da verossimilhança parcial (tratado adiante na seção 2.6)
- com os valores estimados de  $\hat{\beta}$ , estima-se  $S_o(t)$  maximizando-se a verossimilhança completa para  $S_o(t)$ , isto é, maximiza-se :

$$\begin{aligned} L(\hat{\beta}, S_o(t)) &= \prod_{i=1}^n [f(t_i | \mathcal{X}_i)]^{\delta_i} [S(t_i | \mathcal{X}_i)]^{1-\delta_i} = \\ &= \prod_{i=1}^n [\lambda(t_i | \mathcal{X}_i) S(t_i | \mathcal{X}_i)]^{\delta_i} [S(t_i | \mathcal{X}_i)]^{1-\delta_i} = \\ &= \prod_{i=1}^n \left[ \lambda_o(t_i) \exp \{ \mathcal{X}_i \hat{\beta} \} (S_o(t_i))^{exp\{\mathcal{X}_i \hat{\beta}\}} \right]^{\delta_i} \left[ S_o(t_i)^{exp\{\mathcal{X}_i \hat{\beta}\}} \right]^{1-\delta_i} \end{aligned}$$

O estimador de máxima verossimilhança para  $S_o(t)$  é encontrado ser:

$$\hat{S}_o(t) = \prod_{i: t_{(i)} \leq t} \hat{\alpha}_i \quad (2.37)$$

onde:

$$\hat{\alpha}_i = \left( 1 - \frac{\exp\{\mathcal{X}_{(i)} \hat{\beta}\}}{\sum_{l \in R_i} \exp\{\mathcal{X}_l \hat{\beta}\}} \right)^{\exp\{-\mathcal{X}_{(i)} \hat{\beta}\}}, \quad (2.38)$$

quando somente uma falha ocorre em  $t_{(i)}$ , isto é,  $d_i = 1$ . Quando os  $d_i$  ( $i = 1, 2, \dots, n$ ) são maiores que 1, existindo empates, Breslow (1974) sugere :



$$\hat{\alpha}_i = \exp\left(\frac{-d_i}{\sum_{l \in R_i} \exp\{x_l \hat{\beta}\}}\right). \quad (2.39)$$

Assim, 
$$\hat{H}_o(t) = -\log \hat{S}_o(t) = \sum_{i: t(i) \leq t} \left(\frac{d_i}{\sum_{l \in R_i} \exp\{x_l \hat{\beta}\}}\right). \quad (2.40)$$

O resíduo para um tempo de vida censurado  $t_i^*$  é definido (Lawless-1980) por :

$$\hat{e}_i = \hat{H}(t_i^* | x_i) + 1,$$

ou seja, 
$$\hat{e}_i = [-\log \hat{S}_o(t_i^*)] \exp\{x_i \hat{\beta}\} + 1. \quad (2.41)$$

Como já foi mostrado anteriormente, as quantidades  $H(t_i | x_i)$   $i = 1, 2, \dots, n$  são variáveis aleatórias i. i. d. Exponencial padrão. Então, o conjunto completo dos resíduos  $\hat{e}_i$ ,  $i = 1, 2, \dots, n$  correspondendo as observações censuradas e não censuradas podem ser vistos grosseiramente como uma amostra aleatória proveniente de uma distribuição Exponencial padrão. Assim, pode-se estimar a função de sobrevivência do conjunto de resíduos censurados e não censurados pelo produto limite dado por Kaplan-Meier. A estimativa resultante  $\hat{S}(\hat{e})$  deve ser concordante com uma distribuição Exponencial padrão e por exemplo, um plot de  $-\log \hat{S}(\hat{e})$  versus  $\hat{e}$  deve ser aproximadamente linear com inclinação 1.

**OBS:**

- 1) Se um modelo de riscos proporcionais de Cox é ajustado pode-se investigar se algum modelo de riscos proporcionais paramétrico pode ser apropriado e para isso é examinado a estimada função de sobrevivência básica  $\hat{S}_o(t)$ . Por exemplo, o plot de  $\log(-\log \hat{S}_o(t))$  versus  $\log t$  sendo aproximadamente linear, sugere um modelo WEIBULL com  $\lambda_o(t) = \delta \alpha t^{\delta - 1}$ .
- 2) Se um modelo de riscos proporcionais específico (paramétrico) é adequado, o mesmo deve preferencialmente ser usado, pois caso contrário existirá alguma perda de eficiência em usar o modelo semi-paramétrico ao invés do modelo paramétrico.

2.6 - ESTIMAÇÃO DOS COEFICIENTES  $\beta$  POR MÁXIMA VEROSSIMILHANÇA

Suponha que a cada indivíduo esteja associado um tempo de vida ou um tempo de censura  $t_i$  e um vetor de variáveis concomitantes  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ . A notação  $i \in F$  e  $i \in C$  será usada para referenciar um indivíduo  $i$  para o qual seu tempo corresponda a uma falha ou a uma censura, respectivamente.

Se os modelos paramétricos log-linear ou de riscos proporcionais estão sendo supostos, então suas respectivas funções de densidade de probabilidade e funções de sobrevivência são conhecidas, uma vez que está sendo empregada uma família paramétrica aos tempos  $T$  dado as variáveis concomitantes  $\mathbf{x}$ . Assim, a função de verossimilhança para uma amostra censurada baseada em  $n$  indivíduos é:

⇒ Para o modelo log-linear

$$L(\beta) = \prod_{i \in F} f(y | \mathbf{x}) \prod_{i \in C} S(y | \mathbf{x}), \quad \text{onde } y = \log t. \quad (2.42)$$

⇒ Para o modelo de riscos proporcionais

$$L(\beta) = \prod_{i \in F} f(t | \mathbf{x}) \prod_{i \in C} S(t | \mathbf{x}). \quad (2.43)$$

OBS: Claramente, se não houver censuras, as funções de verossimilhança (2.42) e (2.43) reduzem-se a:

$$L(\beta) = \prod_{i=1}^n f(y | \mathbf{x}) \quad \text{e} \quad L(\beta) = \prod_{i=1}^n f(t | \mathbf{x}),$$

respectivamente.

As equações de máxima verossimilhança,

$$\frac{\partial \log L(\beta)}{\partial \beta_r} = 0 \quad (r = 1, 2, \dots, p), \quad (2.44)$$

são então prontamente resolvidas pelo método iterativo de Newton - Raphson ou algum outro método iterativo para obter-se os estimadores de máxima verossimilhança (e.m.v.) de  $\beta$ , isto é,  $\hat{\beta}$ .

A matriz de informação observada p x p é:

$$I_0 = \left( -\frac{\partial^2 \log L(\beta)}{\partial \beta_s \partial \beta_r} \right) \Big|_{\beta = \hat{\beta}} \quad r, s = 1, 2, \dots, p. \quad (2.45)$$

Se o modelo de riscos proporcionais de Cox (2.23) está sendo utilizado, não está sendo assumido nenhuma distribuição aos tempos T dado as variáveis concomitantes  $x$ . Cox propôs então um método de estimar  $\beta$  na ausência do conhecimento de  $\lambda_0(t)$ , conhecido como " **método da verossimilhança parcial**". Este consiste basicamente em: suponha que em uma amostra aleatória de tamanho n, tem-se k observados distintos tempos de falha e n-k tempos censurados. Os k tempos de falha observados são denotados por  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  e seja  $R_i = R(t_{(i)})$  o conjunto de indivíduos vivos e não censurados anterior a  $t_{(i)}$ . Cox sugeriu a seguinte " função de verossimilhança " :

$$L(\beta) = \prod_{i=1}^k \left( \frac{\exp\{x_{(i)}\beta\}}{\sum_{l \in R_i} \exp\{x_l\beta\}} \right), \quad (2.46)$$

onde  $x_{(i)}$  é o vetor de variáveis concomitantes associado ao indivíduo observado falhar em  $t_{(i)}$ . A "verossimilhança" (2.46) não depende de  $\lambda_0(t)$  e pode ser maximizada para produzir as estimativas  $\hat{\beta}$ .

A motivação para o uso de (2.46) é que dado  $R(t)$  e dado que uma falha ocorre em t, a probabilidade de ser o indivíduo i ( $i \in R(t)$ ) que falhe é:

$$\frac{\lambda(t | x_i)}{\sum_{l \in R_i} \lambda(t | x_l)} = \frac{\lambda_0(t) \exp\{x_i\beta\}}{\sum_{l \in R_i} \lambda_0(t) \exp\{x_l\beta\}} = \frac{\exp\{x_i\beta\}}{\sum_{l \in R_i} \exp\{x_l\beta\}}.$$

A verossimilhança (2.46) é formada tomando-se o produto de todos tais fatores sobre os  $k$  tempos de falha observados.

Se empates existem na amostra, alguma modificação da verossimilhança é requerida, e uma sugestão dada por Breslow (1974) é usar:

$$L(\beta) = \prod_{i=1}^k \frac{\exp \{S_i \beta\}}{\left( \sum_{l \in R_i} \exp \{x_l \beta\} \right)^{d_i}}, \quad (2.47)$$

onde:  $d_i$  é o número de tempos de vida igual a  $t_{(i)}$  e

$S_i$  é a soma dos vetores de variáveis concomitantes  $x$  para esses  $d_i$  indivíduos, isto é,  $S_i = \sum_{l \in D_i} x_l$ , com  $D_i$  sendo o conjunto de indivíduos que falharam em  $t_{(i)}$ .

O log da verossimilhança (2.47) é:

$$\log L(\beta) = \sum_{i=1}^k \{S_i \beta\} - \sum_{i=1}^k d_i \log \left( \sum_{l \in R_i} \exp \{x_l \beta\} \right), \quad (2.48)$$

e as derivadas primeiras de (2.48) são:

$$\begin{aligned} \frac{\partial \log L(\beta)}{\partial \beta_r} &= \sum_{i=1}^k S_{ir} - \sum_{i=1}^k \left( d_i \frac{\sum_{l \in R_i} x_{lr} \exp \{x_l \beta\}}{\sum_{l \in R_i} \exp \{x_l \beta\}} \right) = \\ &= \sum_{i=1}^k \left[ S_{ir} - \left( d_i \frac{\sum_{l \in R_i} x_{lr} \exp \{x_l \beta\}}{\sum_{l \in R_i} \exp \{x_l \beta\}} \right) \right], \end{aligned} \quad (2.49)$$

para  $r = 1, 2, \dots, p$  e onde  $S_{ir}$  é a  $r$ -ésima componente em  $S_i = (S_{i1}, S_{i2}, \dots, S_{ip})$ .

As equações de máxima verossimilhança  $\frac{\partial \log L(\beta)}{\partial \beta_r} = 0$  ( $r = 1, 2, \dots, p$ ), podem ser resolvidas geralmente sem dificuldades, usando-se o método iterativo de Newton-Raphson.

A matriz de informação observada  $p \times p$  é:

$$I = \left( - \frac{\partial^2 \log L(\beta)}{\partial \beta_s \partial \beta_r} \right) \Big|_{\beta = \hat{\beta}} \quad r, s = 1, 2, \dots, p, \quad (2.50)$$

$$\text{onde : } - \frac{\partial^2 \log L(\underline{\beta})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^{d_i} d_i \left\{ \frac{\sum_{l \in R_i} x_{lr} x_{ls} \exp \{x_l \underline{\beta}\}}{\sum_{l \in R_i} \exp \{x_l \underline{\beta}\}} - \left( \frac{\sum_{l \in R_i} x_{lr} \exp \{x_l \underline{\beta}\}}{\sum_{l \in R_i} \exp \{x_l \underline{\beta}\}} \right) \left( \frac{\sum_{l \in R_i} x_{ls} \exp \{x_l \underline{\beta}\}}{\sum_{l \in R_i} \exp \{x_l \underline{\beta}\}} \right) / \left( \frac{\sum_{l \in R_i} \exp \{x_l \underline{\beta}\}}{\sum_{l \in R_i} \exp \{x_l \underline{\beta}\}} \right)^2 \right\}, \quad (2.51)$$

com  $r, s = 1, 2, \dots, p$ .

OBS: Quando não existem empates, todos os  $d_i$  ( $i = 1, 2, \dots, n$ ) são iguais a 1 e portanto os resultados acima incluem (2.46) como um caso particular.

## 2.7 - TESTES E INTERVALOS ESTIMADOS PARA OS COEFICIENTES $\underline{\beta}$

### 2.7.1 - Para os modelos paramétricos: log-linear e de riscos proporcionais

Testes e intervalos estimados para os coeficientes  $\underline{\beta}$  podem ser baseados nos métodos da razão de verossimilhança ou em aproximações Normal para  $\underline{\beta}$ . Quando existem censuras, é mais conveniente usar-se a aproximação Normal para  $\underline{\beta}$ , isto é,  $\hat{\underline{\beta}} \sim N(\underline{\beta}, I_0^{-1})$ , onde  $I_0^{-1}$  é o inverso da matriz de informação. Essa aproximação é razoavelmente boa se o tamanho amostral não é muito pequeno.

Muitas hipóteses podem ser feitas em torno de  $\underline{\beta}$  e podem ser colocadas na forma  $H_0: \underline{\beta}_1 = \underline{\beta}_1^0$  onde  $\underline{\beta}'$  é particionado como  $(\underline{\beta}_1, \underline{\beta}_2)'$  e onde  $\underline{\beta}_1$  é  $k \times 1$  ( $k < p$ ) e  $\underline{\beta}_1^0$  é um vetor especificado. Para testar  $H_0$  contra  $H_1: \underline{\beta}_1 \neq \underline{\beta}_1^0$  pode-se usar a estatística da razão de verossimilhança:

$$\gamma = -2 \log \left( \frac{L(\underline{\beta}_1^0, \tilde{\underline{\beta}}_2)}{L(\hat{\underline{\beta}}_1, \hat{\underline{\beta}}_2)} \right), \quad (2.52)$$

onde  $\tilde{\underline{\beta}}_2$  é o e.m.v. sob  $H_0$  e  $\hat{\underline{\beta}} = (\hat{\underline{\beta}}_1, \hat{\underline{\beta}}_2)$  é o e.m.v. de  $\underline{\beta}$ . A estimativa  $\hat{\underline{\beta}}_2$  é encontrada resolvendo-se o sistema de equações:

$$\frac{\partial \log L(\underline{\beta})}{\partial \beta_r} = 0, \quad \text{com } r = k+1, \dots, p \text{ e } \underline{\beta}_1 = \underline{\beta}_1^0.$$

Valores grandes de  $\gamma$  dão evidências contra  $H_0$  e níveis de significância podem ser calculados através da aproximação assintótica  $\gamma \sim \chi^2$  com  $k$  graus de liberdade.

Uma estatística alternativa para testar  $H_0$ , baseada na assintótica aproximação Normal  $\hat{\beta} \sim (\beta, I_0^{-1})$  é :

$$\gamma_1 = (\hat{\beta}_1 - \beta_1^0)' C_{11}^{-1} (\hat{\beta}_1 - \beta_1^0), \quad (2.53)$$

onde  $C_{11}$  é  $k \times k$  e  $C = I_0^{-1}$  é particionada como :

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix},$$

isto é,  $C_{11}$  é assintoticamente a matriz de covariância para  $\hat{\beta}_1$ . Sob  $H_0$ ,  $\gamma_1$  é também aproximadamente  $\chi^2$  com  $k$  graus de liberdade em amostras grandes. A aproximação  $\chi^2$  para  $\gamma$  é tipicamente um tanto melhor do que para  $\gamma_1$  em amostras de tamanho pequeno e moderado.

### 2.7.2 - Para o modelo semi-paramétrico : de riscos proporcionais de Cox

Para fazer inferências em torno de  $\beta$  precisa-se confiar em procedimentos para grandes amostras e uma maneira é tratar  $\hat{\beta} \sim N(\beta, I(\beta)^{-1})$ . Inferências também podem ser feitas através da razão de verossimilhança ou ainda baseadas em testes do vetor score  $p \times 1$ ,

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}, \quad (2.54)$$

o qual assintoticamente pode ser considerado Normalmente distribuído, isto é,

$$U(\beta) \sim N(0, I(\beta)).$$

## CAPÍTULO 3

### MODELOS DE REGRESSÃO EM EXPERIMENTOS DE DOSE-RESPOSTA EM UM CONTEXTO DE RISCOS COMPETITIVOS

#### 3.1 - INTRODUÇÃO

No capítulo 2 foram abordados os modelos de regressão em um contexto de um único risco. É sabido, contudo, que um indivíduo pode falhar por uma dentre muitas possíveis causas de falha. Em muitos problemas se faz necessário considerar essas diversas causas de falha e não somente uma delas. Esses problemas ocorrem na prática em muitas áreas tais como: epidemiologia, demografia, ensaios clínicos, aplicações em confiabilidade industrial e outras e são chamados **problemas de riscos competitivos**. Informações adicionais que descrevem as condições nas quais o experimento foi conduzido, ou seja variáveis concomitantes, podem também ser avaliáveis. No caso particular de experimentos de dose-resposta, considera-se dose como a variável concomitante de interesse e esta assume  $p$  níveis.

Então, os dados avaliáveis em experimentos de dose-resposta são: os tempos de falha ou de censura  $T$ , a causa de falha  $I$  e o nível da variável concomitante dose.

Em geral, os principais objetivos em tais experimentos são :

- estimar as taxas de sobrevivência para cada causa;
- estimar o risco de falha para uma determinada causa em um tempo  $t$  e com uma fixada dose  $d$ ;
- encontrar doses virtualmente seguras;
- encontrar DL50 ( dose letal 50 );
- testar o efeito da dose no tipo de falha;
- testar o efeito da dose no tempo de falha;

As seções seguintes tratam dos problemas de riscos competitivos em modelos de dose-resposta. Inicialmente é introduzido os conceitos básicos e notações utilizadas em tais situações e em seguida é tratado os mesmos modelos de regressão considerados no capítulo 2 em um contexto de riscos competitivos.

### 3.2 - RISCOS COMPETITIVOS - CONCEITOS BÁSICOS E NOTAÇÃO

O problema de riscos competitivos tem sido abordado de duas maneiras distintas. A primeira abordagem a ser tratada aqui deve-se a Prentice et al (1978) e esta leva em consideração somente dados observáveis, isto é, os tempos de falha ou de censuras observados e a causa observada para cada uma das falhas. A segunda abordagem usa o conceito de "tempos de falha latentes" o qual tem sido estudado por vários autores (Cox-1959, Moeschberger e David-1971, Gail-1975, dentre outros ).

Nas seções seguintes estas duas abordagens serão tratadas em maiores detalhes.



### 3.2.1 - Abordagem de Prentice et al para o problema de Riscos Competitivos

Suponha que um indivíduo de uma população esteja sujeito a  $k$  causas (ou riscos) de falha mutuamente exclusivas, as quais serão denotadas por  $C_1, C_2, \dots, C_k$ . Causa de falha é definida aqui como a "doença ou lesão a qual inicia uma sucessão de fatos que leva diretamente a falha". Suponha ainda que quando uma falha ocorre, a mesma é devida a uma única das  $k$  causas distintas consideradas, isto é, duas ou mais causas de falha não ocorrem simultaneamente.

Uma amostra de tamanho  $n$  é então retirada desta população para a realização de um experimento, o qual pode ser um experimento de dose-resposta. Cada indivíduo amostrado pode então ser representado por  $(T, I, \underline{x})$ , onde :

-  $T$  é a variável aleatória não negativa, geralmente contínua, correspondendo aos tempos de falha ou tempos censurados,

-  $I$  é o índice aleatório definido por :

$$I = \begin{cases} i & \text{se falha é devido a } C_i, \quad i = 1, 2, \dots, k \\ 0 & \text{caso contrário,} \end{cases}$$

-  $\underline{x}$  é o vetor de variáveis concomitantes.

As funções de sobrevivência e risco, com tais variáveis concomitantes são definidas por:

$$S_T(t | \underline{x}) = \Pr [ T > t | \underline{x} ] \quad (3.1)$$

e

$$\lambda_T(t | \underline{x}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr [ T \in [t, t + \Delta t) \mid T \geq t, \underline{x} ]}{\Delta t} = \frac{f_T(t | \underline{x})}{S_T(t | \underline{x})}, \quad (3.2)$$

onde:  $f_T(t | \underline{x})$  é a função de densidade de probabilidade da v. a.  $T$  dado  $\underline{x}$ .

Similarmente, pode-se definir uma função risco de causa específica, dada por :

$$\lambda_i(t | \underline{x}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr [ T \in [t, t + \Delta t), I=i \mid T \geq t, \underline{x} ]}{\Delta t}, \quad (3.3)$$

para  $i = 1, 2, \dots, k$ . Em palavras,  $\lambda_i(t | \mathcal{X})$  é a taxa de falha instantânea da causa  $i$  no tempo  $t$  dado  $\mathcal{X}$ , na presença dos outros tipos de falha. Como é assumido que o tipo de falha  $i$  deve ser um único elemento de  $\{1, 2, \dots, k\}$ , a função risco total pode ser escrita como :

$$\lambda_T(t | \mathcal{X}) = \sum_{i=1}^k \lambda_i(t | \mathcal{X}), \quad (3.4)$$

e a função de sobrevivência total (3.1) pode ser escrita por :

$$\begin{aligned} S_T(t | \mathcal{X}) &= \exp \left\{ - \int_0^t \lambda_T(u | \mathcal{X}) du \right\} = \exp \left\{ - \int_0^t \sum_{i=1}^k \lambda_i(u | \mathcal{X}) du \right\} = \\ &= \prod_{i=1}^k \exp \left\{ - \int_0^t \lambda_i(u | \mathcal{X}) du \right\} = \prod_{i=1}^k G_i(t | \mathcal{X}), \end{aligned} \quad (3.5)$$

onde:  $G_i(t | \mathcal{X}) = \exp \left\{ - \int_0^t \lambda_i(u | \mathcal{X}) du \right\}$  representa, segundo Elandt-Johnson

(1980), a distribuição associada a causa  $C_i$  sozinha.

A função  $f_i(t | \mathcal{X})$ , a qual é chamada por Kalbfleisch e Prentice (1980) de **função sub-densidade** do tempo de falha para a causa de falha  $C_i$  dado as variáveis concomitantes  $\mathcal{X}$ , é dada por :

$$f_i(t | \mathcal{X}) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t \leq T < t + \Delta t, I = i | \mathcal{X}]}{\Delta t} = \lambda_i(t | \mathcal{X}) S_T(t | \mathcal{X}), \quad (3.6)$$

para  $i = 1, 2, \dots, k$ . Conseqüentemente, a função risco de causa-específica pode ser expressa por :

$$\lambda_i(t | \mathcal{X}) = \frac{f_i(t | \mathcal{X})}{S_T(t | \mathcal{X})}, \quad i = 1, 2, \dots, k. \quad (3.7)$$

Pode-se portanto notar através de (3.4), (3.5) e (3.6), que a função de verossimilhança pode ser escrita completamente em termos das funções risco de causa-específica. Além disso, as funções risco de causa-específica são **identificáveis**, ou seja, elas podem ser estimadas pelos dados do tipo  $(T, I, \mathcal{X})$  sem outras suposições.

### 3.2.2- O problema de riscos competitivos usando o conceito de tempos de falha latentes

Suponha que em uma amostra de tamanho  $n$ , cada indivíduo esteja sujeito a  $k$  causas de falhas denotadas por  $C_1, C_2, \dots, C_k$ , e que a cada uma dessas  $k$  causas esteja associado um tempo de falha  $T_i$ ,  $i = 1, 2, \dots, k$ . Assim, para cada causa competindo tem-se um tempo associado. Estes tempos são denominados de "tempos de falha latentes" ou "tempos de falha potencial" ou ainda "tempos de falha hipotéticos". Assume-se também que a falha de um indivíduo não ocorre por duas ou mais causas, mas somente por uma única dentre as  $k$  possíveis.

Deve ser enfatizado contudo, que este conceito de tempos de falha latentes é introduzido puramente por "conveniência matemática". Não é exigido que estes tempos existam em qualquer sentido real - meramente sua introdução facilita o desenvolvimento da teoria.

Visto que é impossível observar conjuntamente os  $T_i$ ,  $i = 1, 2, \dots, k$ , uma vez que um indivíduo falha na ocorrência de uma das  $k$  causas, o tempo de falha a ser realmente observado é:

$$T = \min ( T_1, T_2, \dots, T_k ). \quad (3.8)$$

Este tempo  $T$  é chamado "tempo de vida observado" ou "tempo de vida real" do indivíduo.

Neste contexto, tem-se que a função de distribuição de sobrevivência conjunta dos tempos de falha latentes dado as variáveis concomitantes  $\mathcal{X}$ , é dada por:

$$\begin{aligned} S_{T_1, T_2, \dots, T_k} ( t_1, t_2, \dots, t_k | \mathcal{X} ) &= P [ T_1 > t_1, T_2 > t_2, \dots, T_k > t_k | \mathcal{X} ] = \\ &= P \left[ \bigcap_{i=1}^k T_i > t_i | \mathcal{X} \right] \end{aligned} \quad (3.9)$$

onde:  $0 < t_i < \infty$  para  $i = 1, 2, \dots, k$ ;  $S_{T_1, \dots, T_k}(0, 0, \dots, 0 | \mathcal{X}) = 1$  e  $S_{T_1, \dots, T_k}(\infty, \infty, \dots, \infty | \mathcal{X}) = 0$ .

Conseqüentemente, as funções de sobrevivência marginais dos tempos de falha latentes  $T_i$ ,  $i = 1, 2, \dots, k$ , dado as variáveis concomitantes  $\mathcal{X}$ , são dadas por:

$$S_{T_i}(t_i | \mathcal{X}) = \Pr [ T_i > t_i | \mathcal{X} ] = S_{T_1, \dots, T_k} (0, 0, \dots, t_i, \dots, 0 | \mathcal{X} ), \quad (3.10)$$

para  $i = 1, 2, \dots, k$  e onde  $t_i$  em qualquer caso está na  $i$ -ésima posição da expressão. A correspondente função risco de  $S_{T_i}(t_i | \mathcal{X})$  é :

$$h_i(t_i | \mathcal{X}) = - \frac{\partial}{\partial t_i} \log S_{T_i}(t_i | \mathcal{X}) = - \frac{\left( \frac{\partial}{\partial t_i} S_{T_i}(t_i | \mathcal{X}) \right)}{S_{T_i}(t_i | \mathcal{X})} \quad \text{para } i = 1, 2, \dots, k, \quad (3.11)$$

assumindo que tais derivadas existam.

Em geral, os tempos de falha latentes  $T_1, T_2, \dots, T_k$ , bem como os tempos censurados, são assumidos serem independentes. Nas situações onde tal suposição de independência não é realística, Kalbfleisch e Prentice (1980) sugerem o uso de covariáveis tempo-dependentes. Se os  $T_i$ ,  $i = 1, 2, \dots, k$  são assumidos serem independentes, tem-se que a função de distribuição de sobrevivência conjunta (3.9) pode ser expressa por :

$$S_{T_1, \dots, T_k} (t_1, \dots, t_k | \mathcal{X}) = \prod_{i=1}^k S_{T_i}(t_i | \mathcal{X}). \quad (3.12)$$

Como foi visto anteriormente, pode-se observar somente o tempo  $T = \min (T_1, \dots, T_k)$ . Assim, a função de sobrevivência para este tempo de vida  $T$  dado as variáveis concomitantes  $\mathcal{X}$ , é dada por :

$$\begin{aligned} S_T(t | \mathcal{X}) &= \Pr [ T > t | \mathcal{X} ] = P [ \min (T_1, \dots, T_k) > t | \mathcal{X} ] = \\ &= \Pr [ T_1 > t, T_2 > t, \dots, T_k > t | \mathcal{X} ] = S_{T_1, \dots, T_k}(t, t, \dots, t | \mathcal{X}). \end{aligned} \quad (3.13)$$

Esta função de sobrevivência corresponde à função de sobrevivência total dada em (3.1).

A função risco correspondente a (3.13) é portanto :

$$\begin{aligned}\lambda_T(t | \mathcal{X}) &= - \frac{\partial}{\partial t} S_T(t | \mathcal{X}) / S_T(t | \mathcal{X}) = - \frac{\partial}{\partial t} \ln S_T(t | \mathcal{X}) = \\ &= - \frac{\partial}{\partial t} \ln S_{T_1, \dots, T_k}(t, t, \dots, t | \mathcal{X}),\end{aligned}\quad (3.14)$$

que é a taxa de falha instantânea por qualquer causa no tempo  $t$ , dado as variáveis concomitantes  $\mathcal{X}$ , assumindo que o indivíduo sobreviveu até o tempo  $t$ . Esta função risco corresponde a função risco total dada em (3.2).

Neste contexto, as funções risco de causa específica dadas em (3.3) podem ser expressas por :

$$\begin{aligned}\lambda_i(t | \mathcal{X}) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr [ t \leq T_i < t + \Delta t | T > t ; \mathcal{X} ]}{\Delta t} = \\ &= - \frac{\partial}{\partial t_i} \ln S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X}) \Big|_{t_1=t_2=\dots=t},\end{aligned}\quad (3.15)$$

para  $i = 1, 2, \dots, k$ , assumindo que as derivadas existam. A função  $\lambda_i(t | \mathcal{X})$  é chamada "taxa de risco bruta" ou "força bruta de mortalidade". Ela fornece a taxa instantânea de falha pela causa  $C_i$  no tempo  $t$  dado as variáveis concomitantes  $\mathcal{X}$ , quando todas as outras causas estão atuando simultaneamente.

Visto que :

$$\frac{\partial}{\partial t} \ln S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X}) = \sum_{i=1}^k \frac{\partial}{\partial t_i} \ln S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X}) \Big|_{t_1=\dots=t_k=t},$$

$$\text{tem-se :} \quad \lambda_T(t | \mathcal{X}) = \sum_{i=1}^k \lambda_i(t | \mathcal{X}), \quad (3.16)$$

uma vez que falha de diferentes causas são eventos mutuamente exclusivos.

Por outro lado, as funções de sobrevivência marginais dos tempos de falha latentes  $T_i$  ( $i = 1, 2, \dots, k$ ) são dadas por :

$$S_{T_i}(t | \mathcal{X}) = \Pr [ T_i > t | \mathcal{X} ] = S_{T_1, \dots, T_k}(0, 0, \dots, t, \dots, 0 | \mathcal{X}), \quad i = 1, 2, \dots, k, \quad (3.17)$$

com  $t$  ocupando a  $i$ -ésima posição. Suas correspondentes funções risco marginais, assumindo que as derivadas existam, são :

$$h_i(t | \mathcal{X}) = - \frac{\partial}{\partial t} \ln S_{T_i}(t | \mathcal{X}) = - \frac{\partial}{\partial t} \ln S_{T_1, \dots, T_k}(0, 0, \dots, t, \dots, 0 | \mathcal{X}), \quad (3.18)$$

para  $i = 1, 2, \dots, k$ . Esta função é a taxa de falha instantânea associada ao tempo de falha latente  $T_i$ . Em geral não é justificável que ela represente a taxa de falha instantânea para a causa  $C_i$  somente. A função  $h_i(t | \mathcal{X})$  é chamada "taxa de risco líquida" ou "força líquida de mortalidade". Em geral  $\lambda_i(t | \mathcal{X}) \neq h_i(t | \mathcal{X})$ . Agora, se os  $T_i$ ,  $i = 1, 2, \dots, k$  são independentes, tem-se :

$$S_T(t | \mathcal{X}) = \prod_{i=1}^k S_{T_i}(t | \mathcal{X}), \quad (3.19)$$

$$\lambda_i(t | \mathcal{X}) = h_i(t | \mathcal{X}), \quad (3.20)$$

e 
$$G_i(t | \mathcal{X}) = S_{T_i}(t | \mathcal{X}). \quad (3.21)$$

Deve-se notar aqui que : se os  $T_i$ ,  $i=1, 2, \dots, k$  são independentes  $\Rightarrow \lambda_i(t | \mathcal{X}) = h_i(t | \mathcal{X})$ ; mas se  $\lambda_i(t | \mathcal{X}) = h_i(t | \mathcal{X})$  não implica que os  $T_i$ ,  $i = 1, 2, \dots, k$  são independentes (Gail-1975); contudo os eventos  $\{ T_i > t \}$   $i = 1, 2, \dots, k$  são independentes (Elandt-Johnson - 1980).

Sempre que ocorra a falha de um indivíduo, é assumido ser possível observar somente seu tempo de vida real  $T = \min(T_1, \dots, T_k)$  e a causa de sua falha  $I = i$ ,  $i = 1, 2, \dots, k$ , isto é, pode-se observar os valores da variável aleatória bi-dimensional  $(T, I)$ . Como é também assumido que a probabilidade de dois ou mais tempos de falha latentes ser iguais é zero, ou seja,  $\Pr [ T_i = T_j ] = 0$  para  $i \neq j$ , o valor de  $I$  é unicamente determinado com probabilidade 1. Assim, a distribuição de probabilidade conjunta e  $(T, I)$  dado as variáveis concomitantes  $\mathcal{X}$ , pode ser expressa por :

$$\begin{aligned} F_{T,I}(t, i | \mathcal{X}) &= \Pr [ T \leq t ; I=i | \mathcal{X} ] = \Pr [ T_i \leq t ; T_i < T_j, i \neq j | \mathcal{X} ] = \\ &= \int_0^t \left[ \int_{t_i}^{\infty} \dots \int_{t_i}^{\infty} f(t_1, \dots, t_k | \mathcal{X}) \prod_{r \neq i} dt_r \right] dt_i, \end{aligned} \quad (3.22)$$

para  $i = 1, 2, \dots, k$  e onde  $f(t_1, \dots, t_k | \mathcal{X})$  é a f.d.p. conjunta dos tempos de falha latentes  $T_i$ ,  $i = 1, 2, \dots, k$  dado as variáveis concomitantes  $\mathcal{X}$ . A distribuição  $F_{T,I}(t, i | \mathcal{X})$  dada em (3.22) é estimável pelos dados  $(T, I, \mathcal{X})$ .

Consequentemente, a distribuição de probabilidade para o tempo de falha real  $T = \min(T_1, \dots, T_k)$  é :

$$F_T(t | \mathcal{X}) = \Pr [ T \leq t | \mathcal{X} ] = \sum_{i=1}^k F_{T,I}(t, i | \mathcal{X}), \quad (3.23)$$

e a função de sobrevivência para o tempo de falha real  $T = \min(T_1, \dots, T_k)$  é :

$$S_T(t | \mathcal{X}) = \Pr [ T > t | \mathcal{X} ] = 1 - \sum_{i=1}^k F_{T,I}(t, i | \mathcal{X}). \quad (3.24)$$

Também tem-se a expressão :

$$\begin{aligned} \pi_i &= (\text{definição}) = \Pr [ I=i | \mathcal{X} ] = F_{T,I}(\infty, i | \mathcal{X}) = \Pr [ T < \infty, I = i | \mathcal{X} ] = \\ &= \int_0^\infty \left[ \int_{t_i}^\infty \dots \int_{t_i}^\infty f(t_1, \dots, t_k | \mathcal{X}) \prod_{r \neq i} dt_r \right] dt_i, \end{aligned} \quad (3.25)$$

a qual é a probabilidade marginal de que falha ocorrerá em qualquer tempo devido a causa  $C_i$ . Pode ainda ser de interesse, considerar a distribuição condicional de  $T$  dado que  $I=i$ , ou seja,

$$F_{T|I}(t | i; \mathcal{X}) = \Pr [ T \leq t | I=i; \mathcal{X} ] = \frac{\Pr [ T \leq t, I = i | \mathcal{X} ]}{\Pr [ I = i | \mathcal{X} ]} = \frac{F_{T,I}(t, i | \mathcal{X})}{\pi_i}, \quad (3.26)$$

para  $i = 1, 2, \dots, k$ .

OBS: Esta abordagem utilizando tempos de falha latentes tem sido criticada por Kalbfleisch e Prentice (1980) devido as suas suposições e falta de interpretação física dos  $k-1$  tempos latentes não observados.

### 3.2.3 - O problema da identificabilidade

Ao trabalhar-se com riscos competitivos usando o conceito de tempos de falha latentes, surge um problema chamado "problema da identificabilidade". O mesmo pode ser resumido basicamente em:

" Se  $P[I=i | \mathcal{X}]$  e  $F_{T|I}(t | i; \mathcal{X})$  ou  $F_{T,I}(t, i | \mathcal{X})$  são dadas para  $i = 1, 2, \dots, k$ , a distribuição de probabilidade conjunta dos tempos de falha latentes  $T_1, T_2, \dots, T_K$ , a qual é dada pela função de sobrevivência  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  e suas distribuições marginais  $S_{T_i}(t | \mathcal{X})$ ,  $i = 1, 2, \dots, k$  são identificáveis ? "

Se  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  for identificável segue que todas as demais distribuições dadas na seção 3.2.2 serão também identificáveis, uma vez que todas elas seguem de  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$ .

Esta questão foi investigada por alguns autores considerando as observações  $(T, I)$ , ou seja, sem considerar variáveis concomitantes e a conclusão a que chegaram foi a seguinte: sem a suposição de independência entre os tempos de falha latentes, a distribuição de sobrevivência conjunta  $S_{T_1, \dots, T_k}(t_1, \dots, t_k)$  não é identificável, bem como as suas marginais  $S_{T_i}(t)$  para  $i = 1, 2, \dots, k$  não são identificáveis.

Berman (1963) mostrou que se os tempos de falha latentes  $T_1, \dots, T_k$  são independentes, então as probabilidades  $F_{T,I}(t, i)$   $i = 1, 2, \dots, k$  ou as probabilidades  $\Pr[I = i]$  e  $F_{T|I}(t | i)$ ,  $i = 1, 2, \dots, k$  determinam todos os  $S_{T_i}(t)$  e todos os  $F_{T_i}(t)$ , então também a distribuição de probabilidade conjunta dos tempos de falha latentes  $T_1, \dots, T_k$ . Bose (1973) mostrou que se não é assumido  $T_1, \dots, T_k$  independentes, as probabilidades  $\Pr[I=i]$  e  $F_{T|I}(t | i)$  ou  $F_{T,I}(t, i)$   $i = 1, 2, \dots, k$  não podem determinar a distribuição de sobrevivência conjunta de  $T_1, T_2, \dots, T_k$ .

Esses resultados são estendidos diretamente para os dados  $(T, I, \mathcal{X})$  com  $\mathcal{X}$  sendo variáveis concomitantes tempo independentes. Assim, somente se os  $T_i$   $i = 1, 2, \dots, k$  são independentes, pode-se identificar a distribuição de sobrevivência  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  bem como suas marginais  $S_{T_i}(t | \mathcal{X})$ ,  $i = 1, 2, \dots, k$ .

Surge então a seguinte questão: " Como fazer para obter conclusões à respeito dos tempos de falha latentes  $T_1, \dots, T_k$  quando a suposição de independência não pode ser assumida ? ". Segundo Birnbaum (1980), o conhecimento da distribuição de probabilidade de  $(T, I, \mathcal{X})$  deve ser suplementado por alguma informação adicional, seja estabelecida na forma de outras suposições ou pela extração de observações adicionais. Uma maneira efetiva de superar a questão é assumir que  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  pertença a



uma especificada família paramétrica. A escolha apropriada de um modelo paramétrico dependerá de um completo conhecimento do processo de funcionamento de falha dos indivíduos.

Uma vez assumido um modelo paramétrico para  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$ , é possível identificar os parâmetros dessa função de distribuição de sobrevivência pela distribuição de  $(T, I, \mathcal{X})$ ?. De acordo com Elandt-Johnson (1980), a identificação dos parâmetros depende da particular família paramétrica assumida, ou seja, para algumas famílias paramétricas é possível a identificação de seus parâmetros e para outras não.

Mesmo que possa ser possível estimar todos os parâmetros do modelo paramétrico assumido através dos dados  $(T, I, \mathcal{X})$ , pode não ser possível verificar a validade do modelo. Por exemplo, dados do tipo  $(T, I, \mathcal{X})$  não permitem distinguir entre um modelo de riscos competitivos independente e uma infinidade de modelos dependentes obtidos com as mesmas funções risco de causa-específica,  $\lambda_i(t | \mathcal{X})$   $i = 1, 2, \dots, k$ .

Igualmente, se um modelo paramétrico com causas (ou riscos) dependentes é assumido para  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  e todos os parâmetros são identificáveis, será impossível distinguir entre o modelo assumido e um outro modelo com riscos independentes mas com as mesmas funções risco de causa-específica. Em outras palavras, a suposição de dependência assumida no modelo não pode ser verificada através dos dados  $(T, I, \mathcal{X})$ .

### 3.3 - MODELOS DE REGRESSÃO EM UM CONTEXTO DE RISCOS COMPETITIVOS

Uma maneira frequentemente usada para estudar a associação entre os tempos de falhas e variáveis concomitantes fixas  $\mathcal{X}$ , é o uso de modelos de regressão, uma vez que esses modelos permitem incorporar variáveis concomitantes e dados censurados.

Diversos autores, dentre eles: Holt (1978), Lagakos (1978), Kalbfleisch e Prentice (1980), Farewell e Prentice (1977), Larson (1984), têm estudado modelos de regressão em situações de riscos competitivos. Os

modelos de regressão a serem abordados nesta seção são os mesmos tratados no Capítulo 2, uma vez que os mesmos podem ser estendidos a riscos competitivos utilizando-se uma formulação apropriada aos riscos de causa-específica  $\lambda_i(t | \mathcal{X})$ . Serão revistos então os modelos :

- 1 - Modelos Paramétricos : - log-linear  
- de riscos proporcionais
- 2 - Modelo semi-paramétrico : - de riscos proporcionais de Cox,

estendidos a situações de riscos competitivos.

Os modelos citados serão tratados sob a abordagem de tempos de falha latentes. Será assumido que estes tempos são independentes.

### 3.3.1 - Modelo log-linear em situações de riscos competitivos.

Uma vez tendo assumido um modelo paramétrico para  $S_{T_1, \dots, T_k}(t_1, \dots, t_k | \mathcal{X})$  ou equivalentemente para cada uma das funções risco de causa específica  $\lambda_i(t | \mathcal{X})$ ,  $i = 1, 2, \dots, k$ , pode-se alternativamente modelar tais funções risco de causa específica com variáveis concomitantes  $\mathcal{X}$  tempo independente, usando o modelo log-linear, isto é, modelando as funções risco de causa-específica por:

$$\lambda_i(t | \mathcal{X}) = \lambda_{0i} \left( t \exp[-\mathcal{X} \underline{\beta}_i] \right) \exp[-\mathcal{X} \underline{\beta}_i], \quad (3.27)$$

para  $i = 1, 2, \dots, k$  e onde :

$\lambda_{0i}(\cdot)$ : é a função risco para a causa  $i$  a qual tem uma forma paramétrica particular  
 $\underline{\beta}_i$ : é um vetor coluna de parâmetros de regressão associados a causa  $i$ , tal que os parâmetros de  $\underline{\beta}_i$  não estão presentes em nenhuma outra causa  $j$ , ou seja,  $\lambda_i(t | \mathcal{X})$  depende somente dos parâmetros  $\underline{\beta}_i$  e estes parâmetros não estão presentes em nenhuma outra função risco  $\lambda_j(t | \mathcal{X})$ , para todo  $j \neq i$ .

Este modelo foi sugerido por Kalbfleisch e Prentice (1978) e é uma extensão para riscos competitivos do modelo log-linear tratado na seção 2.3.1 onde está sendo suposto modelos paramétricos apropriados para os riscos para cada causa  $i$ . A estimação dos elementos de  $\underline{\beta}$  será tratada mais adiante.

### 3.3.2 - Modelo de riscos proporcionais em situações de riscos competitivos

O modelo de riscos proporcionais apresentado na seção 2.3.2 pode ser facilmente estendido para problemas de riscos competitivos.

Quando é assumido um modelo paramétrico para as funções risco de causa específica  $\lambda_i(t | \underline{x})$  para  $i = 1, 2, \dots, k$  e observa-se que para cada causa  $i$ ,  $i = 1, 2, \dots, k$ , a razão  $\lambda_i(t | \underline{x}_1) / \lambda_i(t | \underline{x}_2)$  para quaisquer dois indivíduos com vetores de variáveis concomitantes  $\underline{x}_1$  e  $\underline{x}_2$  não varia com o tempo, então as funções risco de causa específica podem ser escritas por:

$$\lambda_i(t | \underline{x}) = \lambda_{0i}(t) g_i(\underline{x}, \underline{\beta}_i), \quad (3.28)$$

onde:  $\lambda_{0i} \geq 0$  é uma função risco básica para a causa  $i$ , ou seja, é o risco para a causa  $i$  quando  $g(\underline{x}, \underline{\beta}_i) = 1$ ,

$g_i(\underline{x}, \underline{\beta}_i)$  é uma função que depende de  $\underline{x}$  e  $\underline{\beta}_i$ ,

$\underline{\beta}_i$  é um vetor coluna de coeficientes de regressão da causa específica  $i$ .

É usual tomar-se  $g_i(\underline{x}, \underline{\beta}_i)$  como sendo  $\exp\{\underline{x} \underline{\beta}_i\}$  para  $i = 1, 2, \dots, k$  pelo fato de se estar garantindo que  $g_i(\underline{x}, \underline{\beta}_i) > 0$  para todo  $\underline{x}$  e consequentemente  $\lambda_i(t | \underline{x}) > 0$  para todo  $\underline{x}$  e  $\underline{\beta}_i$ .

Os modelos de riscos proporcionais assumem então que as variáveis concomitantes  $\underline{x}$  tem um efeito multiplicativo na função risco de causa específica.

### 3.3.3 - Modelo de riscos proporcionais de Cox em situações de riscos competitivos

A extensão do modelo de riscos proporcionais de Cox para problemas de riscos competitivos tem sido apresentada por Holt (1978), Prentice e Breslow (1978), Kalbfleisch e Prentice (1980), dentre outros.

Neste modelo, a função risco de causa-específica é dada por :

$$\lambda_i(t | \underline{x}) = \lambda_{0i}(t) \exp \left[ \sum_j x_j \beta_{ij} \right], \quad i = 1, 2, \dots, k \quad (3.29)$$

onde :  $\lambda_{0i}(t) \geq 0$  é uma função risco arbitrária não especificada para a causa  $i$ ,  
 $\underline{\beta}_i$  é um vetor coluna de parâmetros de regressão associados a causa  $i$ .

Este modelo tem sido frequentemente usado pelo fato de não ser necessário assumir nenhuma forma paramétrica particular aos  $\lambda_{0i}(t)$ , o que o torna bastante flexível. A única suposição feita é de que os riscos são proporcionais ou equivalentemente, que as variáveis concomitantes  $\underline{x}$  atuam multiplicativamente na função risco de causa específica.

Os modelos apresentados aqui podem ser aplicados não somente em problemas de estimação de riscos, mas também para modelar e analisar dados que aparecem em uma variedade de experimentos, onde a resposta observada é o tempo (de falha ou de censura).

### 3.4 - INVESTIGAÇÃO DE UM MODELO

Quando tem-se dados do tipo  $(T, I, \underline{x})$ , o problema inicial consiste em decidir qual modelo deve-se assumir. Nos experimentos de dose-resposta onde a variável concomitante de interesse é a dose, alguns gráficos podem fornecer uma boa visualização da relação existente entre os tempos e a dose, ajudando assim, na escolha de um modelo inicial. Por exemplo, para cada causa  $i$ ,  $i = 1, 2, \dots, k$ , gráficos do tempo  $t_i$  contra dose ou  $\ln t_i$  contra dose ou ainda  $\ln t_i$  contra  $\ln$  dose podem sugerir a existência ou não de uma relação linear entre os tempos e a dose.

Pode-se também para cada causa  $i$ ,  $i = 1, 2, \dots, k$  particionar os dados de acordo com os níveis da dose e verificar para essas diferentes partes os modelos ajustados, examinando-se, por exemplo, a constância de  $\sigma$ .

Se um modelo de riscos proporcionais é assumido pode-se verificar essa suposição utilizando-se os dados  $(T, I, d)$  plotando-se :

1-  $\log(-\log \hat{S}_{ji}(t))$  contra  $t$  ou

2-  $\log(-\log \hat{S}_{ji}(t))$  contra  $\log t$ ,

para  $i = 1, 2, \dots, k$ , onde  $\hat{S}_{ji}(t) = \hat{S}_i(t | \underline{x}_j)$  são as estimativas Produto-Limite da função de sobrevivência em cada  $\underline{x}_j$  para cada causa  $i$ .

Se os gráficos 1) forem aproximadamente paralelos para  $i = 1, 2, \dots, k$  então o modelo de riscos proporcionais é apropriado. Ainda, se os gráficos 2) forem aproximadamente lineares então o modelo de riscos proporcionais Weibull é apropriado.

Portanto deve-se inicialmente explorar os dados do tipo  $(T, I, \underline{x})$  a fim de se obter algumas indicações do modelo de regressão a ser assumido.

### 3.5 - ANÁLISE DE RESÍDUOS

Tendo ajustado um modelo de regressão às funções risco de causa específica  $\lambda_i(t | \underline{x})$ ,  $i = 1, 2, \dots, k$ , o método de Cox e Snell (1968) pode ser empregado para definir "quantidades residuais" apropriadas  $\epsilon_{ij}$  ( $i = 1, 2, \dots, k$  e  $j = 1, 2, \dots, n_i$ ). Essas quantidades definem para cada  $i$  uma sequência de variáveis independentes e identicamente distribuídas com função de distribuição  $F_{\epsilon_i}$ .

#### 3.5.1 - Caso dos modelos paramétricos

No caso dos modelos paramétricos log-linear (3.27) e de Riscos proporcionais (3.28) num contexto de riscos competitivos, pode-se através do método de Cox e Snell, definir as seguintes quantidades residuais :

$$\epsilon_{ij} = \left[ t_{ij} \exp \{ - x_{ij} \hat{\beta}_i \} \right]^\delta, \quad \begin{matrix} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n_i \end{matrix} \quad (3.30)$$

onde:  $\epsilon_{ij}$  : quantidade residual para o j-ésimo indivíduo falhando pela causa i ;  
 $t_{ij}$  : tempo observado de falha do j-ésimo indivíduo que falhou pela causa i ;  
 $n_i$  : quantidade de indivíduos que falharam pela causa i e  
 $\delta = 1/ \sigma$ .

Para os tempos censurados  $t_{c_1}, \dots, t_{c_{(n-n_i)}}$  pode-se definir as quantidades residuais  $\epsilon_{c_l}$  para  $l = 1, 2, \dots, (n - n_i)$  por :

$$\epsilon_{c_l} = \left[ t_{c_l} \exp \{ - x_{c_l} \hat{\beta}_i \} \right]^\delta + 1. \quad (3.31)$$

Deste modo, os resíduos para os modelos: log-linear (3.27) e de riscos proporcionais (3.28) podem ser definidos por :

$$\begin{cases} \hat{\epsilon}_{ij} = [t_{ij} \exp\{-x_{ij}\hat{\beta}_i\}]^{\hat{\delta}} & \text{para } t_{ij} \in F_i \text{ (} i = 1, 2, \dots, k \text{ e } j = 1, 2, \dots, n_i) \\ \hat{\epsilon}_{c_l} = [t_{c_l} \exp\{-x_{c_l}\hat{\beta}_i\}]^{\hat{\delta}} + 1 & \text{para } t_{c_l} \in C \text{ (} l = 1, 2, \dots, n - n_i) \end{cases} \quad (3.32)$$

onde :  $F_i$  = conjunto dos tempos de falha pela causa i  
 $C$  = conjunto dos tempos censurados  
 $\hat{\delta} = 1/ \hat{\sigma}$ .

OBS: Considera-se como tempos censurados com relação a causa i, todos os tempos de falha devido a outras causas que não i bem como os tempos dos indivíduos que não falharam por nenhuma das causas.

Para os **casos particulares** onde aos modelos log-linear (3.27) e de riscos proporcionais (3.28) é assumido a forma paramétrica Weibull ou Exponencial aos  $\lambda_{oi}(\cdot)$ , as quantidades  $[t_{ij} \exp \{ - x_{ij} \hat{\beta}_i \}]^\delta$  são os riscos estimados, conforme mostrado na seção 2.5.1. Assim, nestes casos particulares tem-se os resíduos:

$$\begin{cases} \hat{\epsilon}_{ij} = \hat{H}_{oi}(t_{ij}) \exp \{ - x_{ij} \hat{\beta}_i \} = \hat{H}_i(t_{ij} | x_{ij}) & \text{para } t_{ij} \in F_i \\ \hat{\epsilon}_{c_l} = \hat{H}_{oi}(t_{c_l}) \exp \{ - x_{c_l} \hat{\beta}_i \} + 1 = \hat{H}_i(t_{c_l} | x_{c_l}) + 1 & \text{para } t_{c_l} \in C, \end{cases} \quad (3.33)$$

para  $i = 1, 2, \dots, k$ ;  $j = 1, 2, \dots, n_i$  e  $l = 1, 2, \dots, n - n_i$ .

Estes resíduos podem ser tratados como uma amostra aleatória proveniente da distribuição Exp (1). Ainda, nestes casos particulares, se resíduos são baseados em (3.33), os gráficos de  $-\log(\hat{S}_i(\hat{e}_i))$  contra  $\hat{e}_i$  onde  $\hat{e}_i = (\hat{e}_{i1}, \hat{e}_{i2}, \dots, \hat{e}_{in_i}, \hat{e}_{c_1}, \hat{e}_{c_2}, \dots, \hat{e}_{c_{n-n_i}})$  para  $i = 1, 2, \dots, k$  e  $\hat{S}_i(\hat{e}_i)$  é a estimativa produto-limite da função de sobrevivência para os  $\hat{e}_i$  ( $i = 1, 2, \dots, k$ ), devem fornecer aproximadamente uma reta com inclinação 1 se o modelo é adequado para  $i = 1, 2, \dots, k$ .

### 3.5.2 - Caso do modelo semi-paramétrico

De maneira análoga ao que foi tratado na seção 2.5.2, os resíduos para o modelo de riscos proporcionais de Cox podem ser definidos por:

$$\begin{cases} \hat{e}_{ij} = \hat{H}_i(t_{ij} | \mathcal{X}_{ij}) = \hat{H}_{oi}(t_{ij}) \exp\{\mathcal{X}_{ij} \hat{\beta}_i\} = [-\ln \hat{S}_{oi}(t_{ij})] \exp\{\mathcal{X}_{ij} \hat{\beta}_i\} \\ \hat{e}_{c_l} = \hat{H}_i(t_{c_l} | \mathcal{X}_{c_l}) = \hat{H}_{oi}(t_{c_l}) \exp\{\mathcal{X}_{c_l} \hat{\beta}_i\} + 1 = [-\ln \hat{S}_{oi}(t_{c_l})] \exp\{\mathcal{X}_{c_l} \hat{\beta}_i\} + 1, \end{cases} \quad (3.34)$$

para  $i = 1, 2, \dots, k$ ;  $j = 1, 2, \dots, n_i$  e  $l = 1, 2, \dots, n - n_i$ , e onde  $H_{oi}(t)$  é a função risco básica acumulada e  $\hat{H}_{oi}(t) = -\log \hat{S}_{oi}(t)$  é um estimador de  $H_{oi}(t)$ .

O estimador de máxima verossimilhança de  $S_{oi}(t)$  é:

$$\hat{S}_{oi}(t) = \prod_{j: t_{ij} < t} \hat{\alpha}_{ij} \quad \text{para } i = 1, 2, \dots, k \text{ e } j = 1, 2, \dots, n_i \quad (3.34)$$

onde:

$$\hat{\alpha}_{ij} = \left[ 1 - \frac{\exp\{\mathcal{X}_{(j)i} \hat{\beta}_i\}}{\sum_{l \in R(t(j)i)} \exp\{\mathcal{X}_l \hat{\beta}_i\}} \right] \exp\{-\mathcal{X}_{(j)i} \hat{\beta}_i\}, \quad (3.36)$$

quando somente uma falha ocorre em  $t_{(j)i}$ , isto é,  $d_{ij} = 1$ . Se  $d_{ij} > 1$ , existindo portanto empates, tem-se:

$$\hat{\alpha}_{ij} = \exp \left[ \frac{-d_{ij}}{\sum_{l \in R(i;j)} \exp \{ \underline{x}_l \hat{\beta}_i \}} \right] \quad \begin{array}{l} j = 1, 2, \dots, n_i \\ i = 1, 2, \dots, k \end{array} \quad (3.37)$$

onde:  $d_{ij}$  é o número de tempos  $t_{ij}$  empatados e  $t_{(1)i}, t_{(2)i}, \dots, t_{(n_i)i}$  são os  $n_i$  tempos observados de falha pela causa  $i$  ordenados.

OBS : Os  $\hat{\beta}_i, i = 1, 2, \dots, n$  são obtidos pelo método da verossimilhança parcial.

O plot de  $-\log \hat{S}(\hat{e})$  versus  $\hat{e}$  pode ser usado, por exemplo, para investigar se algum modelo de riscos proporcionais paramétrico é apropriado.

### 3.6 - ESTIMAÇÃO DOS ELEMENTOS DE $\beta$ POR MÁXIMA VEROSSIMILHANÇA

#### 3.6.1 - Modelos de regressão paramétricos

Quando um modelo de regressão paramétrico do tipo log-linear (3.27) ou de riscos proporcionais (3.28) é assumido para as funções risco de causa específica  $\lambda_i(t | \underline{x})$  deve-se utilizar uma apropriada função de verossimilhança para a obtenção das estimativas dos  $\hat{\beta}_i, i = 1, 2, \dots, n$ .

Segundo Elandt-Johnson pode-se obter essa função de verossimilhança do seguinte modo: considere,

- $C_1, C_2, \dots, C_k$  as  $k$  causas de falha investigadas no período  $[0, \delta)$ ,
- $n$  o número de indivíduos em observação,
- $w$  o número de censuras,
- $d_i$  o número de falhas pela causa  $i$ ,
- $t_{ir}$  o tempo de falha do  $r$ -ésimo indivíduo que falhou pela causa  $i, r = 1, 2, \dots, d_i, i = 1, 2, \dots, k$ ,
- $t_{wr}$  o tempo do  $r$ -ésimo indivíduo que censurou.

Obs :  $d_1 + d_2 + \dots + d_k + w = n$ .



Seja  $\lambda_i(t | \mathcal{X})$  a função taxa de risco para a causa  $C_i$ .  
Assumindo que as causas de falha são mutuamente exclusivas tem-se que :

$$\lambda_1(t | \mathcal{X}) + \lambda_2(t | \mathcal{X}) + \dots + \lambda_k(t | \mathcal{X}) = \lambda_T(t | \mathcal{X}), \quad (3.38)$$

onde  $\lambda_T(t | \mathcal{X})$  é denominada função risco total, isto é, é a função risco considerando todas as causas. Segue então que a função de sobrevivência total  $S_T(t | \mathcal{X})$  é dada por :

$$\begin{aligned} S_T(t | \mathcal{X}) &= \exp \left[ - \int_0^t \lambda_T(u | \mathcal{X}) du \right] = \exp \left[ - \int_0^t \sum_{i=1}^k \lambda_i(u | \mathcal{X}) du \right] = \\ &= \exp \left[ - \sum_{i=1}^k \int_0^t \lambda_i(u | \mathcal{X}) du \right] = G_1(t | \mathcal{X}) G_2(t | \mathcal{X}) \dots G_k(t | \mathcal{X}), \end{aligned} \quad (3.39)$$

$$\text{onde : } G_i(t | \mathcal{X}) = \exp \left[ - \int_0^t \lambda_i(u | \mathcal{X}) du \right] \quad i = 1, 2, \dots, k.$$

Devido a (3.38) isto significa que,  $S_T(t | \mathcal{X})$  pode ser fatorada no tempo  $t$ .

Também, a função de densidade de probabilidade do tempo de falha da causa  $C_i$  na presença de todas as outras causas de falha pode ser expressa por :

$$f_i(t | \mathcal{X}) = \lambda_i(t | \mathcal{X}) S_T(t | \mathcal{X}) = \lambda_i(t | \mathcal{X}) \prod_{m=1}^k G_m(t | \mathcal{X}), \quad i = 1, 2, \dots, k. \quad (3.40)$$

Deste modo, a contribuição do  $r$ -ésimo indivíduo que falhou pela causa  $C_i$ , para a verossimilhança é :

$$\lambda_i(t_{ir} | \mathcal{X}) \prod_{m=1}^k G_m(t_{ir} | \mathcal{X}). \quad (3.41)$$

Consequentemente, a contribuição dos  $d_i$  indivíduos que falharam pela causa  $C_i$  para a verossimilhança é :

$$\prod_{r=1}^{d_i} \left[ \lambda_i(t_{ir} | \mathcal{X}) \prod_{m=1}^k G_m(t_{ir} | \mathcal{X}) \right], \quad (3.42)$$

e a contribuição das censuras para a verossimilhança é :

$$\prod_{r=1}^w \left[ \prod_{m=1}^k G_m ( t_{wr} | \mathcal{X} ) \right]. \tag{3.43}$$

Logo, a verossimilhança completa é expressa por:

$$\begin{aligned} L(\underline{\beta}) &= \prod_{i=1}^k \left\{ \prod_{r=1}^{d_i} \left[ \lambda_i ( t_{ir} | \mathcal{X} ) \prod_{m=1}^k G_m ( t_{ir} | \mathcal{X} ) \right] \right\} \prod_{r=1}^w \left[ \prod_{m=1}^k G_m ( t_{wr} | \mathcal{X} ) \right] \\ &= \prod_{i=1}^k \left\{ \left[ \prod_{r=1}^{d_i} \lambda_i ( t_{ir} | \mathcal{X} ) \right] \left[ \prod_{l=1}^n G_i ( t_l | \mathcal{X} ) \right] \right\}. \end{aligned} \tag{3.44}$$

Agora, se  $\lambda_i ( t | \mathcal{X} )$  depende somente dos parâmetros  $\beta_i$  e estes parâmetros não estão presentes em nenhuma outra função de taxa de risco, então os parâmetros  $\beta_i$  podem ser estimados usando somente o i-ésimo fator de (3.44), isto é:

$$\begin{aligned} &\left[ \prod_{r=1}^{d_i} \lambda_i ( t_{ir} | \mathcal{X} ) \right] \left[ \prod_{l=1}^n G_i ( t_l | \mathcal{X} ) \right] = \\ &= \left[ \prod_{r=1}^{d_i} \lambda_i ( t_{ir} | \mathcal{X} ) \right] \left\{ \prod_{l=1}^n \exp \left[ - \int_0^{t_l} \lambda_i ( u | \mathcal{X} ) du \right] \right\}, \end{aligned} \tag{3.45}$$

ou seja, usando somente os tempos de falha da causa  $C_i$  e tratando os  $(n - d_i)$  indivíduos como censurados,  $i = 1, 2, \dots, k$ .

Alternativamente, Kalbfleisch e Prentice obtiveram uma função de verossimilhança, a qual é equivalente a função de verossimilhança (3.44) obtida por Elandt - Johnson. Tal função de verossimilhança é obtida do seguinte modo:

Suponha que associado a cada um dos  $n$  indivíduos tem-se os dados  $(t_j, \delta_j, I_j, d_j)$ , onde:

$t_j$ : tempo observado para o  $j$ -ésimo indivíduo,

$\delta_j$ : indicador de censura, isto é,  $\delta_j = \begin{cases} 1 & \text{se } t_j \text{ é tempo de falha} \\ 0 & \text{se } t_j \text{ é tempo de censura,} \end{cases}$

$I_j$ : causa de falha, isto é,  $I_j = \begin{cases} i, & i = 1, 2, \dots, k \text{ se } t_j \text{ é tempo de falha} \\ 0 & \text{se } t_j \text{ é tempo de censura,} \end{cases}$

$d_j$ : dose (ou uma função da dose) que o  $j$ -ésimo indivíduo recebeu,  $j = 1, 2, \dots, n$ .

Sob um mecanismo de censura aleatória independente, a função de verossimilhança baseada em  $n$  observações independentes é escrita por:

$$L(\underline{\beta}) = \prod_{j=1}^n \prod_{i=1}^k \left\{ \lambda_i(t_j | d_j) \delta_{ij} \exp \left[ \int_0^{t_j} \lambda_i(t | d_j) dt \right] \right\}, \quad (3.46)$$

onde:  $\delta_{ij} = 1$  se  $t_j$  corresponde a falha do tipo  $i$ ,  
 $\delta_{ij} = 0$  caso contrário ( $j = 1, 2, \dots, n$ ;  $i = 1, 2, \dots, k$ ),  
 $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$  é o vetor de regressão.

Pode-se portanto observar que a contribuição para a função de verossimilhança de um indivíduo que recebeu dose  $d$  e é observado ter um tipo de falha  $I=i$  no tempo  $t$  é:

$$C = \lambda_i(t | d) \exp \left\{ - \int_0^t \lambda_i(u | d) du \right\} \quad (i = 1, 2, \dots, k). \quad (3.47)$$

Por outro lado, um indivíduo que recebeu a dose  $d$  e sobreviveu ao tempo  $t$  contribue para a função de verossimilhança com:

$$C = \exp \left\{ - \int_0^t \lambda_i(u | d) du \right\} \quad (i = 1, 2, \dots, k). \quad (3.48)$$

Deste modo, a função de verossimilhança (3.46) é o produto de termos da forma (3.47) e (3.48). Se  $d_i = \sum_{j=1}^n \delta_{ij}$  para  $i = 1, 2, \dots, k$ , a função de verossimilhança (3.46) torna-se exatamente a função de verossimilhança (3.44) obtida por Elandt-Johnson.

As estimativas dos  $\underline{\beta}_i$ ,  $i = 1, 2, \dots, k$  são então obtidas resolvendo-se as equações:

$$\frac{\partial \ln L(\underline{\beta})}{\partial \underline{\beta}_i} = 0 \quad (i = 1, 2, \dots, k). \quad (3.49)$$

Estas equações são em geral resolvidas pelo método iterativo de Newton-Raphson o qual fornece os estimadores de máxima verossimilhança para os  $\underline{\beta}_i$ ,  $i = 1, 2, \dots, k$ .

### 3.6.2 - Modelo semi-paramétrico

Quando o modelo semi-paramétrico de riscos proporcionais de Cox é assumido, o método da verossimilhança parcial pode ser utilizado para encontrar as estimativas dos  $\beta_i$ ,  $i = 1, 2, \dots, k$ .

Seja  $t_{i(1)}, t_{i(2)}, \dots, t_{i(n_i)}$  os tempos de falha ordenados e observados ocorrer pela causa  $i$ ,  $i = 1, 2, \dots, k$ , e seja  $d_{i(j)}$ ,  $j = 1, 2, \dots, n_i$  e  $i = 1, 2, \dots, k$ , a variável concomitante dose ou uma função da dose associada ao indivíduo que falhou em  $t_{i(j)}$ . Considere ainda  $R(t_{i(j)})$  como sendo o conjunto dos indivíduos em risco no tempo  $t_{i(j)}$ , ou seja, o conjunto de indivíduos vivos e não censurados em  $t_{i(j)}$ .

A "função de verossimilhança" em uma amostra de tamanho  $n$  é então dada por :

$$L(\beta) = \prod_{i=1}^k \prod_{j=1}^{n_i} \frac{\exp \{ \mathcal{X}_{i(j)} \beta_i \}}{\sum_{l \in R(i(j))} \exp \{ \mathcal{X}_{l} \beta_i \}} \quad (3.50)$$

A verossimilhança (3.50) não depende dos  $\lambda_{oi}(\cdot)$ ,  $i = 1, 2, \dots, k$ . Sua maximização fornece as estimativas dos  $\beta_i$ , isto é, a solução das equações de máxima verossimilhança :

$$\frac{\partial \ln L(\beta)}{\partial \beta_i} = 0, \quad i = 1, 2, \dots, k, \quad (3.51)$$

produzem as estimativas dos  $\beta_i$ ,  $i = 1, 2, \dots, k$ . Em geral as equações (3.51) são resolvidas pelo método iterativo de Newton-Raphson.

Se existirem tempos empatados dentre os  $n_i$ ,  $i = 1, 2, \dots, k$  indivíduos falhando pela causa  $i$ , a verossimilhança a ser usada é dada por :

$$L(\beta) = \prod_{i=1}^k \prod_{j=1}^{n_i} \frac{\exp \{ S_{ij} \beta_i \}}{\left( \sum_{l \in R(i(j))} \exp \{ \mathcal{X}_l \beta_i \} \right)^{d_{ij}}}, \quad (3.52)$$

onde:  $d_{ij} = \#$  de tempos iguais a  $t_{i(j)}$ ,

$S_{ij} = \sum_{p \in D_{i(j)}} \mathcal{X}_p$  com  $D_{i(j)}$  sendo o conjunto de indivíduos que falharam em  $t_{i(j)}$ ,

ou seja,  $S_{i(j)}$  é a soma das variáveis concomitantes  $\mathcal{X}_p$  para os  $d_{i(j)}$  indivíduos com tempos empatados em  $t_{i(j)}$ .

3.7 - TESTES DE HIPÓTESES

Em experimentos onde são considerados riscos competitivos, pode ser de interesse testar hipóteses relacionadas a uma particular variável concomitante. Lagakos(1977) considera algumas destas hipóteses para o modelo log-linear quando  $T_1, T_2, \dots, T_k$  são variáveis aleatórias i. i. d. Exponencial  $(\lambda_j)$ ,  $j = 1, 2, \dots, k$ .

Considerando os experimentos de dose-resposta, a influência da variável concomitante dose pode ser de interesse nos modelos log-linear, de riscos proporcionais e de riscos proporcionais de Cox em um dos seguintes modos:

1) Testar se a dose não tem influência no tipo de falha e nem no tempo até a falha, ou seja,

$$\begin{cases} H_{o1}: \beta_1^{(1)} = \beta_1^{(2)} = \dots = \beta_1^{(k)} = 0 \\ H_{A1}: \text{pelo menos um } \beta_1^{(j)} \neq 0, \end{cases}$$

onde  $\beta_1^{(j)}$  é o coeficiente de regressão associado a dose e causa  $j$ ,  $j = 1, 2, \dots, k$ , ou alternativamente,

2) Testar primeiro se a dose não tem influência no tipo de falha, isto é,

$$\begin{cases} H_{o2}: \beta_1^{(1)} = \beta_1^{(2)} = \dots = \beta_1^{(k)} \\ H_{A2}: \text{pelo menos um } \beta_1^{(j)} \text{ diferente} \end{cases}$$

Se os dados suportarem tal hipótese, pode-se então testar a hipótese condicional de que a dose não tem influência no tempo de falha dado que não tem influência no tipo de falha, isto é,

$$\begin{cases} H_{o3}: \beta_1^{(1)} = \beta_1^{(2)} = \dots = \beta_1^{(k)} = 0 \text{ dado que } \beta_1^{(1)} = \beta_1^{(2)} = \dots = \beta_1^{(k)} \\ H_{A3}: \text{pelo menos um } \beta_1^{(j)} \neq 0 \text{ dado que } \beta_1^{(1)} = \beta_1^{(2)} = \dots = \beta_1^{(k)}, \end{cases}$$

Testes baseados na aproximação  $\chi^2$  para cada uma dessas hipóteses podem ser obtidos comparando os maximizados logs das verossimilhanças como é feito usualmente, isto é, usando:

$$R = -2 \ln \left( \frac{\max_{\tilde{\beta} \in \tilde{\theta}_o} L(\tilde{\beta})}{\max_{\tilde{\beta} \in \tilde{\theta}} L(\tilde{\beta})} \right).$$

Se  $R_i$  denota o resultado da estatística de teste correspondente a  $H_{oi}$ ,  $i = 1, 2, 3$  segue que  $R_1 = R_2 + R_3$  e que as distribuições assintóticas sob  $H_{oi}$  de  $R_1$ ,  $R_2$  e  $R_3$  são respectivamente  $\chi^2_k$ ,  $\chi^2_{k-1}$  e  $\chi^2_1$ .

Embora testes baseados em  $R_i$  sejam conceitualmente diretos, a computação associada pode tornar-se excessiva, e como em geral as análises envolvem diversas variáveis concomitantes, esta aproximação é usualmente não prática. A equivalência assintótica  $Q_1$ ,  $Q_2$  e  $Q_3$  (abaixo) de  $R_1$ ,  $R_2$  e  $R_3$  à distribuição Normal é portanto mais conveniente em situações cujo tamanho amostral seja não muito pequeno.

Especificamente, se  $\hat{\sigma}_{11}^{(j)}$  representa a variância estimada de  $\beta_1^{(j)}$ ,  $j = 1, 2, \dots, k$  tem-se as seguintes expressões para os  $Q_i$ ,  $i = 1, 2, 3$ :

$$Q_1 = \sum_{j=1}^k \left[ \left( \hat{\beta}_1^{(j)} \right)^2 / \hat{\sigma}_{11}^{(j)} \right],$$

$$Q_3 = \left( \sum_{j=1}^k \hat{\beta}_1^{(j)} / \hat{\sigma}_{11}^{(j)} \right)^2 / \sum_{j=1}^k \left( \hat{\sigma}_{11}^{(j)} \right)^{-1}$$

e  $Q_2 = Q_1 - Q_3$ .

Análises baseadas no uso dos estimadores  $Q_1$ ,  $Q_2$  e  $Q_3$  requer somente uma única maximização da função de verossimilhança, independente do número de variáveis concomitantes sendo tratadas.

Hipóteses alternativas restritivas como:

$$H_A : 0 \leq \beta_1^{(1)} < \beta_1^{(2)} < \dots < \beta_1^{(k)}$$

ou

$$H_A : \beta_1^{(1)} < \beta_1^{(2)} < \dots < \beta_1^{(k)},$$

para os modelos tratados podem ser de interesse. Estas hipóteses encontram-se devidamente tratadas no artigo de Wada, Sen e Tarumoto (1993).

# CAPÍTULO 4

## APLICAÇÕES NUMÉRICAS

### 4.1 - INTRODUÇÃO

Neste capítulo são apresentadas aplicações numéricas com a finalidade de ilustrar a teoria exposta nos capítulos 2 e 3.

Inicialmente é realizada na seção 4.2 uma aplicação numérica enfocando os modelos de regressão em um contexto de um único risco. Para essa aplicação foi gerado aleatoriamente, com o auxílio computacional e usando o método da transformação inversa (Kennedy-1980), um conjunto de tempos de vida com distribuição Weibull ( $\exp\{\underline{x} \underline{\beta}\}; \delta$ ). Para cada tempo gerado é conhecido se o mesmo corresponde a uma falha ou a uma censura, bem como o valor de sua respectiva variável concomitante (dose ou uma função da mesma). Neste conjunto de tempos foi então ajustado os modelos de regressão citados no Capítulo 2. Para as estimativas dos coeficientes de regressão  $\underline{\beta}$ , foi utilizado o método da máxima verossimilhança, sendo que as equações de máxima verossimilhança foram resolvidas pelo método iterativo de Newton-Raphson. A

análise residual gráfica foi então realizada para a verificação do ajuste dos modelos. Ainda, a título de ilustração, foi obtido para um determinado tempo  $t$ , doses virtualmente seguras e o tempo médio para a resposta.

Na seção 4.3 é apresentada uma segunda aplicação numérica com o intuito de focar os modelos de regressão em um contexto de riscos competitivos. A título de ilustração, foi considerado  $k = 2$  (duas) causas de falha. Para cada uma das causas de falha foi gerado tempos com distribuição Weibull ( $\exp\{-x^{\beta_i}\}; \delta_i$ ) para  $i = 1, 2$ , obtendo-se em seguida um conjunto correspondendo aos mínimos de cada  $i$ -ésimo par de tempos gerados. A cada um desses tempos mínimos foi associado a causa de falha (0, 1 ou 2) e o respectivo valor da variável concomitante. Causa = 0 indica que o tempo observado corresponde a uma censura, isto é, não ocorreu o evento de interesse até o término pré-fixado do experimento. A esse conjunto de tempos foi então ajustado os modelos de regressão abordados no Capítulo 3. Os procedimentos descritos nas seções do referido capítulo foram então utilizados para a obtenção das estimativas de máxima verossimilhança, bem como para realizar a análise residual gráfica.

É conveniente salientar que a escolha da distribuição Weibull para a geração dos tempos, deve-se ao fato de tal distribuição permitir o ajuste de todos os modelos abordados e sendo assim, foi possível ilustrar os modelos tratados nos capítulos 2 e 3, respectivamente. Além disso, pôde-se também realizar comparações entre os modelos paramétricos e o modelo semi-paramétrico tratados nesta dissertação.

Para a obtenção dos dois conjuntos de dados mencionados nas aplicações numéricas, bem como das estimativas dos coeficientes de regressão, dos valores residuais e de todos os demais cálculos e gráficos necessários, utilizou-se o software estatístico SAS - versão 6.08. Os programas feitos no SAS para a obtenção dos resultados apresentados no decorrer deste capítulo, encontram-se no Apêndice desta dissertação com as respectivas informações necessárias aos interessados em reproduzir as aplicações aqui realizadas.



## 4.2 - APLICAÇÃO NUMÉRICA EM UM CONTEXTO DE UM ÚNICO RISCO DE FALHA

### 4.2.1 - Descrição dos dados gerados

Para ilustrar os modelos de regressão abordados no capítulo 2, foi gerado aleatoriamente um arquivo de 100 observações contendo as seguintes variáveis: tempo - dose - censura. Para cada um dos 4 (quatro) níveis da dose previamente fixados ( $d_1 = 0.3$ ,  $d_2 = 0.7$ ,  $d_3 = 0.8$ ,  $d_4 = 0.9$ ), foi gerado 25 tempos ( $n_1 = n_2 = n_3 = n_4 = 25$ ) com distribuição Weibull( $\exp \{ \underline{x} \underline{\beta} \}; \delta$ ) onde:

$$\underline{x} = (\underline{1}, \underline{d}_i)_{100 \times 2} ; \underline{\beta} = (\beta_0, \beta_1)' = (3.9, 1.1)' \text{ e } \delta = 1.67 .$$

Observe-se que a semente utilizada na geração foi 15747. Tendo fixado  $T = 145$  como o tempo correspondente ao término do experimento, classificou-se os tempos gerados em falha ou censura onde:

$$\text{tempo} = t_i = \begin{cases} t_i & \text{se } t_i \leq 145 \\ T & \text{se } t_i > 145 \end{cases} \quad \text{e} \quad \text{censura} = \begin{cases} 1 & \text{se } t_i \leq 145 \\ 0 & \text{se } t_i > 145, \end{cases}$$

$i = 1, 2, \dots, 100.$

O conjunto de tempos após a classificação resultou em: 83 falhas e 17 censuras. O arquivo de observações aqui utilizado encontra-se no apêndice B-tabela B1 e o programa utilizado para a sua obtenção (PROG1.SAS) encontra-se no apêndice A.

### 4.2.2 - Modelos ajustados e estimação

Uma vez tendo gerado o referido arquivo, foram ajustados os 3 modelos de regressão: log-linear, riscos proporcionais Weibull e riscos proporcionais de Cox, descritos no capítulo 2. Para esses modelos, obteve-se através do método iterativo de Newton-Raphson, as seguintes estimativas de máxima verossimilhança:

Modelos de Regressão	estimativas	desvio-padrão	n <sup>o</sup> . iterações
Log - linear ( $Y = \tilde{x} \tilde{\beta} + \sigma W$ )	$\hat{\beta}_0 = 3.777$ $\hat{\beta}_1 = 1.2124$ $\hat{\sigma} = 0.5979$	dp = 0.1922 dp = 0.2780 dp = 0.0554	5
Riscos Prop. Weibull $\lambda(t \tilde{x}) = \delta t^{\delta-1} \exp\{-\tilde{x} \tilde{\beta}\}$	$\hat{\beta}_0 = 3.777$ $\hat{\beta}_1 = 1.2124$ $\hat{\delta} = 1.6725$	dp = 0.1922 dp = 0.2780 dp = 0.1551	5
Riscos Prop. de Cox $\lambda(t \tilde{x}) = \lambda_0(t) \exp\{d\beta_1\}$	$\hat{\beta}_1 = -2.03445$	dp = 0.4793	4

Tabela 4.1 - Estimativas obtidas através das PROC LIFEREG E PHREG do software estatístico SAS - versão 6.08.

#### 4.2.3 - Testes dos parâmetros e coeficientes de regressão estimados

No modelo log-linear, pode-se testar a hipótese  $H_{01}: \sigma = 1$  (o que implica, no caso de sua não rejeição, que o modelo Exponencial é mais adequado do que o modelo Weibull). Uma estatística a qual pode ser usada para testar  $H_{01}$  é:

$$z^2 = \left( \frac{\hat{\sigma} - 1}{dp(\hat{\sigma})} \right)^2,$$

que possui aproximadamente distribuição  $\chi^2_{(1)}$ . Na aplicação aqui tratada  $z^2 \cong 52.55$ , o qual é significativo na distribuição  $\chi^2_{(1)}$ . Assim, há evidências estatísticas contra o modelo Exponencial e conseqüentemente favoráveis ao modelo Weibull.

No modelo de Riscos Proporcionais Weibull, a hipótese equivalente à  $H_{01}$  é  $H_{02} : \delta = 1$ , uma vez que  $\delta = 1/\sigma$ .

Alternativamente, pode-se usar o teste da razão da verossimilhança (dado em 2.48), para testar  $H_{01}$  e/ou  $H_{02}$ .

Testes relacionados aos coeficientes de regressão também podem ser baseados na estatística  $z^2$  ou na estatística da razão da verossimilhança. Para esta aplicação, utilizou-se a estatística  $z^2$  pois com  $n = 100$  tem-se uma boa aproximação dos elementos de  $\hat{\beta}$  à distribuição Normal. Os resultados obtidos foram:

Modelos de Regressão	Hipóteses	$z^2$	g.l.	p-valor
Log-linear e Riscos	$H_{01} : \beta_0 = 0$	386.27	1	0.0001
Proporcionais Weibull	$H_{02} : \beta_1 = 0$	19.02	1	0.0001
Riscos Proporcionais de Cox	$H_{03} : \beta_1 = 0$	18.02	1	0.0001

Tabela 4.2 - Resultados baseados na Normalidade assintótica de  $\hat{\beta}$ .

Pode-se portanto concluir, baseados nos resultados da tabela 4.2, que existem evidências, estatisticamente significativas, contrárias as hipóteses  $H_{01}$ ,  $H_{02}$  e  $H_{03}$ , ou seja, os elementos de  $\hat{\beta}$  são significativamente diferentes de zero.

#### 4.2.4 - Análise Residual

Com relação ao ajuste dos modelos, utilizou-se a análise residual gráfica abordada na seção 2.5. Para cada um dos modelos de regressão ajustados obteve-se as seguintes análises residuais :

### I - Modelo de regressão log-linear

Neste caso, supõe-se que a distribuição dos erros  $W$  é Valor-extremo padrão, o que implica que :

$$f_W(w) = \exp \{ w - \exp(w) \} \quad ; \quad -\infty < w < \infty \quad (4.1)$$

e

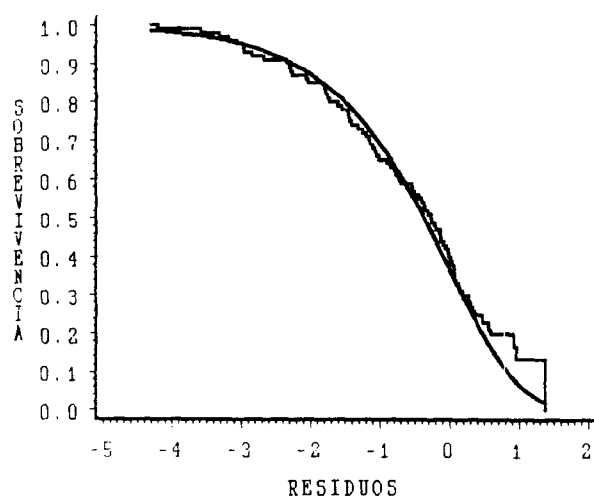
$$S_W(w) = \exp \{ -\exp(w) \} \quad ; \quad -\infty < w < \infty \quad (4.2)$$

são, respectivamente, as funções de densidade de probabilidade e sobrevivência de  $W$ .

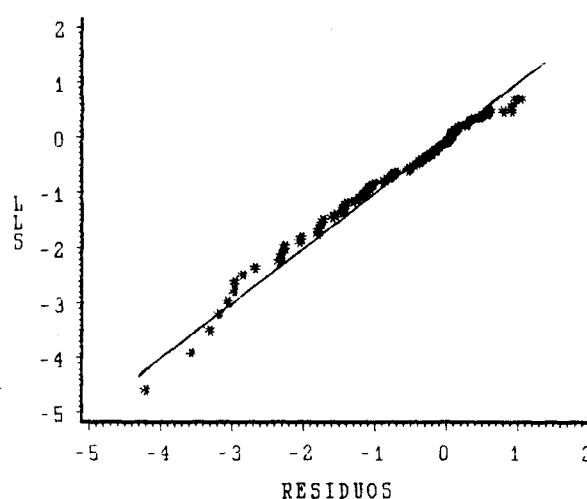
Consequentemente, tem-se de (4.2) a seguinte igualdade:

$$w = \ln ( -\ln ( S_W(w) ) ) \quad ; \quad -\infty < w < \infty . \quad (4.3)$$

Desse modo, os resíduos  $\hat{\epsilon}_i$ ,  $i = 1, 2, \dots, 100$  devem aproximadamente satisfazer (4.1), (4.2) e/ou (4.3) para que o modelo suposto possa ser considerado adequado. Graficamente tem-se :



**Figura 4.1** - Função de sobrevivência estimada e função de sobrevivência suposta para os resíduos.



**Figura 4.2** - Resíduos ( $\hat{\epsilon}_i$ ) versus  $\ln(-\ln \hat{S}(\hat{\epsilon}_i))$  e reta com inclinação 1.

Pela figura 4.1, pode-se observar que a curva de sobrevivência estimada dos resíduos (escada) ajusta-se bem à função de sobrevivência suposta para os erros  $W$  e pela figura 4.2 observa-se claramente que aos pontos

$(\hat{e}_i; \ln(-\ln \hat{S}(\hat{e}_i)))$  pode-se ajustar uma reta com inclinação 1.

Conclui-se portanto, por essa análise residual gráfica, que os resíduos  $\hat{e}_i$ ,  $i = 1, 2, \dots, n$  podem ser considerados como sendo uma amostra aleatória proveniente da distribuição Valor-extremo padrão. Conseqüentemente, o modelo de regressão log-linear pode ser considerado como ajustando-se bem aos dados desta aplicação.

## II - Modelo de Riscos Proporcionais Weibull

De acordo com a teoria apresentada na seção 2.5.1, foi visto que quando é assumido a distribuição Weibull aos tempos de vida  $T | \underline{x}$ , os erros  $W^*$  assumem a distribuição Exponencial padrão. Segue então que:

$$f_{W^*}(w) = \exp \{ -w \} \quad ; \quad w \geq 0 \quad (4.4)$$

e

$$S_{W^*}(w) = \exp \{ -w \} \quad ; \quad w \geq 0 \quad (4.5)$$

são, respectivamente, as funções de densidade de probabilidade e sobrevivência de  $W^*$ .

De (4.5) tem-se válida a seguinte igualdade :

$$w = -\ln ( S_{W^*}(w) ) \quad ; \quad w \geq 0 \quad (4.6)$$

Assim, para que o modelo suposto seja considerado adequado, é necessário que os resíduos  $\hat{e}_i^*$ ,  $i = 1, 2, \dots, 100$  satisfaçam aproximadamente (4.4), (4.5) e/ou (4.6). Graficamente tem-se :

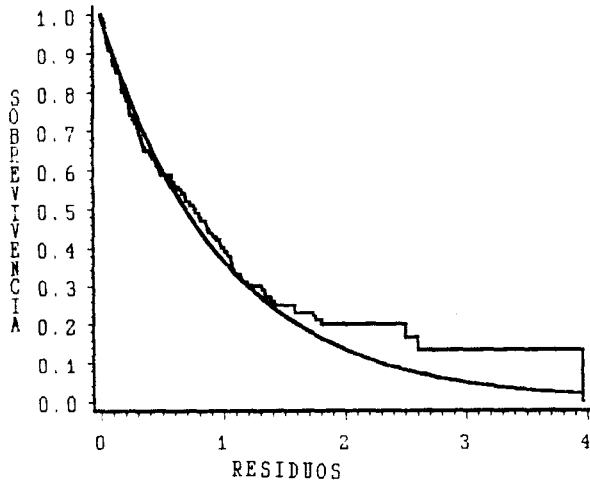


Figura 4.3 - Função de sobrevivência estimada e função de sobrevivência suposta os resíduos.

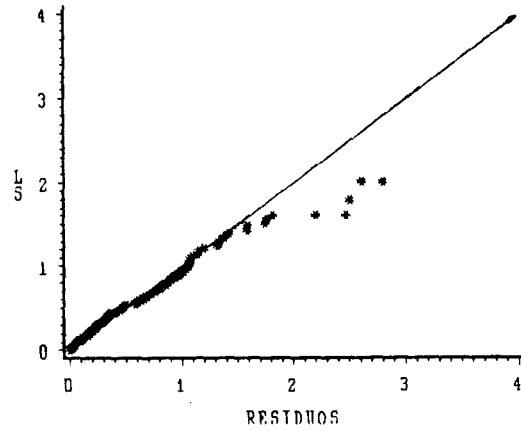


Figura 4.4 - Resíduos ( $\hat{\epsilon}_i^*$ ) versus  $-\ln(\hat{S}(\hat{\epsilon}_i^*))$  e reta de inclinação 1.

Nota-se portanto, que os resíduos  $\hat{\epsilon}_i^*$ ,  $i = 1, 2, \dots, 100$  podem ser vistos como sendo uma amostra aleatória proveniente da distribuição Exponencial padrão, uma vez que a função de sobrevivência estimada dos  $\hat{\epsilon}_i^*$ ,  $i = 1, 2, \dots, n$  (obtida pelo P.L. de Kaplan-Meier), aproxima-se consideravelmente bem da função de sobrevivência suposta para os erros  $W^*$ , conforme pode-se observar na figura 4.3. Ainda, pode-se verificar na figura 4.4, que aos pontos  $(\hat{\epsilon}_i^* ; -\ln(\hat{S}(\hat{\epsilon}_i^*)))$ ,  $i = 1, 2, \dots, 100$  pode-se ajustar uma reta com inclinação 1, satisfazendo assim a igualdade dada em (4.6). Logo, o modelo de regressão de riscos proporcionais Weibull pode ser considerado um modelo adequado para os dados aqui analisados.

### III - Modelo de Riscos Proporcionais de Cox

Os resíduos para este modelo, conforme visto na seção 2.5.2, podem ser obtidos por :

$$\hat{\epsilon}_i = \begin{cases} \hat{H}(t_i | \mathbf{x}_i) & \text{se } i \in F \\ \hat{H}(t_i | \mathbf{x}_i) + 1 & \text{se } i \in C, \end{cases}$$

o que é equivalente a:

$$\hat{\epsilon}_i = \begin{cases} [-\ln \hat{S}_0(t_i)] \exp\{\mathcal{X}_i \hat{\beta}\} & \text{se } i \in F \\ [-\ln \hat{S}_0(t_i^*)] \exp\{\mathcal{X}_i \hat{\beta}\} + 1 & \text{se } i \in C, \end{cases}$$

para  $i = 1, 2, \dots, 100$ . Ainda foi visto que o conjunto completo dos resíduos  $\hat{\epsilon}_i$ ,  $i = 1, 2, \dots, 100$  podem ser tratados "grosseiramente" como sendo uma amostra aleatória proveniente da distribuição Exponencial padrão.

Investigando então, se esta suposição é válida para os resíduos aqui encontrados, tem-se os seguintes gráficos:

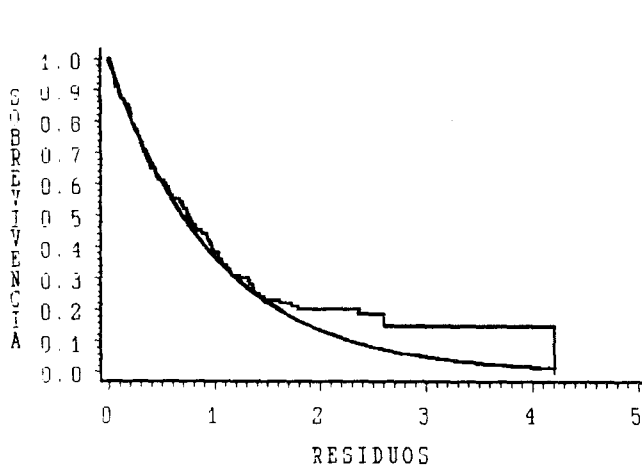


Figura 4.5 - Funções de sobrevivência estimada e suposta para os resíduos.

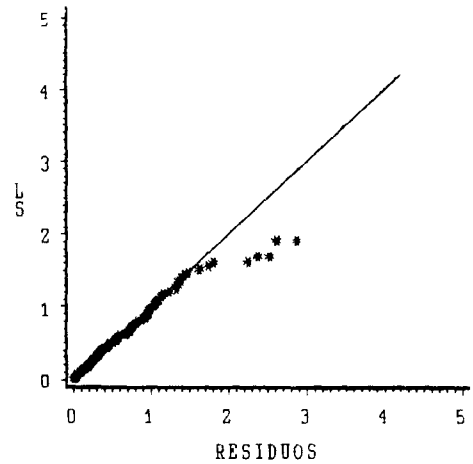


Figura 4.6 - Resíduos ( $\hat{\epsilon}_i$ ) versus  $-\ln(\hat{S}(\hat{\epsilon}_i))$  e reta de inclinação 1.

Observa-se portanto pelas figuras 4.5 e 4.6 que os resíduos podem realmente ser vistos como sendo uma amostra aleatória proveniente da distribuição Exponencial padrão. Logo, pode-se concluir que o modelo de Cox também é apropriado aos dados desta aplicação.

O modelo de Cox pode evidentemente ser usado, uma vez que o mesmo fornece boas estimativas, porém deve-se preferencialmente optar (quando possível) por um modelo paramétrico, pois os mesmos possibilitam a extrapolação tanto para dose quanto para o tempo, enquanto que o modelo semi-paramétrico de Cox não permite a extrapolação para tempos superiores ou inferiores aos observados.

Os programas PROG3.SAS e PROG5.SAS do Apêndice A foram usados para obtenção dos resíduos aqui relatados. Os valores residuais para cada modelo encontram-se na tabela B2 do Apêndice B.

#### 4.2.5 - Funções de sobrevivência e risco estimadas, DVS e $t_{mR}$

Para cada um dos modelos ajustados para esta aplicação, tem-se as seguintes funções de sobrevivência e risco estimadas:

##### I - Modelos Log-linear e de Riscos Proporcionais Weibull

Para estes modelos, as respectivas funções são dadas por:

$$\hat{S}(t | \mathcal{X}) = \exp \left\{ - \left( \frac{t}{\exp\{\hat{\beta}_0 + \hat{\beta}_1 d\}} \right)^{\hat{\delta}} \right\} \quad t \geq 0 \quad (4.7)$$

e

$$\hat{\lambda}(t | \mathcal{X}) = \hat{\delta} t^{\hat{\delta}-1} \left( \exp\{-\hat{\beta}_0 - \hat{\beta}_1 d\} \right)^{\hat{\delta}} \quad t \geq 0 \quad (4.8)$$

onde, pela tabela 4.1, tem-se:  $\hat{\beta}_0 = 3.777$ ,  $\hat{\beta}_1 = 1.2124$  e  $\hat{\delta} = 1.67$ . Relembrando, dose pode assumir os níveis 0.3 - 0.7 - 0.8 e 0.9 ou qualquer outro valor, uma vez que a extrapolação para outros níveis pode ser feita. Graficamente, as curvas das sobrevivências estimadas e dos riscos estimados para as doses 0.3 - 0.7 - 0.8 e 0.9 são:



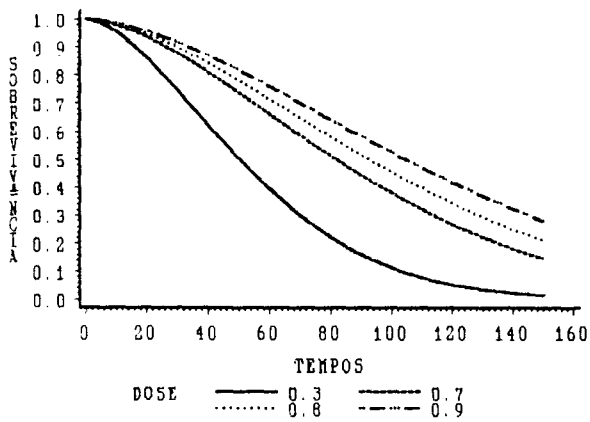


Figura 4.7 - Curvas das Sobrevivências estimadas pelos modelos log-linear e riscos prop. Weibull.

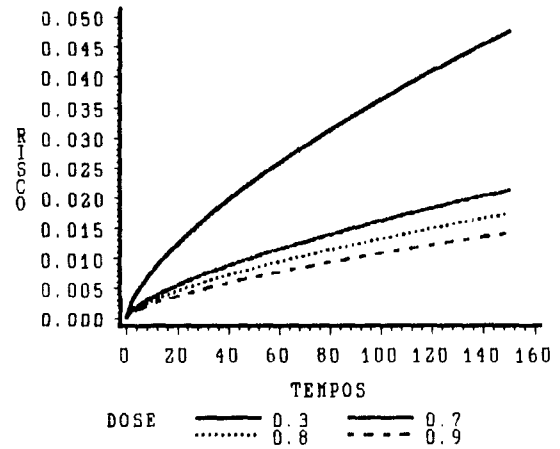


Figura 4.8 - Curvas dos riscos estimados pelos modelos log-linear e de riscos prop. Weibull.

Ainda, por exemplo, para  $t=110$  o "risco reduzido" sobre a resposta "background" obtido por :

$$\begin{aligned} \Pi(t = 110 | d) &= F(t = 110 | 0) - F(t = 110 | d) = \\ &= \exp\left\{-\left(110 / \exp\{3.77 + 1.21 d\}\right)^{1.67}\right\} - \exp\left\{-\left(110 / \exp\{3.77\}\right)^{1.67}\right\}, \end{aligned}$$

produz a seguinte curva:

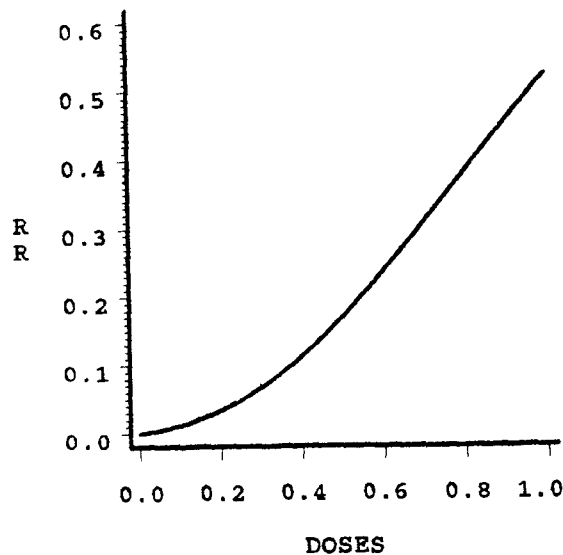


Figura 4.9 - Curva do risco reduzido sobre a resposta "background" para  $t=110$ .

A resposta "background"  $F(t | 0)$  produz, neste exemplo, uma probabilidade de 0.99117 de ocorrência da resposta na dose  $d = 0$ . A figura 4.9 nos fornece doses virtualmente seguras para desejadas reduções nesta probabilidade. Assim, por exemplo, se a redução desejada  $\pi$  é de 0.50, então a correspondente dose virtualmente segura é  $d_0 \cong 0.95$ . Portanto, se um indivíduo receber uma dose  $d_0 \cong 0.95$ , a probabilidade de ocorrer a resposta em  $t = 110$  é reduzida em 0.50 sobre a resposta "background". De modo análogo, pode-se obter valores de  $d_0$  para quaisquer reduções desejadas  $\pi$ , tal que  $0 < \pi < 0.99117$ .

O tempo médio estimado para a ocorrência da resposta para, por exemplo,  $t = 145$  e  $d = 0.8$  é encontrado ser:

$$t_{mR} = \int_0^{145} \exp\left\{-\left(\frac{u}{\exp\{\hat{\beta}_0 + \hat{\beta}_1\}}\right)^{1.67}\right\} du = \int_0^{145} \exp\left\{-\left(\frac{u}{114.43}\right)^{1.67}\right\} du$$

$$t_{mR} \cong 91.03.$$

Portanto, indivíduos que recebem dose 0.8 em um experimento planejado para terminar após 145 dias, produzirão a resposta em  $\pm 91$  dias. O cálculo do  $t_{mR}$  foi possível utilizando a quadratura gaussiana de ordem 7.

## II - Modelo de Riscos Proporcionais de Cox

Neste caso tem-se:

$$\hat{S}(t_i | \mathcal{X}_i) = [\hat{S}_0(t_i)]^{\exp\{d \hat{\beta}_1\}} \quad i = 1, 2, \dots, 100 \quad (4.9)$$

e

$$\hat{\lambda}(t_i | \mathcal{X}_i) = \hat{\lambda}_0(t_i) \exp\{d \hat{\beta}_1\} \quad i = 1, 2, \dots, 100 \quad (4.10)$$

onde pela tabela 4.2,  $\hat{\beta}_1 = -2.03445$ .

Graficamente, tem-se as seguintes curvas estimadas:

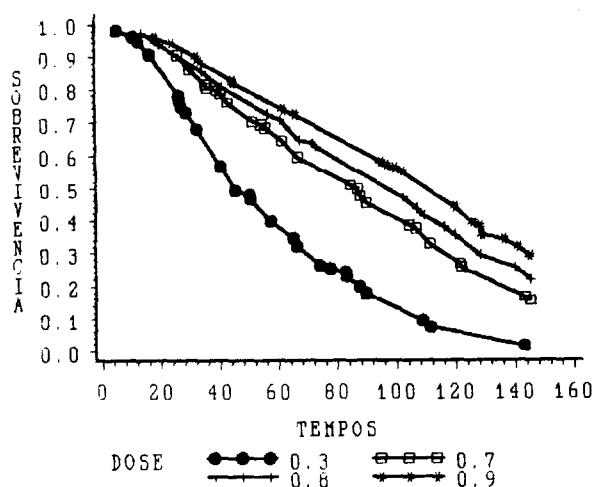


Figura 4.10 - Sobrevivências estimadas pelo modelo de Riscos Prop. de Cox.

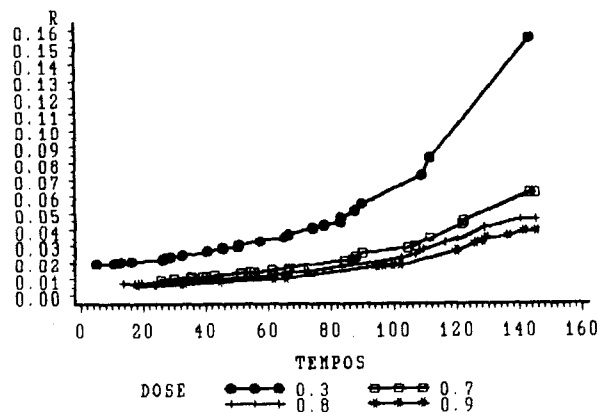


Figura 4.11 - Riscos estimados pelo modelo de Riscos Proporcional de Cox.

Neste modelo pode-se extrapolar para doses, isto é, pode-se estimar  $\hat{S}(t_i | \mathcal{X}_i)$  e  $\hat{\lambda}(t_i | \mathcal{X}_i)$  para outros valores da dose, porém não é permitido extrapolar para o tempo, ou seja, não pode-se estimar  $\hat{S}(t_i | \mathcal{X}_i)$  e  $\hat{\lambda}(t_i | \mathcal{X}_i)$  para tempos superiores ou inferiores aos observados.

Comparando as figuras 4.7 e 4.10 (referente as sobrevivências estimadas), pode-se observar que as estimativas encontram-se bastante próximas para cada um dos 4 (quatro) níveis da variável concomitante dose.

Os valores estimados de  $\hat{S}(t_i | \mathcal{X}_i)$  e  $\hat{\lambda}(t_i | \mathcal{X}_i)$ ,  $i = 1, 2, \dots, 100$  encontram-se na tabela B3 do Apêndice B. Os gráficos aqui apresentados, foram obtidos pela PROC GPLOT do software estatístico SAS - versão 6.08.

Doses virtualmente seguras podem ser obtidas, por exemplo, para  $t = 111,26$  por:

$$\Pi(t_i | d_0) = \pi,$$

onde:  $\Pi(t_i | d) = F(t_i | 0) - F(t_i | d) = [\hat{S}_0(t_i)]^{\exp\{d \hat{\beta}_1\}} - [\hat{S}_0(t_i)] =$   
 $= [0.009557]^{\exp\{-2.03445 d\}} - [0.009557].$

A curva obtida para o risco reduzido  $\Pi(t_i | d)$  é dada na figura 4.12 abaixo:

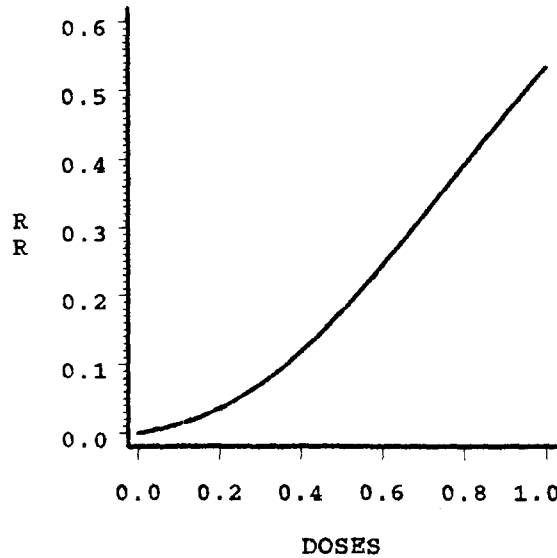


Figura 4.12 - Curva do risco reduzido sobre a resposta "background".

A resposta "background" ocorre com probabilidade  $p = F(t_i|0) = 0.990443$ . Se é desejado reduzir em  $\pi = 0.50$  tal probabilidade, então a dose a qual assegura esta redução é  $d_0 = 0.95$ , ou seja, um indivíduo que receba dose  $d_0 = 0.95$  terá sua probabilidade de falha reduzida em 0.50 no tempo  $t = 111,26$  dias.

O tempo médio estimado para a ocorrência da resposta considerando  $t = 145$  e  $d = 0.8$  é encontrado ser :

$$t_{mR} = \sum_{i: t_i \leq 145} [1 - \hat{F}(t_i | d=0.8)] = \sum_{i: t_i \leq 145} \hat{S}(t_i | d=0.8) \cong$$

$$= \sum_{i: t_i \leq 145} [\hat{S}(t_i)]^{\exp\{(-2.03445)(0.8)\}} \cong 58 \text{ dias.}$$

### 4.3 - APLICAÇÃO NUMÉRICA EM UM CONTEXTO DE RISCOS COMPETITIVOS

#### 4.3.1 Descrição dos dados gerados

Para ilustrar os modelos de regressão apresentados no Capítulo 3, foi gerado aleatoriamente um arquivo de 100 observações contendo as variáveis: tempo - dose - causa - censC1 - censC2.

Para a obtenção desse arquivo foi utilizado o PROG6.SAS, o qual encontra-se no apêndice A. Inicialmente foi gerado para a causa 1, 100 tempos com distribuição Weibull  $(\exp\{\underline{x} \underline{\beta}_{11}\}; \delta_1)$  onde:  $\underline{x} = (\underline{1} \quad \underline{d};)$ ,  $\underline{\beta}_{11} = (\beta_0, \beta_1)' = (4.0, 0.7)$  e  $\delta_1 = 2.0$  e outros 100 tempos com distribuição Weibull  $(\exp\{\underline{x} \underline{\beta}_{12}\}; \delta_2)$  para a causa 2 onde:  $\underline{x} = (\underline{1} \quad \underline{d};)$ ,  $\underline{\beta}_{12} = (3.9, 1.1)'$  e  $\delta_2 = 1.67$ . As sementes utilizadas na geração foram: 14767 e 19749. Em seguida, observou-se o mínimo de cada  $i$ -ésimo par gerado, isto é,  $\min(t_{1i}, t_{2i})$  para  $i = 1, 2, \dots, 100$ , obtendo-se assim um vetor dos mínimos dos tempos gerados. Fixado  $T = 150$ , representando o final do experimento, classificou-se os tempos em falha ou censura, sendo que:

$$\begin{cases} \text{se } t_i = \min(t_{1i}, t_{2i}) > 150 & \text{então } t_i = 150 \\ \text{se } t_i = \min(t_{1i}, t_{2i}) \leq 150 & \text{então } t_i = t_i \end{cases} \quad i = 1, 2, \dots, 100.$$

Estes tempos encontram-se na variável "tempo" do arquivo citado. A variável "dose" contém o nível pré-fixado para cada  $i$ -ésimo indivíduo, isto é,  $d = 0.3, 0.7, 0.8$  ou  $0.9$ . A variável "causa" assume os valores 0, 1 ou 2, onde:

$$\begin{aligned} \text{causa} &= 0 & \text{se } t_i = \min(t_{1i}, t_{2i}) > 150 \\ \text{causa} &= 1 & \text{se } t_i = \min(t_{1i}, t_{2i}) = t_{1i} \\ \text{causa} &= 2 & \text{se } t_i = \min(t_{1i}, t_{2i}) = t_{2i}, \end{aligned} \quad i = 1, 2, \dots, 100.$$

A variável "censC1" assume os valores 0 ou 1, sendo que:

$$\text{censC1} = \begin{cases} 0 & \text{se } \text{causa} = 0 \text{ ou } 2 \\ 1 & \text{se } \text{causa} = 1, \end{cases}$$

ou seja, considera-se censura para a causa 1 os tempos censurados ( $t_i > 150$ ) e os tempos de falha pela causa 2. Analogamente, para a variável "censC2" tem-se:

$$\text{censC2} = \begin{cases} 0 & \text{se causa} = 0 \text{ ou } 1 \\ 1 & \text{se causa} = 2, \end{cases}$$

isto é, as censuras para a causa 2 são os tempos censurados ( $t_i > 150$ ) e os tempos de falha pela causa 1.

Após esta classificação, o conjunto de tempos resultou em:

- 58 falhas e 42 censuras - considerando-se a causa 1 e
- 38 falhas e 62 censuras - considerando-se a causa 2.

Os valores gerados para o arquivo aqui mencionado, encontram-se no apêndice B - tabela B4.

#### 4.3.2 - Modelos ajustados e estimação

Uma vez obtido o arquivo de dados, ajustou-se para cada uma das 2 causas consideradas, os modelos de regressão log-linear, riscos proporcionais Weibull e riscos proporcionais de Cox. As estimativas de máxima verossimilhança obtidas para esses modelos pelo método de Newton-Raphson foram:

MODELO DE REGRESSÃO	CAUSA 1		CAUSA 2	
Log - linear	$\hat{\beta}_0 = 3.9303$	dp = 0.2403	$\hat{\beta}_0 = 3.838$	dp = 0.2339
	$\hat{\beta}_1 = 0.9643$	dp = 0.3287	$\hat{\beta}_1 = 1.4348$	dp = 0.3465
	$\hat{\sigma}_1 = 0.6029$	dp = 0.0662	$\hat{\sigma}_2 = 0.5446$	dp = 0.0726
Riscos Proporcionais	$\hat{\beta}_0 = 3.9303$	dp = 0.2403	$\beta_0 = 3.838$	dp = 0.2339
Weibull	$\hat{\beta}_1 = 0.9643$	dp = 0.3287	$\hat{\beta}_1 = 1.4348$	dp = 0.3465
	$\hat{\delta}_1 = 1.6586$	dp = 0.1821	$\hat{\delta}_2 = 1.8362$	dp = 0.2446
Riscos Proporcionais de Cox	$\hat{\beta}_1 = -1.7518$	dp = 0.60431	$\hat{\beta}_1 = -2.6713$	dp = 0.7122

Tabela 4.3 - Estimativas obtidas pelas PROC LIFEREG E PHREG do software SAS-versão 6.08.

Para os modelos log-linear e riscos proporcionais foi necessário 6 iterações e para o modelo de Cox 4 iterações.

#### 4.3.3 - Testes dos parâmetros e coeficientes de regressão estimados

Os testes relacionados aos parâmetros e coeficientes de regressão dos modelos ajustados para as causas 1 e 2 resultaram em:

Modelo de Regressão	Hipóteses	CAUSA 1			CAUSA 2		
		$z^2$	g.l.	p-valor	$z^2$	g.l.	p-valor
Log-linear e Riscos Proporcionais Weibull	$H_{01}: \sigma_1 = 1$	35.97	1	0.0001	39.27	1	0.0001
	$H_{02}: \beta_0 = 0$	267.44	1	0.0001	269.21	1	0.0001
	$H_{03}: \beta_1 = 0$	8.60	1	0.0034	17.14	1	0.0001
Riscos Proporcionais de Cox	$H_{04}: \beta_1 = 0$	8.40	1	0.0037	14.069	1	0.0002

Tabela 4.4 - Resultados baseados na Normalidade Assintótica de  $\hat{\beta}$ .

Conclui-se portanto pela tabela 4.4, que existem evidências estatísticas (para as duas causas) contra as hipóteses testadas, implicando assim em evidências favoráveis ao modelo Weibull no caso de  $H_{01}$  e nas demais hipóteses implicando que os coeficientes de regressão são estatisticamente diferentes de zero.

#### 4.3.4 - Análise Residual dos modelos ajustados

Analisando os resíduos obtidos para a causa 1 e causa 2 em cada um dos 3 modelos ajustados de forma análoga ao que foi feito na seção 4.2.4, tem-se :

##### I - Para o modelo log-linear

Para este modelo tem-se para as causas 1 e 2, respectivamente, as seguintes figuras :



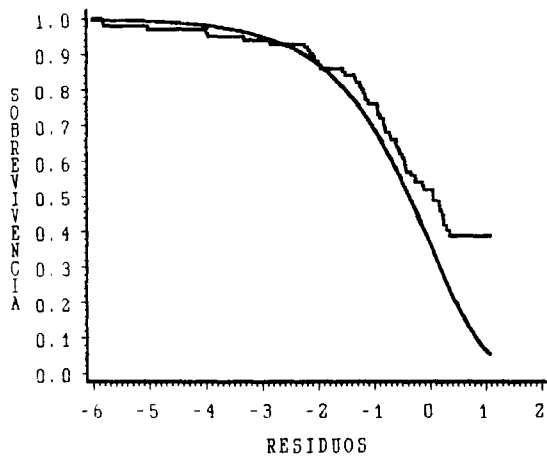


Figura 4.13 - Funções de sobrevivência estimada e suposta para os resíduos da causa 1.

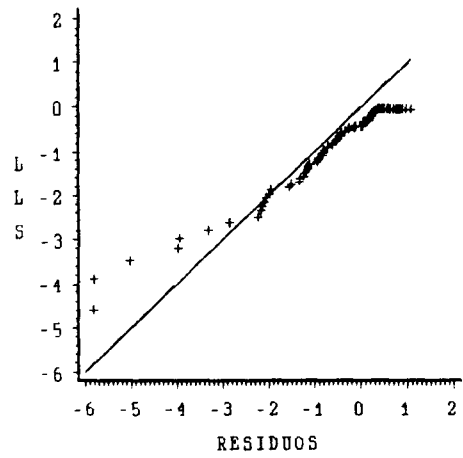


Figura 4.14 - Resíduos  $(\hat{\epsilon}_i)$  versus  $\ln(-\ln\hat{S}(\hat{\epsilon}_i))$  e reta com inclinação 1 para a causa 1.

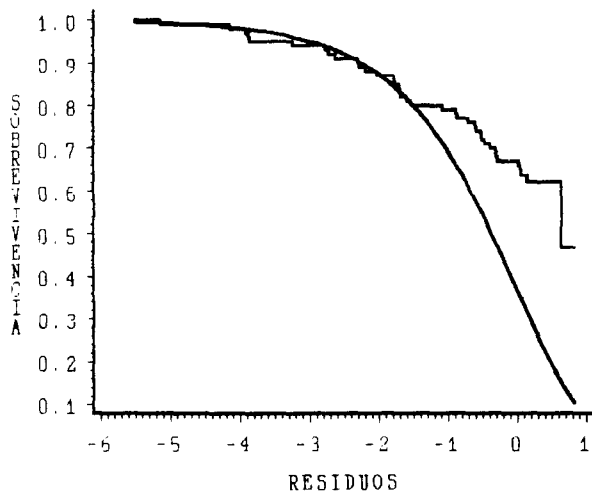


Figura 4.15 - Funções de sobrevivência estimada e suposta para os resíduos da causa 2.

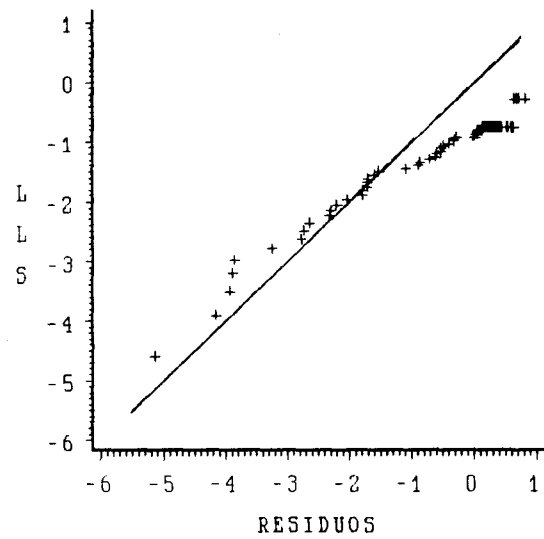


Figura 4.16 - Resíduos  $\hat{\epsilon}_i$  versus  $\ln(-\ln S(\hat{\epsilon}_i))$  e reta com inclinação 1 para a causa 2.

Pode-se portanto concluir para a causa 1, que os resíduos podem ser vistos grosseiramente como uma amostra proveniente da distribuição Valor-extremo padrão. O ajuste não foi melhor devido ao significativo número de censuras observado ( $\pm 40\%$ ).

Para a causa 2, pode-se observar que o ajuste não foi satisfatório. O grande número de censuras ocorridos pela causa 2 ( $\pm 60\%$ ) influenciou consideravelmente no mesmo.

## II - Para o modelo de riscos proporcionais Weibull

Para este modelo obteve-se as seguintes figuras para as causas consideradas:

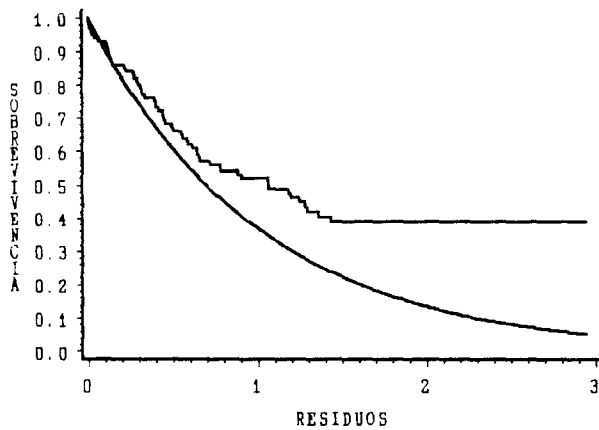


Figura 4.17 - Funções de Sobrevivência estimada e suposta para os resíduos da causa 1.

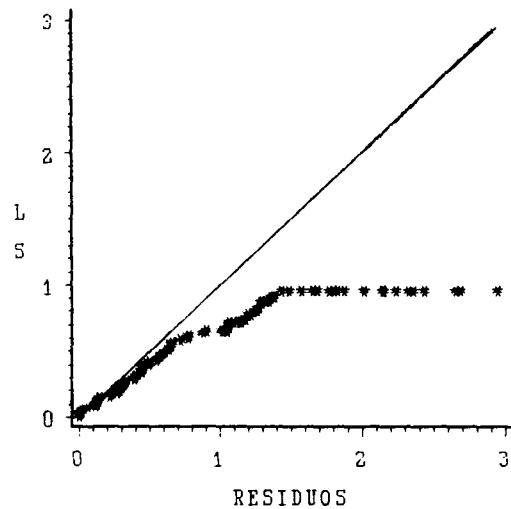


Figura 4.18 - Resíduos( $\hat{\epsilon}_i^*$ ) versus  $-\ln\hat{S}(\hat{\epsilon}_i^*)$  e reta com inclinação 1 para a causa 1.

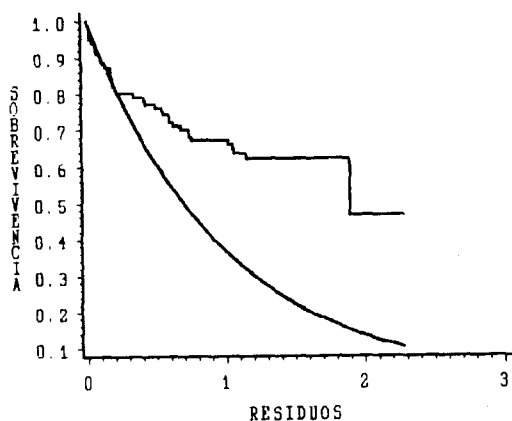


Figura 4.19 - Funções de sobrevivência estimada e suposta para os resíduos da causa 2.

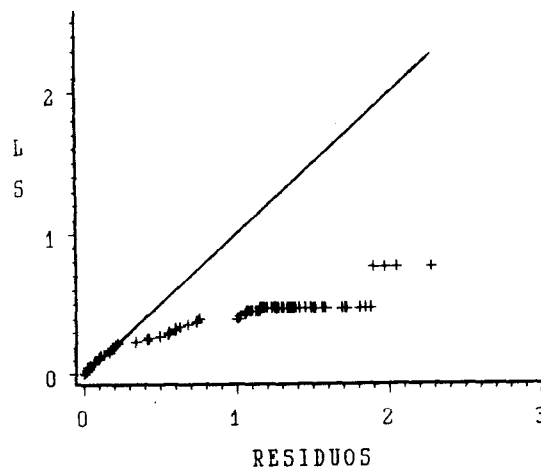


Figura 4.20 - Resíduos( $\hat{\epsilon}_i^*$ ) versus  $-\ln\hat{S}(\hat{\epsilon}_i^*)$  e reta com inclinação 1 para a causa 2.

Para este modelo, são válidas as mesmas conclusões obtidas para o modelo log-linear.

### III - Para o modelo de Riscos Proporcionais de Cox

Investigando se os resíduos obtidos no modelo de Cox para as causas 1 e 2, fornecem evidências de serem uma amostra aleatória proveniente da distribuição Exponencial Padrão, tem-se as seguintes figuras :

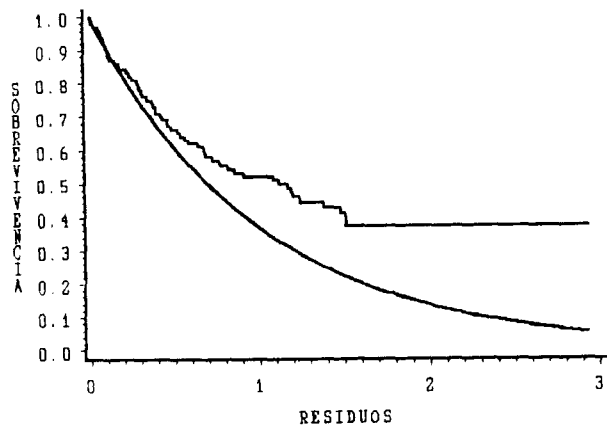


Figura 4.21 - Funções de Sobrevivência estimada e suposta para os resíduos da causa 1.

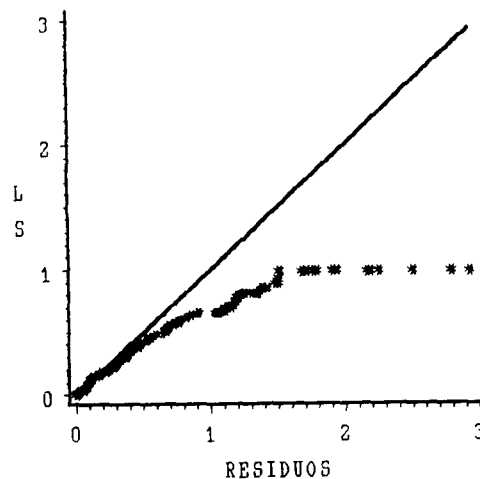
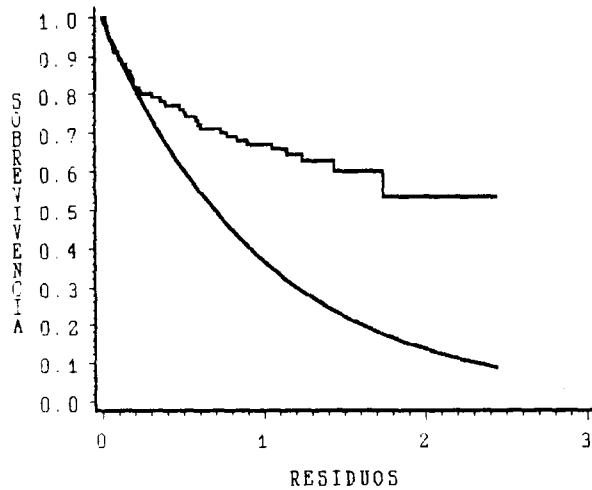
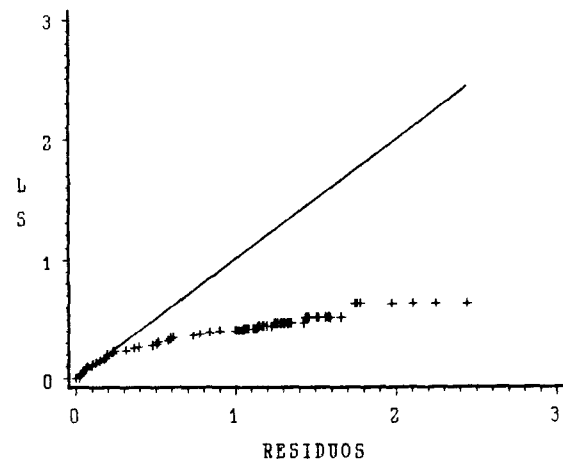


Figura 4.22 - Resíduos( $\hat{\epsilon}_i$ ) versus  $-\ln\hat{S}(\hat{\epsilon}_i)$  e reta com inclinação 1 para a causa 1.



**Figura 4.23 - Funções de sobrevivência estimada e suposta para os resíduos-causa2.**



**Figura 4.24 - Resíduos ( $\hat{\epsilon}_i$ ) versus  $-\ln(\hat{S}(\hat{\epsilon}_i))$  e reta com inclinação 1 para a causa 2.**

Pode-se portanto observar pelas figuras, que para a causa 1, o modelo possui ajuste razoável e para a causa 2 o mesmo apresenta um ajuste não satisfatório. Os programas PROG8.SAS e PROG11.SAS do apêndice A foram usados para a obtenção dos resíduos. Os valores residuais encontram-se na tabela B5 e B7 do apêndice B.

#### 4.3.5 - Funções de Sobrevivência e Risco estimadas, DVS e $t_{mR}$

Para cada uma das causas consideradas, tem-se para os níveis 0.3 - 0.7 - 0.8 e 0.9 da variável concomitante dose, as seguintes funções de sobrevivência e risco estimadas :

##### I - Modelo Log-linear e de Riscos Proporcionais Weibull

As curvas das sobrevivências e riscos estimadas por estes modelos são:

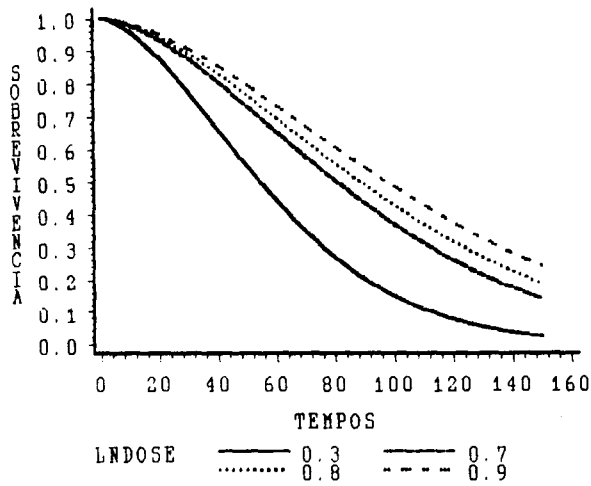


Figura 4.25 - Curvas das Sobrevivências estimadas pelos modelos log-linear e riscos proporcionais Weibull- causa1.

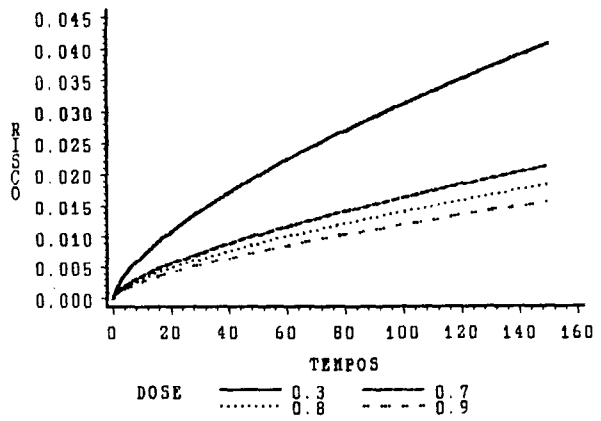


Figura 4.26 - Curvas dos riscos estimados pelos modelos log-linear e riscos proporcionais Weibull- causa1.

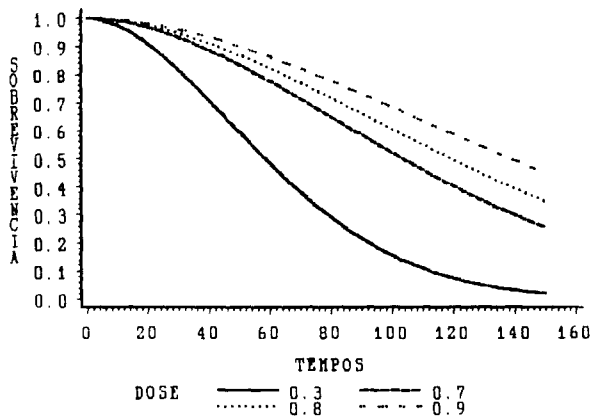


Figura 4.27- Curvas das sobrevivências estimadas pelos modelos log-linear e riscos proporcionais Weibull - causa2.

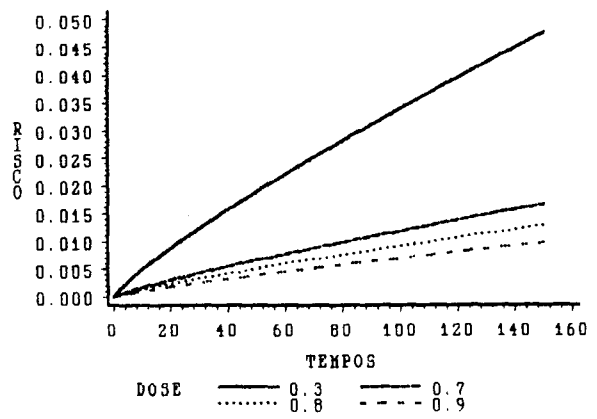


Figura 4.28 - Curvas dos riscos estimados pelos modelos log-linear e riscos proporcionais Weibull - causa 2.

## II - Modelo de riscos proporcionais de Cox

Para este modelo tem-se graficamente as seguintes curvas para os níveis 0.3 -0.7- 0.8 e 0.9 da variável dose:

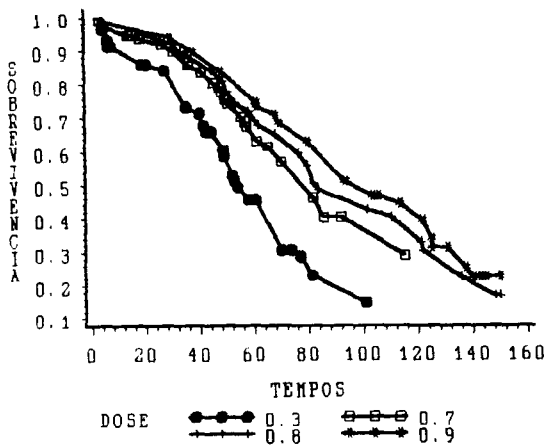


Figura 4.29 - Curvas das sobrevivências estimadas pelo modelo de Cox - causa 1.

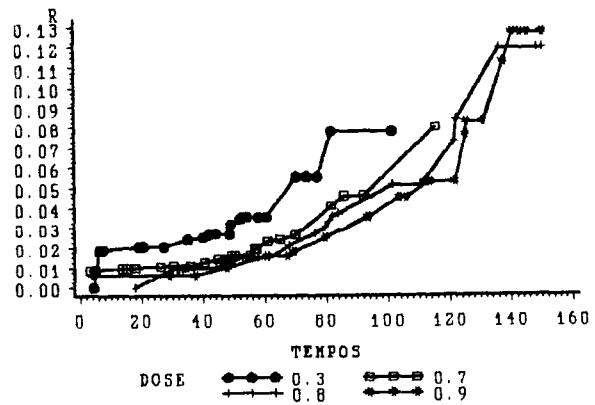


Figura 4.30 - Riscos estimados pelo modelo de Cox - causa 1.

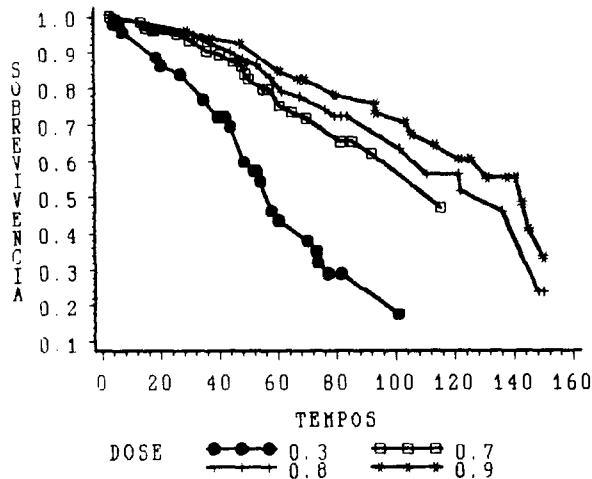


Figura 4.31- Curvas das Sobrevivências estimadas pelo modelo de Cox -causa 2.

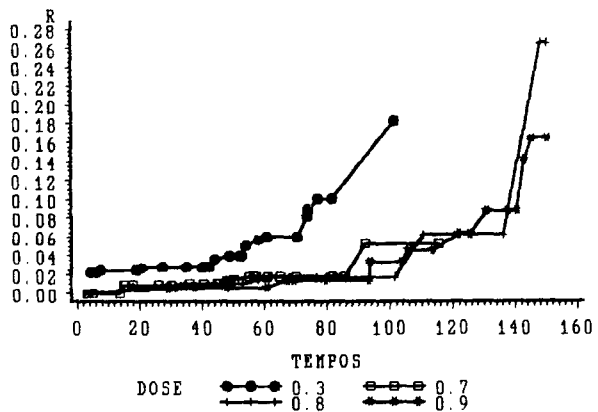


Figura 4.32 - Riscos estimados pelo modelo de Cox - causa 2.

Os valores de  $\hat{\lambda}(t_i | \mathbf{x}_i)$  e  $\hat{S}(t_i | \mathbf{x}_i)$  para as causas 1 e 2 encontram-se, respectivamente, nas tabela B6 e B8 do apêndice B.

Para cada causa, considerando  $t = 150$  e  $d = 0.8$ , obteve-se os seguintes tempos médios para a resposta :

	causal	causa2
<b>Modelo Log-linear e ou R. Proporcionalis</b>	89,69	108,59
<b>Modelo R. Prop. Cox</b>	64,63	74,51

Tabela 4.5 - tempo médio para causas 1 e 2.

Ainda, doses virtualmente seguras para um fixado tempo  $t$ , podem ser obtidas, para cada causa, utilizando-se as curvas dos riscos reduzidos  $\Pi(t | d)$  as quais são obtidas pelas expressões abaixo:

	Modelo log-linear e R.Proporcionais	Modelo R. Propr. Cox
CAUSA 1	$\exp\left\{-\left(\frac{t}{\exp\{3.93\}}\right)^{1.65}\right\} - \exp\left\{-\left(\frac{t}{\exp\{3.93 + 0.964 d\}}\right)^{1.65}\right\}$	$[\hat{S}_0(t)]^{\exp\{-1.7518 d\}} - [\hat{S}_0(t)]$
CAUSA 2	$\exp\left\{-\left(\frac{t}{\exp\{3.838\}}\right)^{1.83}\right\} - \exp\left\{-\left(\frac{t}{\exp\{3.838 + 1.434 d\}}\right)^{1.83}\right\}$	$[\hat{S}_0(t)]^{\exp\{-2.6713 d\}} - [\hat{S}_0(t)]$

Tabela 4.6 - expressões das curvas dos riscos reduzidos para os modelos tratados - causas 1 e 2.

#### 4.3.6 - Testes do efeito da dose no tipo e tempo de falha.

As hipóteses tratadas na seção 3.7 podem ser de interesse para verificar se a dose possui ou não influência no tipo e tempo de falha. As hipóteses que podem ser usadas para verificar tal influência são:

- (1)  $H_{01}: \beta_1^{(1)} = \beta_1^{(2)} = 0$   
 $H_{A1}$ : pelo menos um  $\beta_1^{(j)} \neq 0$ ,  $j = 1, 2$ , ou alternativamente, testa-se :
- (2)  $H_{02}: \beta_1^{(1)} = \beta_1^{(2)}$   
 $H_{A2}$ :  $\beta_1^{(1)} \neq \beta_1^{(2)}$ . Em havendo evidências favoráveis a  $H_{02}$ , pode-se então testar:
- (3)  $H_{03}: \beta_1^{(1)} = \beta_1^{(2)} = 0$  dado que  $\beta_1^{(1)} = \beta_1^{(2)}$   
 $H_{A3}$ : pelo menos um  $\beta_1^{(j)} \neq 0$ .

Os resultados obtidos para estas hipóteses utilizando-se a Normalidade assintótica de  $\hat{\beta}$  foram:

Modelo	Hipótese	$Q_i$	g.l.	p-valor
Log-linear e riscos	$H_{01}$	25.76	2	< 0.005
Proporc. Weibull	$H_{02}$	0.97	1	$0.75 < p < 0.90$
	$H_{03}$	24.79	1	0.0001
Riscos proporc. de	$H_{01}$	22.47	2	<0.005
Cox	$H_{02}$	0.96	1	$0.75 < p < 0.90$
	$H_{03}$	21.51	1	0.0001

Tabela 4.7 - Testes do efeito da dose utilizando a normalidade assintótica.

Pode-se portanto concluir, pelos resultados da tabela 4.7, que existem evidências estatísticas de que a dose tem influência no tempo de falha, mas não tem influência no tipo de falha.



## CONCLUSÕES

Nesta dissertação foi considerado a análise de experimentos de dose-resposta em que foram considerados os tempos de vida como medida associada a resposta dos indivíduos em estudo. Nestes experimentos é comum ter-se um período fixado para acompanhamento dos indivíduos e conseqüentemente é importante não só saber quantos indivíduos responderam ou não ao tratamento, mas também qual o período em que permaneceram livre da resposta. O quão longo foi o período em que permaneceram livre da resposta é sem dúvida mais informativo do que respostas do tipo 0 ou 1, indicando simplesmente se a resposta ocorreu ou não naquele período. Por este fato, trabalhou-se com os tempos até a ocorrência da resposta. Outra informação importante além do tempo livre da resposta é a causa pela qual a resposta ocorreu. É importante portanto, considerar também nos experimentos de dose-resposta as possíveis causas (riscos competitivos) atuando e que podem conduzir a resposta de interesse.

Uma variedade de modelos para experimentos do tipo dose-resposta podem ser usados. Foram abordados os modelos paramétricos log-linear e de riscos proporcionais e o modelo semi-paramétrico de Cox. Estes modelos foram escolhidos devido a sua flexibilidade em modelar dados de tempo para a resposta, bem como pela natureza empírica dos mesmos.

O propósito desta dissertação foi o de fazer uma revisão de alguns modelos de análise de sobrevivência em situações onde uma única causa de falha é considerada e também em situações envolvendo riscos competitivos. Como estes modelos consideram os tempos para a resposta, objetivou-se aplicá-los em experimentos de dose-resposta. Ainda, o comportamento das curvas de dose-resposta em regiões de doses baixas, objetivando a determinação de doses virtualmente seguras e o desenvolvimento de programas computacionais para essas análises, bem como a realização de ilustrações numéricas dos modelos tratados, foram propósitos desta dissertação.

Através das ilustrações numéricas realizadas pôde-se observar que os modelos apresentados adaptam-se bem a análise de dados de sobrevivência em que se verificam:

- a existência de relação linear entre o log dos tempos e a(s) variável(s) concomitante(s) e/ou
- a proporcionalidade dos riscos.

Em particular, os ajustes de tais modelos são satisfatórios quando há a ocorrência de uma quantidade razoável de censuras, ou seja, quando uma quantidade não muito grande de tempos censurados é observada.

Na aplicação 1 onde foi considerado apenas uma causa de falha e onde ocorreram 17 censuras dentre os 100 tempos gerados aleatoriamente, pôde-se observar que os modelos produziram ajustes consideravelmente satisfatórios. Já na aplicação 2, a qual foi considerado riscos competitivos, observou-se que a ocorrência de um grande número de censuras (acentuadamente para a causa 2, onde foi observado 62 censuras dentre os 100 tempos gerados aleatoriamente) influenciou consideravelmente nos ajustes dos modelos.

Nas situações em que riscos competitivos são tratados, muito provavelmente ocorrerão ajustes satisfatórios de determinados modelos para algumas causas e ajustes insatisfatórios (ruins) desses mesmos modelos para as outras causas consideradas. Isto é devido, essencialmente, a forma com que é tratado, nesta teoria, a classificação de um tempo como censurado em relação a uma causa em específico. Como foi visto, os tempos de falha por outras causas, que não a causa considerada, são classificados como tempos censurados para esta respectiva causa. Isto, obviamente, faz com que o número de censuras torne-se acentuadamente grande para algumas causas.

Deve-se ressaltar, no entanto, que em geral, o interesse concentra-se nas causas em que ocorrem um maior número de falhas. Nestes casos, os modelos produzem ajustes razoáveis e conseqüentemente produzem, para estas respectivas causas, boas estimativas das sobrevivências  $S_i(t | \underline{x})$  e dos riscos  $\lambda_i(t | \underline{x})$ .

É importante, portanto, reconhecer a validade de estudos e esforços na busca de modelos que se adaptem às situações de riscos competitivos, uma vez que estas situações são particularmente difíceis de serem tratadas. Finalmente, deve-se também reconhecer a validade dos modelos para experimentos dose-resposta não abordados aqui, ou seja, experimentos que não utilizam o tempo de sobrevivência (provavelmente pela impossibilidade de obtenção de uma medida quantitativa da resposta). Estes ocorrem com acentuada frequência e são igualmente de grande importância no contexto de modelagem dos referidos experimentos.

# APÊNDICE A

PROGRAMAS UTILIZADOS NAS APLICAÇÕES NUMÉRICAS

```

**PROG1.SAS**
ESTE PROGRAMA GERA UM ARQUIVO SAS CONTENDO 100 OBSERVAÇÕES COM AS
VARIÁVEIS TEMPO-DOSE-CENSURA. OS TEMPOS TÊM DISTRIBUIÇÃO WEIBULL E FOI
CONSIDERADO PARA A GERAÇÃO :
- níveis das doses : d1=0.3, d2=0.7, d3=0.8 e d4=0.9
- parâmetro de escala e coef. de regressão:  $\sigma=0.6$  ,  $\beta_0=3.9$  e  $\beta_1=1.1$ 
- semente utilizada : 15747
- tamanhos amostrais para cada nível da dose : 25

```

```

LIBNAME new 'a.';
DATA new.simvep;                               /* GERAÇÃO DE VALORES VALOR-EXTREMO PADRÃO */
n=100;
do j=1 to n ;
  u=ranuni(15747);
  w=log(-log(u));
  output;keep w;
end;

PROC IML;
use new.simvep;
read all var{w};
n1=25; n2=25; n3=25; n4=25;                    /* TAMANHOS AMOSTRAIS*/
n= n1+n2+n3+n4;
um=j(n,1);
dose=j(n1,1,.3)//j(n2,1,.7)//j(n3,1,.8)//j(n4,1,.9); /*MATRIZ DE COVARIÁVEIS*/
x=um||dose;
b={3.9, 1.1}; s=.6;                             /*COEF. REGRESSÃO E PARÂM. ESCALA*/
lntemp= (x*b)+(s*w);
t=145; lt=log(t);                               /*TEMPO FIXADO P/ TÉRMINO DO EXPERIMENTO*/
cens=j(n,1);
do i=1 to n;                                     /*CLASSIFICANDO TEMPOS COMO FALHA OU CENSURA*/

```

```
if lntemp[i,1] >log(t) then do;
    lntemp[i,1]=log(t);
    cens[i,1]=0;
end;
end;
temp= exp(lntemp);
arq=temp||dose||cens;
r={tempo dose censura};
create new.arq from arq[colname=r];
append from arq;
quit;

PROC SORT data=new.arq;
by tempo descending censura;
run;
PROC PRINT data=new.simvcp;
PROC PRINT data=new.arq;
run;
```

\*\*\*FIM DO PROG1.SAS\*\*\*

---

**\*\*PROG2.SAS\*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG1.SAS PARA :

- estimar os coeficientes de regressão do modelo log-linear em um contexto de uma única causa de falha utilizando a PROC LIFEREG do SAS.

```
PROC LIFEREG data=new.arq;
  model tempo*censura(0)=lndose / distribution = Weibull itprint;
run;
```

**\*\* FIM DO PROG2.SAS\*\***

---

**\*\*PROG3.SAS\*\***

ESTE PROGRAMA UTILIZA OS DADOS DO PROG1.SAS PARA:

- calcular os resíduos para os modelos log-linear e de riscos proporcionais Weibull em um contexto de uma única causa de falha.

```
PROC IML iml wrksize=50;
use new.arq;                                /* ARQUIVO GERADO NO PROG1.SAS */
read all into w;
  n=nrow(w);
  t=w[,1];
  x1=w[,2];
  one=j(n,1);
  x=one||x1;
cens= w[,3];
```

```
bch={3.777,1.2124};                               /* ESTIMATIVAS OBTIDAS NO PROG2.SAS */
sch={0.5979};
dch=1/sch;
  xbch=x*bch;
  expxb=exp(xbch);
  expxbd=expxb##-dch;
tdch=t##dch;
aux=tdch#expxbd;
expres=j(n,1,0); res=j(n,1,0);
  do i=1 to n;
    if cens[i,1]=1 then do;
      expres[i,1]=aux[i,1];
      res[i,1]=log(aux[i,1]);
    end;
    else do;
      expres[i,1]=aux[i,1]+1;
      res[i,1]=log(aux[i,1]+1);
    end;
  end;

print res expres;
quit;
```

**\*\*FIM DO PROG3.SAS\*\***

---



**\*\*PROG4.SAS\*\***

**ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG1.SAS PARA:**

**- estimar os coeficientes de regressão no modelo de Riscos Proporcionais de Cox em um contexto de um risco, utilizando a verossimilhança parcial e o método iterativo de Newton-Raphson.**

```
PROC PHREG data=new.arq;  
  MODEL tempo*censura(0)=lndose / covb itprint risklimits;  
  OUTPUT out=new.saidas num_left=numrisco survival=sobrev  
         logsurv=lnsobr loglogs=lnlnsobr / order=data;  
run;  
PROC PRINT data=new.saidas;  
run;
```

**\*\*FIM DO PROG4.SAS\*\***

---

**\*\*PROG5.SAS\*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG1.SAS E OS COEF.  
DE REGRESSÃO ESTIMADOS NO PROG4.SAS PARA :

- calcular os resíduos para o modelo de riscos proporcionais de Cox em um contexto  
de um único risco de falha.

```

PROC IML wrksize=40;
use new.saidas;
read all into w;
n=nrow(w);
t=w[,1];
x=w[,3];
cens=w[,2];
beta={-2.03445};          /* VALOR ESTIMADO PELA VEROS.PARCIAL- PROG4.SAS */
lnSch=w[,6];
  xb=x*beta;
  expxb=exp(xb);
res=j(n,1,0);
do i=1 to n;              /* CÁLCULO DOS RESÍDUOS */
  if cens[i,1]=1 then do;
    res[i,1]=-lnSch[i,1];
  end;
  else do;
    res[i,1]=(-lnSch[i,1]) +1;
  end;
end;
print , "RESIDUOS ", res;
quit;

```

**\*\* FIM DO PROG5.SAS \*\***

**\*\* PROG6.SAS \*\***

ESTE PROGRAMA GERA UM ARQUIVO SAS CONTENDO AS VARIÁVEIS TEMPO-DOSE-CAUSA-CENSC1-CENSC2 . PARA A GERAÇÃO DESTES ARQUIVO FOI UTILIZADO :

- níveis das doses :  $d_1=0.3$ ,  $d_2=0.7$ ,  $d_3=0.8$  e  $d_4=0.9$  .
- parâmetro de escala e regressores para a causa 1:  $\sigma_1=0.6$  ,  $\beta_0=4.0$  e  $\beta_1=0.7$  .
- parâmetro de escala e regressores para a causa 2:  $\sigma_2=0.5$  ,  $\beta_0=3.9$  e  $\beta_1=1.1$  .
- sementes utilizadas: 14767 e 19749.
- tamanhos amostrais para cada nível da dose: 25.

```
LIBNAME novo 'a:';
DATA novo.gera;                               /* GERAÇÃO DE VALORES VALOR-EXTREMO PADRÃO */
n=100;
do j=1 to n ;
  u1=ranuni(14767);
  u2=ranuni(19749);
  w1=log(-log(u1));
  w2=log(-log(u2));
output ;keep w1; keep w2;
end;

PROC IML;
use novo.gera;
read all var{w1 w2};
n1=25; n2=25; n3=25; n4=25;                   /* TAMANHOS AMOSTRAIS */
n= n1+n2+n3+n4;
um=j(n,1);
dose=j(n1,1,.3)//j(n2,1,.7)//j(n3,1,.8)//j(n4,1,.9); /*MATRIZ DAS COVARIÁVEIS*/
x=um||dose;
b1={ 4.0,0.7};
b2={ 3.9,1.1};
s1=0.6;
s2=0.5;
```

```

    lntemp1= (x*b1)+(s1*w1);
    lntemp2= (x*b2)+(s2*w2);
    temp1=exp(lntemp1);
    temp2=exp(lntemp2);
    t=150;
    cens=j(n,1);
    temp=temp1;
    causa=j(n,1);
    do i=1 to n;          /*CLASSIFICAÇÃO DOS TEMPOS POR CAUSA E CENSURA*/
        if temp1[i,1] >temp2[i,1] then do;
            temp[i,1]=temp2[i,1];
            causa[i,1]=2;
        end;
        if temp[i,1] > t then do;
            temp[i,1]=t;
            causa[i,1]=0;
        end;
    end;
    censC1=j(n,1,0);
    censC2=j(n,1,0);
    do i=1 to n;
        if causa[i,1]=1 then censC1[i,1]=1;
        if causa[i,1]=2 then censC2[i,1]=1;
    end;
    arq=temp||dose||causa||censC1||censC2;
    r={tempo dose causa cens1 cens2};
    create novo.arq from arq[colname=r];
    append from arq;

    arq1=temp||dose||censC1;
    r={tempo dose censC1};
    create novo.arq1 from arq1[colname=r];
    append from arq1;

```

```
arq2=temp||dose||censC2;  
r={tempo dose censC2};  
create novo.arq2 from arq2[colname=r];  
append from arq2;
```

```
PROC SORT data=novo.arq1;  
by tempo descending censC1;  
run;
```

```
PROC SORT data=novo.arq2;  
by tempo descending censC2;  
run;
```

```
PROC PRINT data=novo.gera;  
PROC PRINT data=novo.arq;  
PROC PRINT data=novo.arq1;  
PROC PRINT data=novo.arq2;  
quit;
```

**\*\* FIM DO PROG6.SAS \*\***

---

**\*\* PROG 7.SAS \*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG6.SAS PARA:

- estimar os coeficientes de regressão dos modelos log-linear e de riscos proporcionais Weibull em um contexto de riscos competitivos.

```
PROC LIFEREG data= novo.arq1;  
model tempo*censC1(0)=lndose / distribution=Weibull itprint;  
run;
```

```
PROC LIFEREG data=novo.arq2;  
model tempo*censC2(0)=lndose / distribution=Weibull itprint;  
run;
```

**\*\* FIM DO PROG7.SAS \*\***

---

**\*\* PROG8.SAS \*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG6.SAS PARA:

- calcular os resíduos dos modelos log-linear e de riscos proporcionais Weibull em um contexto de riscos competitivos ( CAUSA 1 ).

```
PROC IML wrksize=50;  
use novo.arq1;  
read all into w;  
n=nrow(w);  
t=w[ ,1];  
x1=w[ ,2];
```

```
one=j(n,1);
x=one||x1;
cens= w[ ,3];
bch={3.93,0.96}; /*ESTIMATIVAS OBTIDAS NO PROG7.SAS*/
sch=0.6;
dch=1/sch;
xb=x*bch;
expxb=exp(xb);
expxbd=expxb##-dch;
td=t##dch;
aux=td#expxbd; /* CÁLCULO DOS RESÍDUOS */
expres=j(n,1,0); res=j(n,1,0);
do i=1 to n;
  if cens[i,1]=1 then do;
    expres[i,1]=aux[i,1];
    res[i,1]=log(aux[i,1]);
  end;
  else do;
    expres[i,1]=aux[i,1]+1;
    res[i,1]=log(aux[i,1]+1);
  end;
end;
print res expres;
quit;
```

---

**\*\* FIM DO PROG8.SAS \*\***

**\*\* PROG9.SAS \*\***

**ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PRG6.SAS PARA:**

- calcular os resíduos dos modelos log-linear e de riscos proporcionais Weibull em um contexto de riscos competitivos (CAUSA 2).

```

PROC IML wrksize=50;
use novo.arq2;
read all into w;
n=nrow(w);
t=w[,1];
x1=w[,2];
one=j(n,1);
x=one||x1;
cens= w[,3];
bch={3.838 ,1.4356};          /*ESTIMATIVAS OBTIDAS NO PROG7.SAS*/
sch={0.5446};
dch=1/sch;
xb=x*bch;
expxb=exp(xb);
expxbd=expxb##-dch;
td=t##dch;
aux=td#expxbd;              /*CÁLCULO DOS RESÍDUOS*/
expres=j(n,1,0); res=j(n,1,0);
do i=1 to n;
  if cens[i,1]=1 then do;
    expres[i,1]=aux[i,1];
    res[i,1]=log(aux[i,1]);
  end;

```



```
        else do;
            expres[i,1]=aux[i,1]+1;
            res[i,1]=log(aux[i,1]+1);
        end;
    end;

print res expres;
quit;
```

**\*\* FIM DO PROG9.SAS \*\***

---

**\*\* PROG10.SAS \*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG6.SAS PARA :

- estimar os coeficientes de regressão no modelo de riscos proporcionais de Cox em um contexto de riscos competitivos, utilizando a verossimilhança parcial e o método iterativo de Newton-Raphson ( causas 1 e 2 ) .

```
PROC PHREG data=novo.arq1;
    MODEL tempo*censC1(0)=lndose / covb itprint risklimits;
    OUTPUT out=novo.saidas1 num_left=numrisco survival=sobrev
           logsurv=lnsobr loglogs=lnlnsobr / order=data;
run;

PROC PRINT data=novo.saidas1;
run;
```

```
PROC PHREG data=novo.arq2;
  MODEL tempo*censC2(0)=lndose / covb itprint risklimits;
  OUTPUT out=novo.saidas2 num_left=numrisco survival=sobrev
         logsurv=lnsobr loglogs=lnlnsobr / order=data;
```

```
run;
```

```
PROC PRINT data=novo.saidas2;
```

```
run;
```

**\*\* FIM DO PROG10.SAS \*\***

---

**\*\* PROG11.SAS \*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG6.SAS E O COEFICIENTE DE REGRESSÃO ESTIMADO NO PROG10.SAS PARA:

- calcular resíduos do modelo de riscos proporcionais de Cox em um contexto de riscos competitivos ( causa 1 ).

```
PROC IML wrksize=40;
```

```
use novo.saidas1;
```

```
read all into w;
```

```
n=nrow(w);
```

```
t=w[ ,1];
```

```
x=w[ ,3];
```

```
cens=w[ ,2];
```

```
beta={-1.751856};
```

```
/*VALOR ESTIMADO PELA VEROS.PARCIAL- PROG10.SAS */
```

```
lnSch=w[,6];
xb=x*beta;
expxb=exp(xb);
res=j(n,1,0);                               /*CÁLCULO DOS RESÍDUOS */
do i=1 to n;
  if cens[i,1]=1 then do;
    res[i,1]=-lnSch[i,1];
  end;
  else do;
    res[i,1]=(-lnSch[i,1]) +1;
  end;
end;
print , "RESIDUOS ", res;
quit;
```

**\*\* FIM DO PROG11.SAS \*\***

---

**\*\* PROG12.SAS\*\***

ESTE PROGRAMA UTILIZA OS DADOS GERADOS NO PROG6.SAS E O COEFICIENTE DE REGRESSÃO ESTIMADO NO PROG10.SAS PARA :

- calcular os resíduos do modelo de riscos proporcionais de Cox em um contexto de riscos competitivos ( causa 2).

```
PROC IML wrksize=40;
use novo.saidas2;
read all into w;
```

```
n=nrow(w);
t=w[ ,1];
x=w[ ,3];
cens=w[ ,2];
beta={-2.671393};          /*VALOR ESTIMADO PELA VEROS.PARCIAL- PROG10.SAS */
lnSch=w[ ,6];
xb=x*beta;
expxb=exp(xb);
res=j(n,1,0);
do i=1 to n;                /* CALCULO DOS RESÍDUOS */
  if cens[i,1]=1 then do;
    res[i,1]=-lnSch[i,1];
  end;
  else do;
    res[i,1]=(-lnSch[i,1]) +1;
  end;
end;
print , "RESIDUOS ", res;
quit;
```

**\*\* FIM DO PR0G12.SAS \*\***

---

**APÊNDICE B**

RESULTADOS OBTIDOS NAS APLICAÇÕES NUMÉRICAS

OBS	TEMPO	DOSE	CENSURA	OBS	TEMPO	DOSE	CENSURA
1	5.0936	0.3	1	26	40.8373	0.8	1
2	10.7598	0.3	1	27	42.8191	0.7	1
3	12.7143	0.3	1	28	44.3659	0.9	1
4	13.5813	0.8	1	29	44.4704	0.9	1
5	16.2718	0.3	1	30	45.4211	0.3	1
6	17.2154	0.8	1	31	50.3195	0.3	1
7	17.9588	0.9	1	32	50.5085	0.3	1
8	18.4521	0.8	1	33	51.1984	0.7	1
9	19.5502	0.8	1	34	54.3364	0.7	1
10	23.8063	0.9	1	35	55.5195	0.7	1
11	25.6406	0.7	1	36	56.4847	0.8	1
12	26.2479	0.3	1	37	57.4882	0.3	1
13	26.6251	0.3	1	38	60.6429	0.8	1
14	27.0399	0.3	1	39	61.3056	0.7	1
15	28.6315	0.3	1	40	61.6591	0.9	1
16	29.8417	0.7	1	41	65.183	0.3	1
17	32.1409	0.9	1	42	65.6322	0.9	1
18	32.4973	0.3	1	43	66.3782	0.3	1
19	33.1863	0.9	1	44	66.6845	0.7	1
20	34.2259	0.8	1	45	67.3237	0.8	1
21	35.3322	0.7	1	46	71.7188	0.8	1
22	35.6655	0.7	1	47	72.6588	0.8	1
23	39.0054	0.7	1	48	74.0838	0.3	1
24	40.3647	0.7	1	49	77.7083	0.3	1
25	40.4862	0.3	1	50	82.9211	0.3	1

TABELA B1 - Dados gerados aleatoriamente para realização da aplicação 1 .

cont.

OBS	TEMPO	DOSE	CENSURA	OBS	TEMPO	DOSE	CENSURA
51	82.9924	0.3	1	76	128.254	0.9	1
52	84.2343	0.7	1	77	128.436	0.8	1
53	86.8314	0.7	1	78	128.966	0.9	1
54	87.5874	0.3	1	79	136.16	0.9	1
55	87.756	0.7	1	80	140.141	0.8	1
56	89.683	0.3	1	81	141.185	0.9	1
57	89.77	0.7	1	82	143.166	0.7	1
58	94.737	0.9	1	83	143.222	0.3	1
59	96.68	0.9	1	84	145	0.7	0
60	99.55	0.9	1	85	145	0.7	0
61	101.943	0.9	1	86	145	0.7	0
62	102.228	0.8	1	87	145	0.8	0
63	104.38	0.7	1	88	145	0.8	0
64	106.649	0.7	1	89	145	0.8	0
65	106.821	0.8	1	90	145	0.8	0
66	108.826	0.3	1	91	145	0.8	0
67	109.287	0.8	1	92	145	0.8	0
68	111.26	0.7	1	93	145	0.8	0
69	111.464	0.3	1	94	145	0.9	0
70	116.221	0.8	1	95	145	0.9	0
71	119.774	0.9	1	96	145	0.9	0
72	119.995	0.8	1	97	145	0.9	0
73	121.546	0.7	1	98	145	0.9	0
74	121.976	0.7	1	99	145	0.9	0
75	125.818	0.9	1	100	145	0.9	0

TABELA B1- Dados gerados aleatoriamente para realização da aplicação 1.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
1	0.1857919	1.2041716	0.0192331	26	-0.755558	0.4697483	0.2162286
2	0.573715	1.7748485	0.0388434	27	-3.36713	0.0344885	0.2772345
3	0.832499	2.2990568	0.058846	28	0.2665391	1.3054387	0.192779
4	0.6269205	1.8718374	0.028611	29	-0.969428	0.3793001	0.2010663
5	0.0279522	1.0283465	0.0996868	30	-1.289126	0.2755116	0.7101175
6	0.0114459	1.0115116	0.0435871	31	0.6099513	1.8403417	0.739572
7	1.1680796	3.2158111	0.0417584	32	-0.295061	0.7444861	0.7699204
8	0.4392224	1.5515003	0.0588252	33	0.0409097	1.041758	0.3549631
9	-0.759199	0.4680413	0.0665296	34	-0.374997	0.6872914	0.3689049
10	-2.009981	0.1339912	0.0606128	35	0.3052565	1.356973	0.3830438
11	0.8785508	2.4074085	0.1006338	36	-2.902972	0.0548599	0.3242173
12	0.6618648	1.9384037	0.2490437	37	-0.865121	0.4210004	0.9296738
13	-3.83418	0.0216191	0.2715087	38	-2.277463	0.102544	0.3483912
14	0.1425411	1.1532005	0.29449	39	-0.634049	0.5304395	0.4421831
15	-3.873334	0.0207889	0.3180119	40	-0.985217	0.3733582	0.3046098
16	-0.980828	0.3750006	0.1515404	41	0.2035166	1.2257055	1.0679745
17	0.8362	2.3075815	0.1080052	42	-5.082592	0.0062038	0.3258099
18	-0.293343	0.745766	0.3906038	43	-1.547353	0.2128105	1.1415434
19	-0.321065	0.725376	0.1225927	44	-1.370379	0.2540106	0.5228596
20	-0.420978	0.6564043	0.1593369	45	-2.031766	0.1311037	0.4406556
21	0.501681	1.651495	0.2065347	46	-1.223698	0.2941404	0.4549022
22	0.220342	1.246503	0.2179112	47	0.1615066	1.1752803	0.4693547
23	0.4356737	1.5460042	0.2294187	48	0.039523	1.0403144	1.339097
24	-0.696319	0.4984167	0.2410601	49	0.2493362	1.2831734	1.3819499
25	-0.663016	0.5152949	0.57071	50	-5.867867	0.0028289	1.4267216

TABELA B2 - Resíduos obtidos para os 3 modelos ajustados na aplicação 1.



cont.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
51	-0.573092	0.5637795	1.4735922	76	-0.823186	0.4390307	0.973339
52	-0.212821	0.8083009	0.674562	77	0.0777433	1.0808452	1.2340537
53	0.9253943	2.5228627	0.6965303	78	-0.065891	0.9362329	1.0417262
54	-0.534083	0.5862068	1.6230834	79	-0.229132	0.7952235	1.0778273
55	0.1830897	1.2009221	0.7429784	80	0.8193748	2.2690808	1.3671054
56	-2.257322	0.1046303	1.7320124	81	0.0933701	1.097868	1.1546975
57	0.0852458	1.0889847	0.7932289	82	0.8193748	2.2690808	1.7965405
58	-2.035646	0.1305961	0.5455047	83	-1.2782	0.2785383	4.2096868
59	-0.620262	0.5378037	0.5632525	84	0.1283461	1.1369465	2.8656474
60	-2.175399	0.1135629	0.5813211	85	-1.276581	0.2789895	2.8656474
61	0.1173078	1.1244655	0.5997221	86	0.4547054	1.5757092	2.8656474
62	0.0960506	1.1008148	0.7580546	87	0.7749965	2.1705846	2.5222098
63	0.21054	1.2343444	0.9580471	88	0.8193748	2.2690808	2.5222098
64	0.2784149	1.3210342	0.9878729	89	-0.556594	0.5731581	2.5222098
65	-0.687841	0.5026599	0.8310325	90	-5.972202	0.0025486	2.5222098
66	0.2998595	1.3496691	2.3708462	91	0.1170849	1.1242148	2.5222098
67	-1.395893	0.2476119	0.8849036	92	0.6395715	1.8956685	2.5222098
68	0.043362	1.0443158	1.1194706	93	0.2870888	1.3325425	2.5222098
69	0.2742641	1.3155622	2.6095739	94	-0.382699	0.6820182	2.2419939
70	-1.300947	0.2722739	0.9756411	95	0.5343226	1.706292	2.2419939
71	0.6583724	1.9316458	0.8229537	96	-1.052247	0.3491525	2.2419939
72	0.1273662	1.1358329	1.0426318	97	-0.062518	0.9393962	2.2419939
73	0.937805	2.5543683	1.3211868	98	0.7889429	2.2010685	2.2419939
74	-0.887178	0.4118163	1.3664675	99	0.6980182	2.0097658	2.2419939
75	-2.143882	0.117199	0.9410036	100	-1.679361	0.1864932	2.2419939

TABELA B2 - Resíduos obtidos para os 3 modelos ajustados na aplicação 1.

OBS	DOSE	TEMPO	SOBREV.	RISCO	OBS	DOSE	TEMPO	SOBREV.	RISCO
1	0.3	5.0936	0.9809	0.01923	26	0.8	40.8373	0.80555	0.00986
2	0.3	10.7598	0.9619	0.01961	27	0.7	42.8191	0.75788	0.01222
3	0.3	12.7143	0.9428	0.02000	28	0.9	44.3659	0.82466	0.00821
4	0.8	13.5813	0.9717	0.00733	29	0.9	44.4704	0.81786	0.00828
5	0.3	16.2718	0.9051	0.02056	30	0.3	45.4211	0.49159	0.02861
6	0.8	17.2154	0.9573	0.00754	31	0.3	50.3195	0.47732	0.02945
7	0.9	17.9588	0.9591	0.00619	32	0.3	50.5085	0.46305	0.03034
8	0.8	18.4521	0.9428	0.00764	33	0.7	51.1984	0.7012	0.01375
9	0.8	19.5502	0.9356	0.00770	34	0.7	54.3364	0.69149	0.01394
10	0.9	23.8063	0.9411	0.00633	35	0.7	55.5195	0.68178	0.01413
11	0.7	25.6406	0.9042	0.00958	36	0.8	56.4847	0.72309	0.01168
12	0.3	26.2479	0.7795	0.02197	37	0.3	57.4882	0.39468	0.03304
13	0.3	26.6251	0.7622	0.02246	38	0.8	60.6429	0.70582	0.01222
14	0.3	27.0399	0.7449	0.02298	39	0.7	61.3056	0.64263	0.01518
15	0.3	28.6315	0.7275	0.02352	40	0.9	61.6591	0.73741	0.01024
16	0.7	29.8417	0.8593	0.01060	41	0.3	65.1830	0.3437	0.03551
17	0.9	32.1409	0.8976	0.00712	42	0.9	65.6322	0.72194	0.01072
18	0.3	32.4973	0.6766	0.02452	43	0.3	66.3782	0.31933	0.03722
19	0.9	33.1863	0.8846	0.00735	44	0.7	66.6845	0.59282	0.01695
20	0.8	34.2259	0.8527	0.00908	45	0.8	67.3237	0.64361	0.01404
21	0.7	35.3322	0.8134	0.01124	46	0.8	71.7188	0.63451	0.01424
22	0.7	35.6655	0.8042	0.01137	47	0.8	72.6588	0.62541	0.01445
23	0.7	39.0054	0.7950	0.01150	48	0.3	74.0838	0.26208	0.04109
24	0.7	40.3647	0.7857	0.01164	49	0.3	77.7083	0.25109	0.04285
25	0.3	40.4862	0.5651	0.02677	50	0.3	82.9211	0.24009	0.04477

TABELA B3 - Sobrevivências e riscos estimados pelo modelo de Cox - Aplicação 1.

cont.

OBS	DOSE	TEMPO	SOBREV.	RISCO	OBS	DOSE	TEMPO	SOBREV.	RISCO
51	0.3	82.9924	0.2291	0.04687	76	0.9	128.254	0.37782	0.03234
52	0.7	84.2343	0.5093	0.02149	77	0.8	128.436	0.29111	0.04111
53	0.7	86.8314	0.4983	0.02196	78	0.9	128.966	0.35285	0.03484
54	0.3	87.5874	0.1972	0.05141	79	0.9	136.16	0.34033	0.03610
55	0.7	87.7564	0.4756	0.02366	80	0.8	140.141	0.25484	0.04610
56	0.3	89.6830	0.1769	0.05554	81	0.9	141.185	0.31515	0.03926
57	0.7	89.7700	0.4523	0.02564	82	0.7	143.166	0.16587	0.06202
58	0.9	94.7370	0.5795	0.01744	83	0.3	143.222	0.01485	0.15593
59	0.9	96.6800	0.5693	0.01775	84	0.7	145	0.15480	0.06202
60	0.9	99.5500	0.5591	0.01807	85	0.7	145	0.15480	0.06202
61	0.9	101.943	0.5489	0.01840	86	0.7	145	0.15480	0.06202
62	0.8	102.228	0.4685	0.02302	87	0.8	145	0.21823	0.04610
63	0.7	104.380	0.3836	0.02896	88	0.8	145	0.21823	0.04610
64	0.7	106.649	0.3723	0.02983	89	0.8	145	0.21823	0.04610
65	0.8	106.821	0.4356	0.02501	90	0.8	145	0.21823	0.04610
66	0.3	108.826	0.0934	0.07262	91	0.8	145	0.21823	0.04610
67	0.8	109.287	0.4127	0.02761	92	0.8	145	0.21823	0.04610
68	0.7	111.260	0.3264	0.03492	93	0.8	145	0.21823	0.04610
69	0.3	111.464	0.0735	0.08358	94	0.9	145	0.28881	0.03926
70	0.8	116.221	0.3769	0.03203	95	0.9	145	0.28881	0.03926
71	0.9	119.774	0.4391	0.02691	96	0.9	145	0.28881	0.03926
72	0.8	119.995	0.3525	0.03401	97	0.9	145	0.28881	0.03926
73	0.7	121.546	0.2668	0.04332	98	0.9	145	0.28881	0.03926
74	0.7	121.976	0.2550	0.04528	99	0.9	145	0.28881	0.03926
75	0.9	125.818	0.3902	0.03132	100	0.9	145	0.28881	0.03926

TABELA B3 - Sobrevivências e riscos estimados pelo modelo de Cox - Aplicação 1.

OBS TEMPO DOSE CAUSACENSC1 CENSC2						OBS TEMPO DOSE CAUSACENSC1 CENSC2					
1	3.056	0.7	1	1	0	26	36.035	0.7	2	0	1
2	3.659	0.9	1	1	0	27	37.378	0.9	2	0	1
3	4.355	0.3	2	0	1	28	39.916	0.3	1	1	0
4	4.888	0.7	1	1	0	29	40.468	0.7	1	1	0
5	6.201	0.3	1	1	0	30	41.446	0.3	1	1	0
6	6.348	0.3	1	1	0	31	42.279	0.3	1	1	0
7	7.465	0.3	2	0	1	32	44.031	0.3	2	0	1
8	13.63	0.7	1	1	0	33	44.140	0.8	2	0	1
9	14.96	0.7	2	0	1	34	44.985	0.7	1	1	0
10	15.28	0.7	2	0	1	35	47.224	0.7	1	1	0
11	17.87	0.8	2	0	1	36	47.658	0.9	1	1	0
12	17.99	0.7	1	1	0	37	48.005	0.7	2	0	1
13	18.89	0.3	1	1	0	38	48.394	0.8	2	0	1
14	20.47	0.3	2	0	1	39	48.644	0.3	2	0	1
15	26.15	0.7	1	1	0	40	48.862	0.3	1	1	0
16	27.20	0.3	2	0	1	41	49.108	0.7	1	1	0
17	28.91	0.9	2	0	1	42	50.014	0.8	1	1	0
18	29.88	0.8	1	1	0	43	50.227	0.7	2	0	1
19	30.29	0.7	1	1	0	44	51.870	0.3	1	1	0
20	31.38	0.8	1	1	0	45	52.737	0.3	1	1	0
21	31.97	0.8	1	1	0	46	52.936	0.8	1	1	0
22	33.15	0.8	2	0	1	47	53.957	0.3	2	0	1
23	34.11	0.8	1	1	0	48	55.135	0.7	2	0	1
24	34.92	0.8	2	0	1	49	56.634	0.7	1	1	0
25	34.96	0.3	1	1	0	50	57.171	0.7	1	1	0

TABELA B4 - Dados gerados aleatoriamente para a realização da aplicação 2.

cont.

OBS TEMPO DOSE CAUSACENSC1 CENSC2						OBS TEMPO DOSE CAUSACENSC1 CENSC2					
51	57.818	0.8	2	0	1	76	93.258	0.9	1	1	0
52	57.900	0.3	2	0	1	77	93.506	0.9	2	0	1
53	60.503	0.3	2	0	1	78	101.13	0.3	2	0	1
54	60.577	0.9	1	1	0	79	101.46	0.8	1	1	0
55	60.636	0.9	1	1	0	80	103.47	0.9	1	1	0
56	60.850	0.7	1	1	0	81	105.66	0.9	2	0	1
57	61.477	0.8	2	0	1	82	110.45	0.8	2	0	1
58	64.970	0.7	1	1	0	83	113.42	0.9	1	1	0
59	67.348	0.9	2	0	1	84	115.28	0.7	1	1	0
60	67.794	0.8	1	1	0	85	121.25	0.8	1	1	0
61	69.340	0.9	1	1	0	86	121.78	0.9	2	0	1
62	69.866	0.7	1	1	0	87	122.04	0.8	1	1	0
63	70.232	0.3	1	1	0	88	125.05	0.9	1	1	0
64	73.490	0.3	2	0	1	89	125.30	0.9	1	1	0
65	73.778	0.3	2	0	1	90	130.83	0.9	2	0	1
66	76.375	0.8	1	1	0	91	136.09	0.8	1	1	0
67	77.096	0.3	2	0	1	92	137.55	0.9	1	1	0
68	79.517	0.9	1	1	0	93	140.45	0.9	1	1	0
69	79.524	0.8	1	1	0	94	142.92	0.9	2	0	1
70	81.549	0.3	1	1	0	95	145.14	0.9	2	0	1
71	81.570	0.7	1	1	0	96	148.17	0.8	2	0	1
72	81.799	0.8	1	1	0	97	150	0.8	0	0	0
73	83.730	0.8	1	1	0	98	150	0.9	0	0	0
74	85.564	0.7	1	1	0	99	150	0.9	0	0	0
75	91.990	0.7	2	0	1	100	150	0.9	0	0	0

TABELA B4 - Dados gerados aleatoriamente para realização da aplicação 2.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
1	-5.807863	0.0030038	0.0088126	26	0.1684347	1.1834510	1.1548795
2	-5.827700	0.0029448	0.0124626	27	0.1324192	1.1415868	1.1091013
3	0.0102273	1.0102798	1.0356536	28	-0.885342	0.4125732	0.3373580
4	-5.025336	0.0065694	0.0268032	29	-1.502471	0.2225795	0.1801663
5	-3.988577	0.0185260	0.0726327	30	-0.822682	0.4392522	0.3893075
6	-3.949561	0.0192632	0.0916036	31	-0.789495	0.4540738	0.4162417
7	0.0249205	1.0252336	1.0916036	32	0.3959953	1.4858623	1.4162417
8	-3.315879	0.0363021	0.0551915	33	0.1982098	1.2192182	1.1733544
9	0.0415365	1.0424112	1.0551915	34	-1.326116	0.2655064	0.2207380
10	0.0429872	1.0439246	1.0551915	35	-1.245147	0.2878986	0.2351343
11	0.0474633	1.0486077	1.0463223	36	-1.549883	0.2122729	0.1759018
12	-2.853345	0.0576511	0.0653060	37	0.2591873	1.2958765	1.2497092
13	-2.131746	0.1186300	0.1523070	38	0.2275708	1.2555463	1.2095812
14	0.1271377	1.1355734	1.1523070	39	0.4533795	1.5736213	1.5032261
15	-2.230025	0.1075257	0.0862339	40	-0.548305	0.5779286	0.5350164
16	0.1970277	1.2177778	1.1737827	41	-1.179948	0.3072946	0.2816423
17	0.0882767	1.0922903	1.0607454	42	-1.309474	0.2699621	0.2501488
18	-2.167888	0.1144190	0.0816812	43	0.2769164	1.3190560	1.2980441
19	-1.985189	0.1373547	0.1085213	44	-0.448741	0.6384312	0.6350599
20	-2.086251	0.1241517	0.1005807	45	-0.421117	0.6563135	0.6707146
21	-2.054845	0.1281127	0.1101705	46	-1.214860	0.2967516	0.2945687
22	0.1275372	1.1360272	1.1101705	47	0.5198743	1.6818163	1.7072896
23	-1.946988	0.1427033	0.1199477	48	0.3167683	1.3726845	1.3509689
24	0.1383741	1.1484051	1.1199477	49	-0.942302	0.3897296	0.3705141
25	-1.106195	0.3308154	0.3121208	50	-0.926568	0.3959102	0.3904489

TABELA B5 - Resíduos obtidos para os 3 modelos ajustados - causa 1 da aplicação 2.

cont.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
51	0.2954671	1.3437539	1.3277043	76	-0.431046	0.6498292	0.6741121
52	0.5691947	1.7668437	1.7868516	77	0.5024172	1.6527114	1.6741121
53	0.6016792	1.8251810	1.7868516	78	1.0792979	2.9426128	2.9285238
54	-1.150126	0.3165969	0.2908930	79	-0.130543	0.8776190	0.8546297
55	-1.148513	0.3171080	0.3069986	80	-0.257760	0.7727811	0.7626132
56	-0.822627	0.4392763	0.4591320	81	0.5878907	1.8001873	1.7626132
57	0.3226421	1.3807711	1.3853500	82	0.6986515	2.0110390	1.9086287
58	-0.713447	0.4899523	0.4834954	83	-0.104730	0.9005676	0.8156003
59	0.3204486	1.3777456	1.3405870	84	0.2422710	1.2741394	1.2382172
60	-0.802539	0.4481894	0.4270886	85	0.1664807	1.1811408	1.1120904
61	-0.924964	0.3965458	0.3766817	86	0.7000713	2.0138964	1.9333788
62	-0.592364	0.5530187	0.5611925	87	0.1772478	1.1939269	1.1962229
63	0.0563416	1.0579590	1.1864862	88	0.0579396	1.0596510	1.0805140
64	0.7612824	2.1410202	2.1864862	89	0.0613008	1.0632187	1.1633812
65	0.7647613	2.1484815	2.1864862	90	0.7620083	2.1425749	2.1633812
66	-0.603901	0.5466751	0.5213642	91	0.3589738	1.4318593	1.5056744
67	0.8046187	2.2358437	2.2518489	92	0.2166433	1.2419011	1.3764089
68	-0.696705	0.4982244	0.4626992	93	0.2514969	1.2859489	1.5034357
69	-0.536557	0.5847579	0.5820656	94	0.8432298	2.3238604	2.5034357
70	0.3053488	1.3570983	1.4756061	95	0.8579044	2.3582136	2.5034357
71	-0.334218	0.7158978	0.7733112	96	0.9744836	2.6497986	2.7912946
72	-0.489553	0.6129004	0.6848816	97	0.9872373	2.6838096	2.7912946
73	-0.450680	0.6371947	0.7220546	98	0.8898843	2.4348479	2.5034357
74	-0.254562	0.7752562	0.9064783	99	0.8898843	2.4348479	2.5034357
75	0.6284501	1.8747027	1.9064783	100	0.8898843	2.4348479	2.5034357

TABELA B5 - Resíduos obtidos para os 3 modelos ajustados - causa 1 da aplicação 2.

OBS	DOSE	TEMPO	SOBREV. RISCO		OBS	DOSE	TEMPO	SOBREV. RISCO	
1	0.7	3.0565	0.99123	0.008813	26	0.7	36.0359	0.85652	0.011201
2	0.9	3.6597	0.98761	0.006255	27	0.9	37.3785	0.89664	0.006255
3	0.3	4.3556	0.96497	0.000000	28	0.3	39.9167	0.71365	0.025237
4	0.7	4.8881	0.97355	0.009111	29	0.7	40.4682	0.83513	0.012764
5	0.3	6.2019	0.92994	0.018618	30	0.3	41.4460	0.67753	0.026228
6	0.3	6.3488	0.91247	0.018971	31	0.3	42.2795	0.65952	0.026934
7	0.3	7.4652	0.91247	0.018971	32	0.3	44.0314	0.65952	0.026934
8	0.7	13.632	0.94630	0.009736	33	0.8	44.1409	0.84084	0.009777
9	0.7	14.966	0.94630	0.009736	34	0.7	44.9850	0.80193	0.014192
10	0.7	15.284	0.94630	0.009736	35	0.7	47.2245	0.79046	0.014396
11	0.8	17.878	0.95473	0.000000	36	0.9	47.6589	0.83870	0.010267
12	0.7	17.993	0.93678	0.010114	37	0.7	48.0053	0.77903	0.014396
13	0.3	18.896	0.85872	0.020699	38	0.8	48.3946	0.81092	0.009777
14	0.3	20.471	0.85872	0.020699	39	0.3	48.6440	0.60458	0.026934
15	0.7	26.153	0.91738	0.010657	40	0.3	48.8628	0.58566	0.03179
16	0.3	27.205	0.84048	0.020699	41	0.7	49.1084	0.75454	0.016158
17	0.9	28.913	0.94106	0.006255	42	0.8	50.0146	0.77868	0.013766
18	0.8	29.882	0.92157	0.009305	43	0.7	50.2277	0.74227	0.016158
19	0.7	30.292	0.89716	0.011201	44	0.3	51.8708	0.52990	0.034427
20	0.8	31.382	0.90431	0.009499	45	0.3	52.7377	0.51134	0.035655
21	0.8	31.979	0.89568	0.009590	46	0.8	52.9360	0.74485	0.015233
22	0.8	33.150	0.89568	0.009590	47	0.3	53.9578	0.49298	0.035655
23	0.8	34.117	0.88697	0.009777	48	0.7	55.1351	0.70401	0.016158
24	0.8	34.928	0.88697	0.009777	49	0.7	56.6345	0.69038	0.019545
25	0.3	34.962	0.73189	0.024114	50	0.7	57.1717	0.67675	0.019935

TABELA B6 - Sobrevivências e riscos estimados pelo modelo de Cox - Causa 1 para a aplicação 2.



cont.

OBS	DOSE	TEMPO	SOBREV.	RISCO	OBS	DOSE	TEMPO	SOBREV.	RISCO
51	0.8	57.818	0.72058	0.015233	76	0.9	93.258	0.50961	0.035565
52	0.3	57.900	0.45528	0.035655	77	0.9	93.506	0.50961	0.035365
53	0.3	60.503	0.45528	0.035655	78	0.3	101.13	0.14536	0.078007
54	0.9	60.577	0.74760	0.015850	79	0.8	101.46	0.42544	0.051447
55	0.9	60.636	0.73565	0.016106	80	0.9	103.47	0.46645	0.045321
56	0.7	60.850	0.63183	0.023318	81	0.9	105.66	0.46645	0.045321
57	0.8	61.477	0.68021	0.015233	82	0.8	110.45	0.40308	0.051447
58	0.7	64.970	0.61662	0.024363	83	0.9	113.42	0.44237	0.052987
59	0.9	67.348	0.71135	0.016106	84	0.7	115.28	0.28990	0.080396
60	0.8	67.794	0.65241	0.021290	85	0.8	121.25	0.32887	0.072853
61	0.9	69.340	0.68613	0.018226	86	0.9	121.78	0.39322	0.052987
62	0.7	69.866	0.57053	0.026457	87	0.8	122.04	0.30233	0.084133
63	0.3	70.232	0.30529	0.055544	88	0.9	125.05	0.33942	0.076523
64	0.3	73.490	0.30529	0.055544	89	0.9	125.30	0.31243	0.082867
65	0.3	73.778	0.30529	0.055544	90	0.9	130.83	0.31243	0.082867
66	0.8	76.375	0.59371	0.027222	91	0.8	136.09	0.22187	0.119540
67	0.3	77.096	0.28598	0.055544	92	0.9	137.55	0.25248	0.112690
68	0.9	79.517	0.62958	0.025118	93	0.9	140.45	0.22236	0.127030
69	0.8	79.524	0.55874	0.030775	94	0.9	142.92	0.22236	0.127030
70	0.3	81.549	0.22864	0.078007	95	0.9	145.14	0.22236	0.127030
71	0.7	81.570	0.46148	0.041091	96	0.8	148.17	0.16674	0.119540
72	0.8	81.799	0.50415	0.035841	97	0.8	150	0.16674	0.119540
73	0.8	83.730	0.48575	0.037173	98	0.9	150	0.22236	0.127030
74	0.7	85.564	0.40394	0.046174	99	0.9	150	0.22236	0.127030
75	0.7	91.990	0.40394	0.046174	100	0.9	150	0.22236	0.127030

TABELA B6 - Sobrevivências e riscos estimados pelo modelo de Cox - Causa 1 - aplicação 2.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
1	0.0010684	1.0010690	1.0000000	26	-2.310695	0.0991923	0.10207
2	0.0008779	1.0008783	1.0000000	27	-2.770741	0.0626156	0.066161
3	-5.136292	0.0058795	0.0226509	28	0.2953050	1.3435361	1.328630
4	0.0025284	1.0025315	1.0077807	29	0.1157717	1.1227395	1.112885
5	0.0111873	1.0112501	1.0226509	30	0.3134153	1.3680896	1.328630
6	0.0116758	1.0117442	1.0226509	31	0.3233847	1.3817968	1.328630
7	-4.146949	0.0158126	0.0471579	32	-0.888318	0.4113470	0.364586
8	0.0165092	1.0166463	1.0161989	33	-2.201786	0.1106054	0.105551
9	-3.924211	0.0197577	0.0248315	34	0.1389455	1.1490614	1.137873
10	-3.885584	0.0205358	0.0335394	35	0.1509769	1.1629698	1.137873
11	-3.861299	0.0210406	0.0323947	36	0.0933300	1.0978239	1.080806
12	0.0273324	1.0277094	1.0423146	37	-1.784076	0.1679521	0.151089
13	0.0834367	1.0870164	1.1231853	38	-2.032852	0.1309615	0.125905
14	-2.294574	0.1008043	0.1500905	39	-0.705383	0.4939195	0.518687
15	0.0536016	1.0550641	1.0515567	40	0.4041356	1.4980071	1.518687
16	-1.772398	0.1699250	0.1780048	41	0.1613590	1.1751067	1.178171
17	-3.242251	0.0390759	0.0415521	42	0.1302594	1.1391238	1.136401
18	0.0526254	1.0540347	1.0542763	43	-1.700979	0.1825048	0.193251
19	0.0696320	1.0721135	1.0708969	44	0.4419533	1.5557431	1.562588
20	0.0574383	1.0591199	1.0542763	45	0.4529319	1.5729171	1.562588
21	0.0594016	1.0612013	1.0542763	46	0.1435882	1.1544086	1.147946
22	-2.727544	0.0653797	0.0620426	47	-0.515015	0.5974918	0.612990
23	0.0666521	1.0689236	1.0620426	48	-1.529809	0.2165771	0.228481
24	-2.631593	0.0719637	0.0699315	49	0.2049919	1.2275151	1.228481
25	0.2384967	1.2693395	1.2659246	50	0.2082276	1.2314934	1.228481

TABELA B7 - Resíduos obtidos para os 3 modelos ajustados - Causa2 da aplicação 2.

cont.

RESÍDUOS							
Obs	Log-linear	R.P.Weibull	Cox	Obs	Log-linear	R.P.Weibull	Cox
51	-1.706165	0.1815607	0.1893810	76	0.2893558	1.3355668	1.2811314
52	-0.385537	0.6800854	0.7771404	77	-1.087059	0.3372069	0.3141193
53	-0.304759	0.7373014	0.8375780	78	0.6385243	1.8936843	1.7424870
54	0.1414614	1.1519561	1.1686252	79	0.4120463	1.5099043	1.4582305
55	0.1416960	1.1522264	1.1686252	80	0.3408590	1.4061550	1.3508059
56	0.2307786	1.2595803	1.2877112	81	-0.862625	0.4220527	0.3961784
57	-1.593489	0.2032153	0.2375938	82	-0.517615	0.5959404	0.5800447
58	0.2567793	1.2927597	1.3103504	83	0.3925435	1.4807423	1.4440628
59	-1.689608	0.1845919	0.1956856	84	0.6092589	1.8390680	1.7576674
60	0.2176853	1.2431958	1.2556089	85	0.5349159	1.7073047	1.5800447
61	0.1779275	1.1947387	1.1956856	86	-0.601835	0.5478053	0.5076777
62	0.2885859	1.3345389	1.3338821	87	0.5398473	1.7157449	1.6631399
63	0.6777741	1.9694890	1.9719895	88	0.4543206	1.5751028	1.5076777
64	0.0522836	1.0536745	1.0540750	89	0.4556742	1.5772364	1.5076777
65	0.0594645	1.0612681	1.1435059	90	-0.470197	0.6248791	0.5953096
66	0.2644321	1.3026910	1.3007134	91	0.6282878	1.8743985	1.7776067
67	0.1402357	1.1505449	1.2443339	92	0.5217566	1.6849849	1.5953096
68	0.2234791	1.2504195	1.2505153	93	0.5375450	1.7117993	1.5953096
69	0.2821691	1.3260029	1.3272287	94	-0.307949	0.7349531	0.7361405
70	0.8222073	2.2755170	2.2443339	95	-0.279724	0.7559922	0.9001076
71	0.3678276	1.4445929	1.4274334	96	0.0218728	1.0221138	1.4406660
72	0.2951516	1.3433301	1.3272287	97	0.7155702	2.0453525	2.4406660
73	0.3062732	1.3583533	1.3272287	98	0.5895181	1.8031192	2.1029255
74	0.3956661	1.4853733	1.4274334	99	0.5895181	1.8031192	2.1029255
75	-0.589869	0.5543998	0.4796711	100	0.5895181	1.8031192	2.1029255

TABELA B7 - Resíduos obtidos para os 3 modelos ajustados - Causa 2 da aplicação 2.

OBS	DOSE	TEMPO	SOBREV.	RISCO	OBS	DOSE	TEMPO	SOBREV.	RISCO
1	0.7	3.0565	1.00000	0.000000	26	0.7	36.035	0.90297	0.010724
2	0.9	3.6597	1.00000	0.000000	27	0.9	37.378	0.93598	0.006339
3	0.3	4.3556	0.97760	0.022651	28	0.3	39.916	0.71991	0.027914
4	0.7	4.8881	0.99225	0.000000	29	0.7	40.468	0.89325	0.010724
5	0.3	6.2019	0.97760	0.022651	30	0.3	41.446	0.71991	0.027914
6	0.3	6.3488	0.97760	0.022651	31	0.3	42.279	0.71991	0.027914
7	0.3	7.4652	0.95394	0.024507	32	0.3	44.031	0.69448	0.035955
8	0.7	13.632	0.98393	0.000000	33	0.8	44.140	0.89983	0.009674
9	0.7	14.966	0.97547	0.008633	34	0.7	44.985	0.87121	0.010724
10	0.7	15.284	0.96702	0.008708	35	0.7	47.224	0.87121	0.010724
11	0.8	17.878	0.96812	0.006718	36	0.9	47.658	0.92237	0.006339
12	0.7	17.993	0.95857	0.008708	37	0.7	48.005	0.85977	0.013216
13	0.3	18.896	0.88410	0.024507	38	0.8	48.394	0.88170	0.010237
14	0.3	20.471	0.86063	0.026905	39	0.3	48.644	0.59530	0.039914
15	0.7	26.153	0.94975	0.008708	40	0.3	48.862	0.59530	0.039914
16	0.3	27.205	0.83694	0.027914	41	0.7	49.108	0.83680	0.013216
17	0.9	28.913	0.95930	0.005715	42	0.8	50.014	0.87249	0.010237
18	0.8	29.882	0.94717	0.006718	43	0.7	50.227	0.82427	0.01508
19	0.7	30.292	0.93156	0.008708	44	0.3	51.870	0.56973	0.039911
20	0.8	31.382	0.94717	0.006718	45	0.3	52.737	0.56973	0.039911
21	0.8	31.979	0.94717	0.006718	46	0.8	52.936	0.86248	0.010237
22	0.8	33.150	0.93984	0.007766	47	0.3	53.957	0.54173	0.050402
23	0.8	34.117	0.93984	0.007766	48	0.7	55.135	0.79574	0.017917
24	0.8	34.928	0.93246	0.007889	49	0.7	56.634	0.79574	0.017917
25	0.3	34.962	0.76650	0.027914	50	0.7	57.171	0.79574	0.017917

TABELA B8 - Sobrevivências e riscos estimados pelo modelo de Cox - Causa 2 - Aplicação 2.

cont.

OBS	DOSE	TEMPO	SOBREV. RISCO		OBS	DOSE	TEMPO	SOBREV. RISCO	
51	0.8	57.818	0.82747	0.01446	76	0.9	93.258	0.75493	0.01379
52	0.3	57.900	0.45972	0.05699	77	0.9	93.506	0.73043	0.03299
53	0.3	60.503	0.43276	0.06043	78	0.3	101.13	0.17508	0.18223
54	0.9	60.577	0.84483	0.00633	79	0.8	101.46	0.63240	0.01733
55	0.9	60.636	0.84483	0.00634	80	0.9	103.47	0.70412	0.03299
56	0.7	60.850	0.74998	0.01792	81	0.9	105.66	0.67289	0.04537
57	0.8	61.477	0.78852	0.01733	82	0.8	110.45	0.55987	0.06255
58	0.7	64.970	0.73319	0.01792	83	0.9	113.42	0.64143	0.04537
59	0.9	67.348	0.82227	0.01379	84	0.7	115.28	0.46876	0.05224
60	0.8	67.794	0.77444	0.01733	85	0.8	121.25	0.55987	0.06255
61	0.9	69.340	0.82227	0.01379	86	0.9	121.78	0.60189	0.06361
62	0.7	69.866	0.71614	0.01792	87	0.8	122.04	0.51523	0.06255
63	0.3	70.232	0.37833	0.06044	88	0.9	125.05	0.60189	0.06361
64	0.3	73.490	0.34851	0.08209	89	0.9	125.30	0.60189	0.06361
65	0.3	73.778	0.31870	0.08943	90	0.9	130.83	0.55139	0.08763
66	0.8	76.375	0.74029	0.01733	91	0.8	136.09	0.45950	0.06255
67	0.3	77.096	0.28813	0.10083	92	0.9	137.55	0.55139	0.08763
68	0.9	79.517	0.77840	0.01379	93	0.9	140.45	0.55139	0.08763
69	0.8	79.524	0.72092	0.01733	94	0.9	142.92	0.47896	0.14083
70	0.3	81.549	0.28813	0.10083	95	0.9	145.14	0.40653	0.16397
71	0.7	81.570	0.65218	0.01792	96	0.8	148.17	0.23677	0.26493
72	0.8	81.799	0.72092	0.01733	97	0.8	150	0.23677	0.26493
73	0.8	83.730	0.72092	0.01733	98	0.9	150	0.33190	0.16397
74	0.7	85.564	0.65218	0.01792	99	0.9	150	0.33190	0.16397
75	0.7	91.990	0.61899	0.05224	100	0.9	150	0.33190	0.16397

TABELA B8 - Sobrevivências e riscos estimados pelo modelo de Cox - causa 2 - Aplicação 2.

Abcissas ( $\pm x_i$ ) and weight factors ( $w_i$ ) for Gaussian Integration

$$\int_A^B f(y) dy \simeq \frac{(B+A)}{2} \sum_{i=1}^n w_i f\left(\frac{A+B+x_i(B-A)}{2}\right).$$

$\pm x_i$	$w_i$	$\pm x_i$	$w_i$
<b>n = 2</b>			
0.57735 02691 89626	1.00000 00000 00000	0.18343 46424 95650	0.36268 37833 78362
<b>n = 8</b>			
0.00000 00000 00000	0.88888 88888 88889	0.52553 24099 16329	0.31370 66458 77887
0.77459 66692 41483	0.55555 55555 55556	0.79666 64774 13627	0.22238 10344 53374
<b>n = 4</b>			
0.33998 10435 84856	0.65214 51548 62546	0.96028 98564 97536	0.10122 85362 90376
0.86113 63115 94053	0.34785 48451 37454	<b>n = 9</b>	
<b>n = 5</b>			
0.00000 00000 00000	0.56888 88888 88889	0.00000 00000 00000	0.33023 93550 01260
0.53846 93101 05683	0.47862 86704 99366	0.32425 34234 03809	0.31234 70770 40003
0.90617 98459 38664	0.23692 68850 56189	0.61337 14327 00590	0.26061 06964 02935
<b>n = 6</b>			
0.23861 91860 83197	0.46791 39345 72691	0.83603 11073 26636	0.18064 81606 94857
0.66120 93864 66265	0.36076 15730 48139	0.96816 02395 07626	0.08127 43883 61574
0.93246 95142 03152	0.17132 44923 79170	<b>n = 10</b>	
<b>n = 7</b>			
0.00000 00000 00000	0.41795 91836 73469	0.14887 43389 81631	0.29552 42247 14753
0.40584 51513 77397	0.38183 00505 05119	0.43339 53941 29247	0.26926 67193 09996
0.74153 11855 99394	0.27970 53914 89277	0.67940 95682 99024	0.21908 63625 15982
0.94910 79123 42759	0.12948 49661 68870	0.86506 33666 88985	0.14945 13491 50581
<b>n = 12</b>			
<b>n = 16</b>			
0.09501 25098 37637 440185	0.18945 06104 55068 496285	0.97390 65285 17172	0.24914 70458 13403
0.28160 35507 79258 913230	0.18260 34150 44923 588867	0.12523 34085 11469	0.23349 25365 38355
0.45801 67776 57227 386342	0.16915 65193 95002 538189	0.36783 14989 98180	0.20316 74267 23066
0.61787 62444 02643 748447	0.14959 59888 16576 732081	0.58731 79542 86617	0.16007 83285 43346
0.75540 44083 55003 033895	0.12462 89712 55533 872052	0.76990 26741 94305	0.10693 93259 95318
0.86563 12023 87831 743880	0.09515 85116 82492 784810	0.90411 72563 70475	0.04717 53363 86512
0.94457 50230 73232 576078	0.06225 35239 38647 892863	0.98156 06342 46719	
0.98940 09349 91649 932596	0.02715 24594 11754 094852		
<b>n = 20</b>			
0.07652 65211 33497 333755	0.15275 33871 30725 850698		
0.22778 58511 41645 078080	0.14917 29864 72603 746788		
0.37370 60887 15419 560673	0.14209 61093 18382 051329		
0.51086 70019 50827 098004	0.13168 86384 49176 626898		
0.63605 36807 26515 025453	0.11819 45319 61518 417312		
0.74633 19064 60150 792614	0.10193 01198 17240 435037		
0.83911 69718 22218 823395	0.08327 67415 76704 748725		
0.91223 44282 51325 905868	0.06267 20483 34109 063570		
0.96397 19272 77913 791268	0.04060 14298 00386 941331		
0.99312 85991 85094 924786	0.01761 40071 39152 118312		
<b>n = 24</b>			
0.06405 68928 62605 626085	0.12793 81953 46752 156974		
0.19111 88674 73616 309159	0.12583 74563 46828 296121		
0.31504 26796 96163 374387	0.12167 04729 27803 391204		
0.43379 35076 26045 138487	0.11550 56680 53725 601353		
0.54542 14713 88839 535658	0.10744 42701 15965 634783		
0.64809 36519 36975 569252	0.09761 86521 04113 888270		
0.74012 41915 78554 364244	0.08619 01615 31953 275917		
0.82000 19859 73902 921954	0.07334 64814 11080 305734		
0.88641 55270 04401 034213	0.05929 85849 15436 780746		
0.93827 45520 02732 758524	0.04427 74388 17419 806169		
0.97472 85559 71309 498198	0.02853 13886 28933 663181		
0.99518 72199 97021 360180	0.01234 12297 99987 199547		

Compiled from P. Davis and P. Rabinowitz, Abcissas and weights for Gaussian quadratures of high order, J. Research NBS 56, 35-37, 1956, RP2645; P. Davis and P. Rabinowitz, Additional abcissas and weights for Gaussian quadratures of high order. Values for n=64, 80, and 96, J. Research NBS 60, 613-614, 1958, RP2875; and A. N. Lowan, N. Davids, and A. Levenson, Table of the zeros of the Legendre polynomials of order 1-16 and the weight coefficients for Gauss' mechanical quadrature formula, Bull. Amer. Math. Soc. 48, 739-743, 1942 (with permission).

Tabela B9 - Abcissas e fatores de ponderação para integração Gaussiana.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BERMAN, S. M. (1963). *Note on extreme values, competing risks and semi-Markov processes*, Ann. Math. Statist., **34**, 1104-1106.
- BIRNBAUM, Z. William (1980). *On the mathematics of competing risks*, DheW Publications, US department of Health, Education and Welfare, 1-58.
- BRESLOW, N. (1971). *Covariance analysis of censored survival data*, Biometrics, **30**, 89-99.
- COX, D. R. and SNELL, E. J. (1968). *A general definition of residuals*, Journal of the Royal Statistical Society, B, **30**, 2, 248-265.
- COX, D. R. (1972). *Regression models and life tables (with discussion)*, Journal of the Royal Statistical Society, B, **34**, 187-220.
- COX, D. R. (1975). *Partial likelihood*, Biometrika, **62**, 269-276.
- ELANDT-JOHNSON, R. C. and JOHNSON, N. L. (1980). *Survival Models and Data Analysis*, John Wiley & Sons, New York.
- FAREWELL, V. T. and PRENTICE, R. L. (1977). *A study of distributional shape in life testing*, Technometrics, **18**, 69-76.
- FINNEY, D. J. (1978). *Statistical Method in Biological Assay*, Charles Griffin & Company, London and High Wycombe.
- GAIL, M (1975). *A review and critique of some models used in competing risk analysis*, Biometrics, **31**, 209-222.
- GART, J. J. et al (1986). *The Design and Analysis of Long-term Animal Experiments*, **3**, IARC Scientific Publications.
- GLASSER, M. (1967). *Exponential survival with covariance*, Journal American Statistical Association, **62**, 561-568.

- GREENWOOD, M. (1926). *The natural duration of cancer*, Reports on Public Health and Medical Subjects, 33, 1-26.
- HOLT, J. D. (1978). *Competing risk analyses with special reference to matched pair experiments*, Biometrika, 65, 159-165.
- JONHSON, N.L. and KOTZ, S. . *Competing Risks*, Encyclopedia of Statistical Sciences, 2, 75-81.
- KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.
- KALBFLEISCH, J.D. ; KREWSKI, D.R. and VAN RYZIN, J.(1983). *Dose-response models for time-to-response toxicity data (with discussion)*, Canadian Journal Statistics, 11, 1, 25-49.
- KAPLAN, E. L. and MEIER, P. (1958). *Nonparemetric estimation from incomplete observations*, Journal of the American Statistical Association, 53, 457-481.
- KAY, R. (1977). *Proportional hazard regression models and the analysis of censored survival data*, Applied Statistics, 26, 3, 227-237.
- KENNEDY, Williams J. and GENTLE, James E. (1980). *Statistical Computing: Statistical, Textbooks and Monographs*, 33, Marcel Dekker, Inc., New York.
- LAGAKOS, W. (1978). *A covariate model for partially censored data subject to competing causes of failure*, Applied Statistics, 27, 3, 235-241.
- LAWLESS, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.
- LEE, Elisa T. (1980). *Statistical Methods for Survival Data Analysis*, Lifetime Bearing Publications, Belmont, California.
- MOESHBERGER, M.L. and DAVID, H. A. (1971). *Life tests under competing causes of failure and the theory of competing risks*, Biometrics, 27, 909-933.
- MYERS, M. H. ; HANKEL, B. J. and MANTEL, N. ( 1973 ). *A logistic exponential model for use with response-time data involving regressor variables*, Biometrics, 29, 257-269.
- NELSON, Wayne (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, John Wiley & Sons, New York.



- NELSON, Wayne (1982). *Applied Life Data Analysis*, John Wiley & Sons.
- PRENTICE, R.L. (1973). *Exponential survivals with censoring and explanatory variables*, *Biometrika*, 60, 279-284.
- PRENTICE, R.L. et al (1978). *The analysis of failure times in the Presence of competing risks*, *Biometrics*, 34, 541-554.
- PRENTICE, R.L. and KALBFLEISCH, J.D. (1979). *Hazard rate models with covariates*, *Biometrics*, 35, 25-39.
- SALDIVA, Carmem D. (1980). *Ensaio biológicos com resposta quantal*, Tese de mestrado - USP.
- SAS Institute Inc. SAS User's Guide : *Statistics, Version 5 edition*. Cary, N.C.: SAS Institute Inc, 1985, 956 pp.
- ZIPPIN, C. and ARMITAGE, P. (1966). *Use of concomitant variables and incomplete survival parameter*, *Biometrics*, 22, 665-672.
- WADA, Cicilia Y. ; SEN, P. K. and TARUMOTO, M. H. (1993). *On the restricted alternatives tests in a proportional hazard model with competing risk data*, em impressão - *Revista Brasileira de Probabilidade e Estatística*.
-