Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Análise de influência local nos modelos de riscos múltiplos

Juliana Betini Fachini

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração: Estatística
e Experimentação Agronômica

Piracicaba
2006

Juliana Betini Fachini
Bacharel em Estatística

Análise de influência local nos modelos de riscos múltiplos

Orientador:
Prof. Dr. Edwin Moises Marcos Ortega

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração: Estatística
e Experimentação Agronômica

Piracicaba
2006

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Fachini, Juliana Betini

Análise de influência local nos modelos de riscos múltiplos / Juliana Betini Fachini. - -
Piracicaba, 2006.
77 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2007.
Bibliografia.

1. Dados censurados 2. Logística (estatística) 3. Modelos estatísticos I. Título

CDD 519.5

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

*A Deus,
pois sem Ele nada teria acontecido.*

*Aos meus pais,
Edson Fachini e Maria de Fátima Betini Fachini, que
sempre lutaram para dar uma boa educação e formação aos
seus filhos.*

*Ao meu irmão,
Ricardo Betini Fachini e noiva **Erica Gonçalves**, pela
amizade e grandes incentivos.*

*Aos meus avós,
Ezio Fachini, Aparecida Barozi Fachini (in memo-
riam), **Serafim Betini e Aparecida Meroti Betini** (in
memoriam), pelo incentivo e grande amor.*

AGRADECIMENTOS

Ao professor Dr. **Edwin Moises Marcos Ortega**, pela compreensão, e principalmente pela orientação à elaboração deste trabalho.

À Capes pela concessão de bolsa de estudo.

Ao conselho do programa de Pós-graduação em Estatística e Experimentação Agronômica, os professores Dra. **Clarice Garcia Borges Demétrio**, Dr. **Décio Barbin** e Dra. **Roseli Aparecida Leandro**, pelas valiosas sugestões e confiança.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, Dr. **Carlos Tadeu dos Santos Dias**, Dr. **César Gonçalves de Lima**, Dr. **Gerson Barretos Mourão**, Dr. **Silvio Sandoval Zocchi** e Dra. **Sônia Maria Stefano Piedade**, pela amizade e formação.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, as secretárias **Solange de Assis Paes Sabadin** e **Luciane Brajão** e aos técnicos em informática **Jorge Alexandre Wiendl** e **Eduardo Bonilha**, pelos auxílios permanentes.

Aos colegas de estudo do mestrado e do doutorado, pela amizade, companheirismo e paciência.

Um agradecimento especial aos colegas, **Alexandre**, **Ana Paula**, **Cristiane**, **Fernanda**, **Joseane** e **Pâmela**, pelo incentivo, motivação, companheirismo, paciência e principalmente a grande amizade dedicada.

Ao namorado **Júnior**, pelo amor e paciência.

À amiga **Elizabeth**, pela amizade incondicional e formação.

Aos primos **Rogério** e **Andréia**, pelo carinho e cuidados; e aos primos **Bruno** e **Bianca** pelas alegrias.

À todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

SUMÁRIO

	Página
RESUMO	8
ABSTRACT	9
1 INTRODUÇÃO	10
2 REVISÃO BIBLIOGRÁFICA	13
2.1 Notação e conceitos básicos	13
2.1.1 Tempo de falha	13
2.1.2 Censura	13
2.1.3 Variáveis explicativas	15
2.1.4 Representação dos dados de sobrevivência	15
2.1.5 Especificando o tempo de sobrevivência	16
2.1.5.1 Função densidade de probabilidade	16
2.1.5.2 Função de sobrevivência	16
2.1.5.3 Função de risco	16
2.1.6 Relações entre as funções	17
2.1.7 Distribuições importantes em análise de sobrevivência	18
2.1.7.1 Distribuição Weibull	18
2.1.7.2 Distribuição log-logística	20
2.2 Modelos de riscos múltiplos	21
2.2.1 Modelo Weibull múltiplo	25
2.2.2 Modelo log-logístico múltiplo	26
2.2.3 Inferência	27
2.3 Influência local	29
2.3.1 Método de influência local	30

	6
2.4	Análise de resíduos 33
2.4.1	Resíduo martingal 34
2.4.2	Resíduo deviance 35
2.5	Impacto das observações influentes detectadas 35
3	MATERIAL E MÉTODOS 37
3.1	Material 37
3.2	Métodos 37
3.2.1	Modelos de riscos múltiplos 37
3.2.2	Influência local 39
3.2.2.1	Perturbação de casos 39
3.2.2.2	Perturbação de uma covariável 41
3.2.2.3	Perturbação da variável resposta 44
3.2.3	Análise de resíduos 46
3.2.3.1	Resíduo martingal 46
3.2.3.2	Resíduo deviance 47
4	RESULTADOS E DISCUSSÃO 49
4.1	Modelo de riscos múltiplos 49
4.2	Influencia local 52
4.2.1	Perturbação de casos 52
4.2.2	Perturbação de uma covariável 54
4.2.3	Perturbação da variável resposta 58
4.3	Análise de resíduo 59
4.4	Impacto das observações influentes detectadas 60
4.5	Reanálise dos dados 61
4.6	Qualidade de ajuste 61

		7
5	CONSIDERAÇÕES FINAIS	64
5.1	PERSPECTIVAS PARA TRABALHOS FUTUROS	64
	REFERÊNCIAS	66
	ANEXOS	69

RESUMO

Análise de influência local nos modelos de riscos múltiplos

Neste trabalho, é apresentado vários métodos de diagnóstico para modelos de riscos múltiplos. A vantagem desse modelo é sua flexibilidade em relação aos modelos de risco simples, como, os modelos Weibull e log-logístico, pois acomoda uma grande classe de funções de risco, função de risco não-monótona, por exemplo, forma de “banheira” e curvas multimodal. Alguns métodos de influência, assim como, a influência local, influência local total de um indivíduo são calculadas, analisadas e discutidas. Uma discussão computacional do método do afastamento da verossimilhança, bem como da curvatura normal em influência local são apresentados. Finalmente, um conjunto de dados reais é usado para ilustrar a teoria estudada. Uma análise de resíduo é aplicada para a seleção do modelo apropriado.

Palavras-chaves: Modelo de risco múltiplo; distribuição Weibull múltipla; distribuição log-logística múltipla; análise de influência local; dados censurados; análise de resíduo

ABSTRACT

Influence diagnostics for polyhazard models in the presence of covariates

In this paperwork is present various diagnostic methods for polyhazard models. Polyhazard models are a flexible family for fitting lifetime data. Their main advantage over the single hazard models, such as the Weibull and the log-logistic models, is to include a large amount of nonmonotone hazard shapes, as bathtub and multimodal curves. Some influence methods, such as the local influence, total local influence of an individual are derived, analyzed and discussed. A discussion of the computation of the likelihood displacement as well as the normal curvature in the local influence method are presented. Finally, an example with real data is given for illustration. A residual analysis is performed in order to select an appropriate model.

Keywords: Polyzard model; Poly-Weibull distribution; Poly-log-logistic distribution; influence diagnostics; censored data; residual analysis

1 INTRODUÇÃO

Para situações de estudo em que o tempo de ocorrência de um evento é a variável resposta (ou variável de observação), a análise de sobrevivência é uma técnica apropriada.

Este tempo pode ser definido como a resposta a um determinado tratamento, a recidiva (recorrência) da doença ou da causa em estudo, ou a morte do paciente, sendo em análise de sobrevivência, denominado tempo de falha. Geralmente, é considerado o tempo do diagnóstico da doença até a morte do paciente, animal ou objeto em estudo, ou da remissão (após o tratamento, o paciente não apresenta mais os sintomas) até a recidiva da doença.

Em dados de sobrevivência, existe a presença de outro tempo considerado como a principal característica dessa análise, que é a observação censurada, representada por δ_i , podendo haver também covariáveis ou variáveis explicativas, representadas por um vetor $\mathbf{x}_i$.

À modelagem de riscos múltiplos, supõe que um indivíduo ou sistema de indivíduos está sujeito a $k \geq 2$ motivos diferentes e independentes que levarão à falha do indivíduo, sendo que os motivos responsáveis pela ocorrência do evento de interesse ou falha são completamente ou parcialmente desconhecidos. O evento de interesse será descrito pelo primeiro motivo que causar a falha do indivíduo ou do objeto observado.

A vantagem do modelo de riscos múltiplos é sua flexibilidade em relação ao modelo de risco usual, pois, acomoda não apenas função risco crescente, decrescente ou constante, como também a função de risco não-monótona, como por exemplo, forma de “banheira” e curvas multimodais.

No presente trabalho é discutido a modelagem de riscos múltiplos proposta por Berger e Sun (1993) e Louzada-Neto (1999), sendo que os parâmetros do modelo são estimados através do método de máxima verossimilhança para o qual foram utilizados os procedimentos NLMIXED do SAS, assim como, o MAXBFGS do OX.

Quando se está ajustando um modelo a um conjunto de dados é imprescindível que as estimativas obtidas a partir do modelo proposto sejam resistentes a pequenas pertur-

bações nas observações ou no modelo. Se o modelo ajustado não apresentar uma boa descrição dos dados que foram observados, o mesmo pode conduzir a inferências errôneas.

Assim, é importante que se faça um estudo sobre a robustez dos resultados obtidos em termos dos vários aspectos que envolvem a formulação do modelo, e as estimativas dos seus parâmetros, ou seja, fazer uma análise de diagnóstico, que consiste de métodos para avaliar o grau de sensibilidade das inferências a pequenas perturbações nos dados, ou mesmo, no modelo proposto.

Dentre as medidas geralmente utilizadas para detectar influência de observações sobre o modelo ajustado, duas se destacam. Uma delas é a distância D_i , sugerida por Cook (1977), e a outra medida de influência é conhecida por $DFFITS_i$, que foi proposta por Belsley; Kuh e Welsch (1980), ambas medidas utiliza a estratégia da exclusão de casos.

Uma abordagem mais moderna é o estudo de influência local introduzido por Cook (1986), que se baseia na análise das curvaturas das seções normais de uma determinada superfície, denominada gráfico de influência, numa vizinhança de um ponto especial.

A teoria de influência local a qual tem como objetivo verificar, através de alguma medida de influência, o quanto as estimativas obtidas a partir do modelo proposto são resistentes a pequenas perturbações nas observações. Se essas perturbações causarem efeitos desproporcionais em determinados resultados, significa que há indícios de que o modelo está mal ajustado. A identificação das observações responsáveis por essa discrepância pode ajudar na escolha de um modelo mais adequado.

Por outro lado, quando se procura ajustar um modelo a um conjunto de dados, a validação desse ajuste passa pela análise de uma estatística especial, os resíduos, que tem a finalidade de medir a qualidade do modelo ajustado. Uma vez que os resíduos são usados para identificar discrepâncias entre um modelo ajustado e o conjunto de dados, é conveniente buscar uma definição para resíduo que leve em consideração a construção de cada observação sobre essa medida de qualidade de ajuste.

Em análise de sobrevivência os resíduos martingal e deviance introduzidos por

Therneau et al. (1990) são muito utilizados. Portanto, aos modelos de riscos múltiplos serão propostos dois resíduos utilizando o trabalho de Therneau et al. (1990)

Na revisão de literatura são apresentados alguns conceitos básicos relacionados a análise de sobrevivência, a descrição da teoria de riscos múltiplos, influência local e resíduos.

Em materiais e métodos é descrito o conjunto de dados a ser usado, o procedimento de estimação dos parâmetros de riscos múltiplos, é desenvolvida a teoria de influência local considerando os três esquemas de perturbação para os modelos Weibull múltiplo e log-logístico múltiplo e desenvolve a teoria de resíduos proposta por Therneau et al. (1990) aos modelos Weibull múltiplo e log-logístico múltiplo.

No contexto de resultados e discussão, é ajustado o modelo bi-log-logístico aos dados de peixes da espécie “Notropis Dourado, crysoleucas de Notemigonus”, é aplicado influência local e análise de resíduos ao modelo ajustado, é analisado o impacto dos possíveis pontos influentes, é feito a reanálise dos dados e verificado a qualidade de ajuste através do método gráfico.

2 REVISÃO BIBLIOGRÁFICA

2.1 Notação e conceitos básicos

2.1.1 Tempo de falha

O tempo de falha é uma característica muito importante em dados de sobrevivência, e por isso, tem que ser bem definido para evitar problemas como, por exemplo, a ambigüidade.

Para definir o tempo de falha é preciso fixar o tempo de início do estudo, a escala de medida a ser usada e estabelecer o evento de interesse que freqüentemente é indesejável e conhecido como falha. Na maioria das vezes, a falha (evento de interesse) é a morte do indivíduo ou recidiva da doença.

A falha pode ocorrer devido a um motivo ou a vários motivos. Situações de estudo, em que vários motivos podem causar a falha, são conhecidas como riscos competitivos e, conseqüentemente, como riscos múltiplos.

2.1.2 Censura

A censura é a principal característica de dados de sobrevivência. Sua ocorrência se dá, geralmente, em estudos clínicos que, mesmo sendo longos, terminam antes que todos os indivíduos tenham falhado, resultando em observações parciais ou incompletas, ou seja, em censura.

As observações censuradas, mesmo sendo parciais, possuem informações do tempo de vida dos pacientes e o seu não uso poderá gerar conclusões viciadas.

As censuras podem ser classificadas como do Tipo I, Tipo II e Aleatória.

(i) Censura tipo I

No início do experimento é pré-estabelecido um tempo τ em que o experimento será terminado. Logo, se o indivíduo falhar no tempo t , então $t < \tau$, t é dito ser tempo de falha, mas se o indivíduo não falhou ao término do experimento, então $t = \tau$ é tempo de censura.

Este tipo de censura pode ocorrer em dois tipos de situações:

1. Quando os n indivíduos entram no estudo ao mesmo tempo e é pré-estabelecido o término. Então, todos os indivíduos têm o mesmo tempo de censura potencial, isto é, são seguidos por um tempo τ . Quando os indivíduos entram em tempos diferentes no estudo, mas também são observados pelo mesmo tempo pré-estabelecido τ , ocorre a **censura simples do Tipo I**. Logo, os tempos observados podem ser representados pelo par (t_i, δ_i) , em que $i = 1, 2, \dots, n$ e

$$t_i = \min(T_i, \tau) \quad \text{e} \quad \delta_i = \begin{cases} 1 & \text{se } t_i = T_i \\ 0 & \text{se } t_i = \tau. \end{cases}$$

2. Quando os n indivíduos entram em tempos diferentes no estudo e é pré-estabelecido o tempo do término do experimento τ , então, para cada indivíduo é definido um tempo de observação $L_1, L_2, \dots, L_n$, pois, o tempo pré-estabelecido τ começa a ser contado no momento em que cada indivíduo entra no estudo. Esta situação caracteriza as **censuras múltiplas do Tipo I**. Os tempos observados podem ser representados pelo par (t_i, δ_i) , em que $i = 1, 2, \dots, n$, L_i são tempos potenciais limitados por τ e

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

(ii) Censura tipo II

A censura Tipo II é caracterizada por pré-estabelecer o número de falhas que irá ocorrer no estudo no início do experimento. Então, ao ocorrer o número de falhas desejado, o pesquisador encerra o estudo e todos os indivíduos que não falharam nesse período são considerados censurados.

(iii) Censura aleatória

A censura aleatória ocorre quando o indivíduo em estudo não falha por motivos relacionados ao acaso. Por exemplo, quando ele morre por uma causa diferente daquela estudada, ou seja, por uma causa que não tem influência sob o evento de interesse; ou quando o indivíduo deixa o estudo por um motivo qualquer, como: mudar de cidade, não ter condições financeiras, ou mesmo deixar o estudo por vontade própria.

2.1.3 Variáveis explicativas

Na análise de dados de sobrevivência, encontram-se o tempo de sobrevivência, a variável indicadora de censura e, muitas vezes, também, leva-se em consideração a presença de variáveis explicativas ou covariáveis, que podem influenciar o tempo de sobrevivência ou de censura. É importante detectar se as covariáveis estão relacionadas ao tempo de sobrevivência ou de censura, ou mesmo se estão relacionadas entre si. Frequentemente, as variáveis explicativas ou covariáveis são: sexo, idade, dependência química, entre outros fatores.

2.1.4 Representação dos dados de sobrevivência

Em análise de sobrevivência cada indivíduo é representado pelo par (t_i, δ_i) , sendo que t_i é o tempo de falha ou censura e δ_i é a variável indicadora de falha ou censura, ou seja,

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é tempo de falha} \\ 0 & \text{se } t_i \text{ é tempo de censura.} \end{cases}$$

Ao considerar situações de estudo em que há a presença de variáveis explicativas, estas são indicadas por $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$. Então os dados são representados por $(t_i, \delta_i, \mathbf{x}_i^T)$, sendo que, $i = 1, \dots, n$ indica o número de indivíduos, e $r = 0, \dots, p$ indica o número de variáveis explicativas no modelo.

2.1.5 Especificando o tempo de sobrevivência

A variável aleatória não-negativa, e geralmente, contínua T , representa o tempo de vida, ou seja, tempo de sobrevivência de um indivíduo. Pode ser representada pela função densidade de probabilidade, $f(t)$; função de sobrevivência, $S(t)$; função risco, $h(t)$; e por relações existentes entre essas três funções.

2.1.5.1 Função densidade de probabilidade

A função densidade de probabilidade é definida como o limite da probabilidade de um indivíduo morrer em um intervalo de tempo $[t + \Delta t)$ por unidade de Δt (comprimento do intervalo), ou simplesmente por unidade de tempo. É expressa por (LEE, 1992):

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}. \quad (1)$$

2.1.5.2 Função de sobrevivência

A função de sobrevivência, denotada por $S(t)$, é definida como a probabilidade de um indivíduo sobreviver a um tempo t , ou seja, a probabilidade de um indivíduo não falhar até um certo tempo t . Ela é expressa por:

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx, \quad (2)$$

sendo que $S(t)$ é uma função contínua monótona decrescente, com as seguintes propriedades: $S(0) = 1$ e $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ (LAWLESS, 2003). É conhecida, também, como a taxa de sobrevivência acumulada.

2.1.5.3 Função de risco

A função de risco é definida como o limite da probabilidade de um indivíduo falhar no intervalo de tempo $[t, \Delta t)$, assumindo que este mesmo indivíduo sobreviveu até o tempo t , dividida pelo comprimento do intervalo e é representada por (LAWLESS, 2003):

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (3)$$

Pode, ainda, ser expressa em termos da função densidade de probabilidade e da função de sobrevivência, isto é,

$$h(t) = \frac{f(t)}{S(t)}. \quad (4)$$

A função de risco descreve como a probabilidade instantânea de falha (taxa de falha) se modifica com o passar do tempo. É conhecida como taxa de falha instantânea, força de mortalidade e taxa de mortalidade condicional (COX; OAKES, 1984).

A função de risco é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções taxa de falha podem diferir drasticamente. Dessa forma, a modelagem da função taxa de falha é um importante método para dados de sobrevivência (COLOSIMO; GIOLO, 2006), pois pode ter forma crescente, decrescente, constante ou não monótona.

2.1.6 Relações entre as funções

As funções definidas anteriormente estão matematicamente relacionadas entre si. Algumas relações importantes:

$$f(t) = \frac{dF(t)}{dt}.$$

Sabe-se que a função de distribuição acumulada é $F(t) = 1 - S(t)$, então

$$f(t) = \frac{d[1 - S(t)]}{dt} = -S'(t).$$

Considerando $f(t) = -S'(t)$ e a expressão (4) tem-se

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d[\log S(t)]}{dt}.$$

Logo,

$$\log S(t) = -\int_0^t h(u) du. \quad (5)$$

Uma função importante é a função de risco acumulada, expressa por

$$H(t) = \int_0^t h(u)du. \quad (6)$$

Então das expressões (5) e (6) tem-se que

$$S(t) = \exp\{H(t)\}. \quad (7)$$

As expressões (4) e (7) são muito utilizadas para a realização de cálculos importantes neste trabalho.

2.1.7 Distribuições importantes em análise de sobrevivência

O comportamento dos tempos de falhas ou ocorrência de algum evento, em geral, pode ser acomodado ou aproximado por uma distribuição de probabilidade teórica. Na literatura, várias distribuições têm sido amplamente usadas para descrever tempos de sobrevivência.

2.1.7.1 Distribuição Weibull

A distribuição Weibull é muito utilizada em estudos relacionados ao tempo de falha devido à grande aplicabilidade tanto na área de confiabilidade, bem como na análise de sobrevivência. É, talvez, a distribuição mais amplamente utilizada em análise de tempos de vida, porque a função de risco pode ser uma função monótona, ou seja, crescente, decrescente ou constante.

Uma variável aleatória T com distribuição Weibull tem a função densidade de probabilidade expressa por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}, \quad (8)$$

sendo que $t \geq 0$, $\alpha > 0$ é o parâmetro de escala e $\gamma > 0$ é o parâmetro de forma. As funções de sobrevivência, risco e o p -ésimo percentil são, respectivamente, expressos por:

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\},$$

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1},$$

e

$$t_p = \alpha[-\log(1-p)]^{1/\gamma},$$

com $t \geq 0$, $\alpha > 0$ e $\gamma > 0$. Para $\gamma = 1$, tem-se a distribuição exponencial, como caso particular.

A esperança e a variância da distribuição Weibull são representadas, respectivamente, por: $E[T] = \alpha\Gamma[1 + (1/\gamma)]$ e $Var[T] = \alpha^2[\Gamma[1 + (2/\gamma)] - \Gamma[1 + (1/\gamma)]^2]$, sendo que $\Gamma[k]$ é a função gama que é definida como $\Gamma[k] = \int_0^\infty x^{k-1} \exp\{-x\} dx$.

A distribuição Weibull possui uma importante relação com a distribuição do valor extremo, ou seja, se T é uma variável aleatória com distribuição Weibull, tem-se que $Y = \log(T)$, tem distribuição do valor extremo com função densidade dada por:

$$f(y) = \frac{1}{\sigma} \exp\left\{\left(\frac{y - \mu}{\sigma}\right) - \exp\left\{\frac{y - \mu}{\sigma}\right\}\right\}, \quad (9)$$

sendo que y e μ , parâmetro de locação, $\in \Re$ e $\sigma > 0$, parâmetro de escala. Quando $\mu = 0$ e $\sigma = 1$ tem-se a distribuição valor extremo padrão.

A função de sobrevivência, risco e o p -ésimo percentil da distribuição do valor extremo são dados, respectivamente, por:

$$S(y) = \exp\left\{-\exp\left\{\frac{y - \mu}{\sigma}\right\}\right\},$$

$$h(y) = \frac{1}{\sigma} \exp\left\{\frac{y - \mu}{\sigma}\right\},$$

e

$$t_p = \mu + \sigma \log[-\log(1-p)].$$

Em algumas situações é coerente utilizar o logaritmo do tempo, e devido a isso, alguns *softwares* trabalham com a distribuição do valor extremo em lugar da Weibull. Portanto, é importante saber a relação existente entre elas, ou seja, $\gamma = 1/\sigma$ e $\alpha = \exp\{\mu\}$.

2.1.7.2 Distribuição log-logística

Uma variável aleatória T com distribuição log-logística tem a função densidade de probabilidade dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} [1 + (t/\alpha)^\gamma]^{-2}, \quad (10)$$

sendo que $t > 0$, $\alpha > 0$ é o parâmetro de forma e $\gamma > 0$ é o parâmetro de escala. As funções de sobrevivência, risco e o p -ésimo percentil são, respectivamente, expressas por:

$$S(t) = \frac{1}{1 + (t/\alpha)^\gamma},$$

$$h(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^\gamma]},$$

e

$$t_p = \alpha \left[\frac{p}{(1-p)} \right]^{\frac{1}{\gamma}}.$$

A esperança e a variância da distribuição log-logística são representadas por:

$$E[T] = [\pi\alpha \csc(\pi/\gamma)]/\gamma \text{ para } \gamma > 1 \text{ e } Var[T] = [(2\pi\alpha^2 \csc(2\pi/\gamma))/\gamma] - E[T]^2.$$

Em algumas situações é conveniente trabalhar com o logaritmo do tempo. Então, se T é uma variável aleatória com distribuição log-logística, o logaritmo de T , $Y = \log(T)$, segue uma distribuição logística com função densidade expressa por:

$$f(y) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \left[1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right]^{-2}, \quad (11)$$

sendo que $-\infty < \mu < \infty$ e $\sigma > 0$, parâmetros de locação e escala, respectivamente. Consequentemente, as funções de sobrevivência e risco são dadas por:

$$S(t) = \frac{1}{1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}},$$

e

$$h(t) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma}\right\} \left[1 + \exp\left\{\frac{y-\mu}{\sigma}\right\}\right]^{-1}.$$

Os parâmetros das distribuições estão relacionados da seguinte forma: $\gamma = 1/\sigma$ e $\alpha = \exp\{\mu\}$. O que é importante, pois freqüentemente, alguns pacotes estatísticos trabalham com essa relação.

2.2 Modelos de riscos múltiplos

Os modelos de riscos múltiplos, dentre eles, Weibull múltiplo e o log-logístico múltiplo (LOUZADA-NETO, 1999), devido às suas flexibilidades em acomodar uma rica classe de formas de riscos são modelos alternativos importantes que podem ser utilizados nos mais variados problemas de modelagem de dados de sobrevivência e confiabilidade. Em geral, os modelos de riscos múltiplos são aplicados em problemas específicos, como em riscos competitivos, sistemas mascarados e riscos complementares nos quais existe informação total ou parcial a respeito das causas das falhas. A vantagem desse modelo é sua flexibilidade em relação ao modelo de risco usual, pois acomoda não apenas funções risco crescente, decrescente ou constante, mas também função risco não-monótona, como, por exemplo, forma de “banheira” e curvas multimodal. Uma forma de verificar a distribuição adequada aos dados, é fazer o gráfico *TTT plot* (LAWLESS, 2003; MUDHOLKAR et al., 1995). Alguns exemplos podem ser apreciados nas Figuras 1, 2 e 3.

Na Figura 1, por exemplo, observa-se uma curva inicialmente convexa e depois côncava, indicando que a função de risco tem forma de “U”. Nesse caso, as distribuições Weibull Exponenciada, Weibull aditiva, Nova Weibull extendida e Burn XII aditiva são candidatas possíveis a modelar os dados.

Na Figura 2, observa-se uma curva inicialmente côncava e depois convexa. Assim a função de risco parece ser unimodal e, possíveis candidatas para o ajuste aos dados são as distribuições log-logística, log-normal e Burn XII.

Já na Figura 3, observa-se uma forma convexa, com uma reta constante no final. Nesse caso, o modelo poly log-logístico é mais adequado.

O modelo de riscos múltiplos supõe que um indivíduo ou sistema de indivíduos está sujeito a $k \geq 2$ motivos diferentes e independentes que levará à falha, sendo que o motivo responsável pela ocorrência do evento de interesse ou falha é completa ou parcialmente desconhecido.

Figura 1 - Gráfico TTT *plot* para pacientes em um estudo oncológico

Figura 2 - Gráfico TTT *plot* para pacientes com câncer submetidos a radioterapia

Figura 3 - Gráfico TTT *plot* para tempo de sobrevivência de peixes

O tempo de falha (ou tempo de vida) relacionado ao j -ésimo motivo ocorrer é representado por Z_j , $j = 1, 2, \dots, k$, ou seja, $\mathbf{s}_i^T = (Z_1, Z_2, \dots, Z_k)$, sendo que $i = 1, 2, \dots, n$ e $\mathbf{s}_i^T$ representa os k motivos de falha associados ao i -ésimo indivíduo ou objeto em estudo. A quantidade observada para cada indivíduo é $T_i = \min(Z_1, Z_2, \dots, Z_k)$, ou seja, é utilizado o tempo do primeiro motivo que causou a falha do indivíduo ou objeto em estudo, e os Z_i 's são assumidos independentes.

Dado que, tempo de vida Z_j , tem função densidade $f_j(z)$ e, conseqüentemente, uma função risco $h_j(z)$, e assumindo os motivos de falhas independentes, a função risco global do modelo de riscos múltiplos é representada por:

$$h(t) = \sum_{j=1}^k h_j(t),$$

e Louzada (1999) propõe

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j - 1}}{\mu_j^{\omega_j}} q_j(t; \mu_j, \omega_j, \kappa), \quad (12)$$

sendo que $t > 0$ e μ_j, ω_j, κ são parâmetros positivos e não conhecidos e $q_j(\cdot)$ são funções monótonas de parâmetro de forma conhecido igual a um quando todos os demais argumentos são iguais a zero.

Na literatura, o método de estimação mais conhecido é o método de mínimos quadrados, mas este é inapropriado para estudos de tempo de vida, pois é incapaz de incorporar censuras no seu processo de estimação. Então, uma opção para este tipo de dados é o método de máxima verossimilhança, que incorpora as observações censuradas e possui propriedades ótimas para grandes amostras.

O método de máxima verossimilhança escolhe o valor do parâmetro que maximiza a probabilidade de obter a amostra observada, ou seja, o valor que melhor explica a amostra observada. Sendo assim, considere uma amostra de variáveis aleatórias independentes $T_1, T_2, \dots, T_n$, de uma particular população de interesse, tal que $T_i = \min(Z_1, Z_2, \dots, Z_k)$, $i = 1, 2, \dots, n$, $k \geq 2$ e associado a cada T_i tem-se uma variável indicadora de censura δ_i , sendo que $\delta_i = 1$ se t_i é uma observação do tempo de falha e, $\delta_i = 0$ se t_i é uma observação censurada.

Assim, a função de verossimilhança baseada nos tempos de falhas e nos tempos de censuras, $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$, podem ser divididas em dois grupos, as r primeiras ordenadas são as não-censuradas e as $n - r$ seguintes são as censuradas. Considerando censura não informativa, a função de verossimilhança é expressa por:

$$L(t_i; \boldsymbol{\theta}) = \prod_{i=1}^r f(t_i; \boldsymbol{\theta}) \prod_{i=r+1}^n S(t_i; \boldsymbol{\theta}),$$

ou equivalente a:

$$\begin{aligned} L(t_i; \boldsymbol{\theta}) &= \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{\delta_i} \\ &= \prod_{i=1}^n [h(t_i; \boldsymbol{\theta})]^{\delta_i} S(t_i; \boldsymbol{\theta}), \end{aligned} \tag{13}$$

em que $h(t_i)$ é definida em (12) e $\boldsymbol{\theta}$ é o vetor de parâmetros. Nesse caso, a contribuição de cada observação censurada é sua função de sobrevivência, sendo essa, representada por $S(t_i) = \prod_{j=1}^k S_j(t_i)$, que é a função de sobrevivência global do modelo de riscos múltiplos.

Os estimadores de máxima verossimilhança são os valores de $\boldsymbol{\theta}$ que maximizam $L(\boldsymbol{\theta})$, ou equivalentemente, o logaritmo de $L(\boldsymbol{\theta})$. Assim, eles são encontrados resolvendo o

seguinte sistema de equações:

$$U(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0.$$

2.2.1 Modelo Weibull múltiplo

Como caso particular da equação (12), o modelo Weibull múltiplo é caracterizado quando a função de risco global é representada por:

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1}}{\mu_j^{\omega_j}}, \quad (14)$$

pois,

$$q_j(t; \mu_j, \omega_j, \kappa) = \frac{\Gamma(\kappa)^{-1} u_j^{\kappa-1} \exp\{-u_j\}}{1 - I(\kappa; u_j)},$$

e, sendo que $u_j = (\frac{t}{\mu_j})^{\omega_j}$ e, $I(\kappa; u_j) = \Gamma(\kappa)^{-1} \int_0^{u_j} a^{\kappa-1} \exp\{-a\} da$ é a função Gama Incompleta. Portanto, quando se tem $\kappa = 1$, então $q_j(\cdot) = 1$.

Berger e Sun (1993) dizem que, quando os parâmetros de forma, $\omega_1, \omega_2, \dots, \omega_k$, são iguais, o modelo Weibull múltiplo torna-se um modelo Weibull simples e os parâmetros $\mu_1, \mu_2, \dots, \mu_k$ não são identificáveis. Para evitar esse problema, Davison e Louzada (2000) propõem a reparametrização $\alpha_j = \mu_j^{\omega_j}$, resultando em

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1}}{\alpha_j}.$$

Geralmente, na prática existe a presença de covariáveis (ou variáveis explicativas), representadas por $\mathbf{x}_i^T$, ficando a função de risco global do modelo Weibull múltiplo dada por:

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1}}{\alpha_j} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}.$$

Mazucheli; Louzada e Achcar (2001) propõem uma transformação no modelo de riscos múltiplos na presença de covariáveis, representada por $\beta_{0j} = -\log(\alpha_j)$. Logo a função

de risco global do modelo Weibull múltiplo é,

$$h(t) = \sum_{j=1}^k \omega_j t^{\omega_j-1} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}, \quad (15)$$

sendo que $\boldsymbol{\beta}_j^T = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \dots, \beta_{pj})$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_k)$, $\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ip})$, $\boldsymbol{\omega}^T = (\omega_1, \omega_2, \dots, \omega_k)$ e $\mathbf{x}_i^T \boldsymbol{\beta}_j = \beta_{0j} + \beta_{1j}x_{i1} + \dots + \beta_{pj}x_{ip}$.

Considere-se um amostra aleatória $t_1, t_2, \dots, t_n$, com função de risco dada em (15), tal que associado a t_i exista uma variável indicadora de censura δ_i e um vetor de variáveis regressoras $\mathbf{x}_i^T$. Então, a função de verossimilhança do modelo Weibull múltiplo é representada por:

$$L(t_i; \boldsymbol{\omega}, \boldsymbol{\beta}) = \prod_{i=1}^n \left[\sum_{j=1}^k \omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right]^{\delta_i} \exp\left\{ - \sum_{i=1}^n \sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right\},$$

sendo que $\boldsymbol{\theta} = (\boldsymbol{\omega}^T, \boldsymbol{\beta}^T)^T$, $S_j(t_i) = \exp\{-t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}\}$ e $S(t_i) = \prod_{j=1}^k S_j(t_i)$ que representa a função de sobrevivência global do modelo Weibull múltiplo.

Assim, o logaritmo da função de verossimilhança pode ser dado por

$$\begin{aligned} l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta}) &= \sum_{i:\delta_i=1} \log \left[\sum_{j=1}^k \omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] - \sum_{i:\delta_i=1} \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] \\ &\quad - \sum_{i:\delta_i=0} \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right]. \end{aligned} \quad (16)$$

2.2.2 Modelo log-logístico múltiplo

Como caso particular da equação (12), o modelo log-logístico múltiplo é caracterizado quando a função risco global é representada por:

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1}}{\mu_j^{\omega_j} + t^{\omega_j}}, \quad (17)$$

para

$$q_j(t; \mu_j, \omega_j, \kappa) = \frac{1}{1 + u_j}.$$

Considerando a presença de variáveis regressoras, a expressão (17) pode ser escrita como,

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}{\mu_j + t^{\omega_j} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}.$$

Mazucheli; Louzada e Achcar (2001) propõem uma transformação no modelo de riscos múltiplos na presença de variáveis explicativas, representada por $\beta_{0j} = -\log(\alpha_j)$. Logo, a função de risco global do modelo log-logístico múltiplo é,

$$h(t) = \sum_{j=1}^k \frac{\omega_j t^{\omega_j-1} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}{1 + t^{\omega_j} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}, \quad (18)$$

sendo que $\boldsymbol{\beta}_j^T = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \dots, \beta_{pj})$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_k)$, $\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ip})$, $\boldsymbol{\omega}^T = (\omega_1, \omega_2, \dots, \omega_k)$ e $\mathbf{x}_i^T \boldsymbol{\beta}_j = \beta_{0j} + \beta_{1j}x_{i1} + \dots + \beta_{pj}x_{ip}$.

Considere uma amostra aleatória $t_1, t_2, \dots, t_n$, com função risco dada em (26), tal que associado a t_i exista uma variável indicadora de censura δ_i e um vetor de variáveis regressoras $\mathbf{x}_i^T$. Então, a função de verossimilhança do modelo log-logístico múltiplo é representada por:

$$L(t_i; \boldsymbol{\omega}, \boldsymbol{\beta}) = \prod_{i=1}^n \left[\sum_{j=1}^k \frac{\omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}}{1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}} \right]^{\delta_i} \prod_{j=1}^k \left[1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right]^{-1},$$

sendo que $\boldsymbol{\theta} = (\boldsymbol{\omega}^T, \boldsymbol{\beta}^T)^T$, $S_j(t_i) = (1 + t_i^{\omega_j} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j})^{-1}$ e $S(t_i) = \prod_{j=1}^k S_j(t_i)$ que representa a função de sobrevivência global do modelo log-logístico múltiplo.

Assim, o logaritmo da função de verossimilhança fica:

$$\begin{aligned} l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta}) &= \sum_{i:\delta_i=1} \log \left[\sum_{j=1}^k \frac{\omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}}{1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}} \right] - \sum_{i:\delta_i=1} \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right] \\ &\quad - \sum_{i:\delta_i=0} \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right]. \end{aligned} \quad (19)$$

2.2.3 Inferência

A Inferência Estatística tem como objetivo estudar métodos que permitam tirar conclusões sobre os parâmetros desconhecidos de uma população, a partir da análise de dados

amostrais. Assim, as inferências de interesse são baseadas em informações ou quantidades obtidas de uma amostra aleatória selecionada da população.

Através de cálculos das estimativas da variância de $\hat{\boldsymbol{\theta}}$ pode-se construir testes de hipóteses e intervalos de confiança para os parâmetros, utilizando o fato que $\hat{\boldsymbol{\theta}}$ tem distribuição assintótica normal multivariada, sob certas condições de regularidade, com média $\boldsymbol{\theta}$ e matriz de variância e covariância $\mathbf{I}^{-1}(\boldsymbol{\theta})$, ou seja,

$$\hat{\boldsymbol{\theta}} \sim N_{[k+k(p+1)]}(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta})),$$

em que $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$, tal que

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

Por outro lado, sabe-se que $\mathbf{I}(\boldsymbol{\theta})$, denominada matriz de informação de Fisher, não é possível calcular devido a presença de observações censuradas, então pode-se utilizar, alternativamente, a matriz $[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, denominada matriz de informação observada, a qual é uma estimativa consistente de $\mathbf{I}(\boldsymbol{\theta})$. Nesse caso, a matriz $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ tem a seguinte forma,

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} -\ddot{\mathbf{L}}_{\omega\omega} & -\ddot{\mathbf{L}}_{\omega\beta} \\ -\ddot{\mathbf{L}}_{\beta\omega} & -\ddot{\mathbf{L}}_{\beta\beta} \end{pmatrix},$$

sendo cada submatriz expressa de forma fechada para os modelos de riscos múltiplos.

Assim, um intervalo de confiança de $(1 - \alpha)100\%$ para um θ_j é expresso por:

$$\hat{\theta}_j \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta}_j)}.$$

Ao obter as estimativas e intervalos de confiança para os parâmetros, em geral, tem-se o interesse em testar hipóteses relacionadas ao vetor $\boldsymbol{\theta}$ de parâmetros. Sendo assim, uma hipótese de interesse é representada por:

$$\begin{aligned} H_0 : \theta_j &= \theta_{j0} \\ H_1 : \theta_j &\neq \theta_{j0}. \end{aligned} \tag{20}$$

Esse teste é baseado na distribuição assintótica dos estimadoras, e é uma generalização do teste t de Student. Então, para grandes amostras a distribuição t aproxima-se da distribuição normal, sendo a estatística para testar as hipóteses (20)

$$z = \frac{\hat{\theta}_j - \theta_{j0}}{\sqrt{\hat{Var}(\hat{\theta}_j)}},$$

em que $Z \sim N(0, 1)$.

Os procedimentos inferências descritos serão utilizados na obtenção das estimativas dos modelos de riscos múltiplos na seção (3.2.1).

2.3 Influência local

Uma importante etapa na análise estatística de modelos é verificar o quanto as estimativas obtidas a partir do modelo proposto são resistentes a pequenas perturbações nos dados, ou no modelo. Se o modelo ajustado não apresentar uma boa descrição dos dados que foram observados, o mesmo pode conduzir a inferências errôneas.

Devido a isso, destaca-se a importância de realizar um estudo sobre a robustez dos resultados obtidos, a nível de vários aspectos que envolvem a formulação do modelo, e as estimativas dos seus parâmetros, ou seja, realizar uma análise de diagnóstico.

Uma observação pode ser considerada influente quando altera os resultados da análise estatística do modelo ao ser omitida no ajuste do modelo ou ao ser submetida a uma perturbação.

Cook (1986) propõe avaliar a influência conjunta das observações sob pequenas mudanças (perturbações) no modelo, ao invés da avaliação pela retirada individual ou conjunta de pontos, através do afastamento pela verossimilhança. Existe inúmeras formas alternativas de se perturbar o modelo proposto, cuja escolha deve levar em consideração quais os aspectos da análise se deseja monitorar. Obviamente, deve-se pensar em esquemas de perturbação que sejam interpretáveis.

Lesafre e Verbeke (1998) sugerem expressar a função de verossimilhança em

função das perturbações, teoria conhecida como ponderação de casos, e também propõe calcular a curvatura normal para o i -ésimo indivíduo. Na próxima seção é descrito a teoria de influência proposta por Cook (1986), Lesafre e Verbeke (1998), entre outros.

2.3.1 Método de influência local

O método de influência local introduzido por Cook (1986) propõe o estudo de uma perturbação dos componentes do modelo, ou dos dados, utilizando a função de afastamento pela verossimilhança do modelo estatístico.

Seja $l(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança do modelo postulado, e considere um vetor de perturbação $\mathbf{m}$ pertencente a um subconjunto aberto $\Omega \subseteq \mathbf{R}^b$, então, o logaritmo da função de verossimilhança no modelo postulado é $l(\boldsymbol{\theta}|\mathbf{m})$. Assume-se que $l(\boldsymbol{\theta}|\mathbf{m})$ é duas vezes continuamente diferenciável em $(\boldsymbol{\theta}, \mathbf{m})^T$, e que o modelo postulado está encaixado no modelo perturbado. Ou seja, é suposto que existe $\mathbf{m}_0 \in \Omega$ tal que $l(\boldsymbol{\theta}|\mathbf{m}_0) = l(\boldsymbol{\theta}) \forall \boldsymbol{\theta} \in \mathbf{R}^{k+k(p+1)}$. Em geral, a dimensão b , do vetor de perturbação, está relacionada com a dimensão do vetor $\boldsymbol{\theta}$, ou com o tamanho da amostra, depende do esquema de perturbação.

No modelo perturbado, será denotado o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}_m$, e o objetivo, agora, é comparar $\hat{\boldsymbol{\theta}}$ com $\hat{\boldsymbol{\theta}}_m$, quando $\mathbf{m}$ varia em Ω . Se a distância entre eles permanecer pequena quando $\mathbf{m}$ varia em Ω , isso é uma indicação de que existe uma estabilidade do modelo ajustado no que diz respeito ao particular esquema de perturbação que foi utilizado. E por tanto, ao correspondente aspecto da análise que esta sendo monitorado.

Uma comparação direta entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_m$ pode não ser simples devido a diversos fatores, tais como, diferença de escala, unidade de medida, erro de estimativa, etc. Cook (1986) sugere que essa comparação seja feita através da função

$$LD(\mathbf{m}) = 2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_m)],$$

que é uma generalização do afastamento pela verossimilhança e $\mathbf{m} \in \Omega$. Mas na análise de influência usa-se muito a função

$$LD(i) = 2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{-i})],$$

assim, o sentido da distância entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_m$ passa a depender da concavidade do logaritmo da função de verossimilhança, Cook (1986).

Por outro lado, como $\hat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança para $\boldsymbol{\theta}$ no modelo sem perturbação, segue-se que $LD(\mathbf{m}) \geq 0 \forall \mathbf{m} \in \Omega$. Além disso, como $LD(\mathbf{m}_0) = 0$, conclui-se que $\mathbf{m}_0$ é um ponto de mínimo local da função afastamento pela verossimilhança.

Uma análise sobre o comportamento geométrico da função $LD(\mathbf{m})$ quando $\mathbf{m}$ varia em Ω pode fornecer informações bastantes relevantes a respeito de características importantes do modelo sob investigação.

Assim sendo, a idéia de Cook (1986) consiste em estudar o comportamento da função $LD(\mathbf{m})$ em uma vizinhança de $\mathbf{m}_0$, que é o ponto em que as duas verossimilhanças são iguais. O procedimento consiste em escolher uma direção unitária $\mathbf{d}$, $\|\mathbf{d}\| = 1$, e assim considerar o gráfico de $LD(\mathbf{m}_0 + a\mathbf{d})$ e a , sendo que $a \in \Re$. Esse gráfico é chamado de linha projetada. Cada linha pode ser investigada pela curvatura normal $C_{\mathbf{d}}$ em torno de $a = 0$, sendo esta considerada da superfície geométrica expressa por:

$$\boldsymbol{\alpha}(\mathbf{m}) = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ LD(\mathbf{m}) \end{pmatrix}.$$

O autor sugere considerar a direção $\mathbf{d}_{max}$ correspondente à maior curvatura $C_{\mathbf{d}_{max}}$. Um gráfico de $\mathbf{d}_{max}$ contra a ordem das observações pode revelar os elementos que sob pequenas perturbações exercem maior influência sobre $LD(\mathbf{m})$.

Segundo Cook (1986) a curvatura normal na direção $\mathbf{d}$ pode ser representada por:

$$C_{\mathbf{d}} = 2\|\mathbf{d}^T \boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta} \mathbf{d}\|,$$

em que

$$\boldsymbol{\Delta} = \frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{m})}{\partial \boldsymbol{\theta} \partial \mathbf{m}^T} \quad (21)$$

e

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}, \quad (22)$$

ambas matrizes avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$. Através da equação (21) é possível avaliar a influência que pequenas perturbações podem exercer sobre os componentes do modelo, como estimativas dos parâmetros e outros resultados da análise.

Uma informação importante é saber a direção que produz a maior influência local na estimativa dos parâmetros que é dada pelo $\mathbf{d}_{max}$, sendo este o autovetor normalizado correspondente ao maior autovalor $C_{\mathbf{d}_{max}}$ da matriz $F = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$. O autovetor $\mathbf{d}_{max}$ identifica as observações mais influentes para o esquema de perturbação considerado.

Lesaffre e Verbeke (1998) sugerem considerar a direção do i -ésimo indivíduo que corresponderia a $\mathbf{d}_i = (0, \dots, 1, \dots, 0)^T$, tal que o i -ésimo elemento é um. Sendo assim, a curvatura normal chamada influência local total do i -ésimo indivíduo, é representada por:

$$C_i = 2|\boldsymbol{\Delta}_i^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}_i|,$$

considerando $\mathbf{d}_{max}$ e o gráfico de C_i contra a ordem das observações como diagnóstico em influência local. Os possíveis pontos influentes são aqueles tal que $C_i \geq 2\bar{C}$ em que $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$.

Em alguns casos, existe a necessidade de trabalhar com um subconjunto do vetor de parâmetros $\boldsymbol{\theta}$. Então, faz-se o uso da partição, ou seja, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ e o interesse está em $\boldsymbol{\theta}_1$. Logo, a curvatura na direção $\mathbf{d}$ é representada por:

$$C_{\mathbf{d}} = 2\|\mathbf{d}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta} \mathbf{d}\|$$

sendo

$$\mathbf{B}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix}, \quad (23)$$

em que $\ddot{\mathbf{L}}_{22}$ é determinada pela partição de $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ considerando a partição de $\boldsymbol{\theta}$.

A influência local total do i -ésimo indivíduo é dada por:

$$C_i = 2|\boldsymbol{\Delta}_i^T (\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}_i|,$$

sendo que $\mathbf{B}_{22}$ considerado como em (23).

Na literatura existem diversas metodologias de influência local que surgiram do artigo de Cook (1986), por exemplo, Thomas e Cook (1989,1990), Paula (2004), Ortega (2001), Ortega et al.(2003), Ortega et al.(2006), dentre outros.

Cook; Peña e Weisberg (1988) comparam o afastamento da verossimilhança com medidas tradicionais de deleção de pontos, tais como a distância de Cook e o DFFITS_i; esse último proposto por Belsley; Kuh e Welsch (1980).

Fung e kwan (1997) apresentam uma discussão interessante sobre a aplicação de influência local para outras medidas de influência diferentes do afastamento pela verossimilhança. Eles mostram que uma medida de influência, digamos $\hat{T}_{\mathbf{m}}$ é invariante de escala se

$$\frac{d(\hat{T}_{\mathbf{m}})}{d\mathbf{m}} = 0,$$

avaliado em $\mathbf{m} = \mathbf{m}_0$.

Quando esse derivada é diferente de zero a ordem entre as componentes de $\mathbf{d}_{max}$ não é necessariamente preservada sob mudanças de escala. Em particular para o afastamento pela verossimilhança, tem-se que

$$\frac{d(LD(\mathbf{m}))}{d\mathbf{m}} = 0,$$

avaliado em $\mathbf{m} = \mathbf{m}_0$.

Essa propriedade segue também, por exemplo, para as medidas de influência proposta por Thomas e Cook (1990). Porém ela não vale para outras medidas de influência discutidas por Fung e Kwan (1997).

2.4 Análise de resíduos

A análise de resíduos tem a finalidade de avaliar a adequação da distribuição proposta para a variável resposta, verificar as suposições do erro, ou seja, verificar as suposições

básicas do modelo, bem como detectar a presença de pontos extremos, identificar a relevância de um fator adicional omitido e analisar a forma funcional do modelo.

Como sabe-se quando temos observações censuradas, devemos ter métodos estatísticos especiais. A metodologia empregada para dados lineares não faz sentido para dados com presença de censuras. Essa particularidade está presente também quando se avalia os resíduos referente a modelos com presença de censuras.

Para modelos lineares, o resíduo mais utilizado é o de Cox e Snell (1968) que são resíduos calculado a partir de $(Y - \hat{X}\boldsymbol{\theta})$, em que Y é a variável resposta, X é a matriz de dados e $\hat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança. Logo, a partir deste resultado muitos resíduos foram calculados.

Em análise de sobrevivência os resíduos mais utilizados são os de martingal e deviance introduzidos por Therneau et al. (1990).

2.4.1 Resíduo martingal

Quando se procura ajustar um modelo a um conjunto de dados, a avaliação desse ajuste passa pela análise de uma estatística especial, que tem a finalidade de medir a qualidade do modelo ajustado, uma vez que os resíduos são usados para identificar discrepâncias entre um modelo ajustado e o conjunto de dados, é conveniente buscar uma definição para resíduo que leve em consideração a contribuição de cada observação sobre medida de qualidade de ajuste.

Inicialmente, os resíduos martingal foram introduzidos nos processos de contagem, depois reescritos para modelos de regressão paramétricos na presença de dados censurados com a seguinte forma:

$$\hat{r}_{M_i} = \delta_i - \hat{H}_i(ti|x_i), \quad (24)$$

em que δ_i é a variável indicadora de falha e $\hat{H}_i(ti|x_i)$ é a função de risco acumulada obtida do modelo ajustado.

Esses resíduos são vistos como uma estimativa do número de falhas em excesso observada nos dados mas não predito pelo modelo. Os mesmos são usados, em geral, para examinar a melhor forma funcional (linear, quadrática etc.) para uma dada covariável em um modelo de regressão assumido para os dados sob estudo (COLOSIMO; GIOLO, 2006).

2.4.2 Resíduo deviance

Para os modelos de regressão paramétricos, os resíduos deviance são definidos por:

$$\hat{r}_{Di} = \text{sin}(\hat{r}_{M_i}) \left[-2 \left(\hat{r}_{M_i} + \delta_i \log(\delta_i - \hat{r}_{M_i}) \right) \right]^{1/2}. \quad (25)$$

Esses resíduos são uma tentativa de tornar os resíduos martingal mais simétricos em torno de zero, facilitam, em geral, a detecção de pontos atípicos (outliers). Se o modelo for apropriado, esses resíduos devem apresentar um comportamento aleatório em torno de zero. Gráficos dos resíduos martingal, ou deviance, contra os tempos fornecem uma forma de verificar a adequação do modelo ajustado, bem como auxiliam na detecção de observações atípicas (COLOSIMO; GIOLO, 2006).

2.5 Impacto das observações influentes detectadas

Para revelar o impacto das observações influentes detectadas, estima-se, novamente, os parâmetros sem as observações influentes. Seja, $\hat{\theta}$ e $\hat{\theta}^0$ as estimativas de máxima verossimilhança dos modelos que são obtidos do conjunto de dados com e sem as observações influentes, respectivamente. Lu e Song (2006), definem as seguintes quantidades para a diferença medida entre $\hat{\theta}$ e $\hat{\theta}^0$:

$$TRC = \sum_{i=1}^{n_p} \left| \frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i} \right| \quad \text{e} \quad MRC = \max_i \left| \frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i} \right|,$$

em que TRC é a chance relativa total, MRC é a chance relativa máxima, n_p é o número de parâmetros, e $LD_I(\theta) = 2\{l(\hat{\theta}) - l(\hat{\theta}_{(I)})\}$ é o afastamento da verossimilhança, sendo que

$\hat{\boldsymbol{\theta}}_{(I)}$ denota o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ depois de remover o conjunto de observações (I). (ver, COOK; PEÑA; WEISBERG, 1988).

A fim de comparar o impacto das observações influentes, é realizada a análise removendo o mesmo número de observações (I) aleatoriamente do modelo completo, encontrando novos valores para TRC , MRC e $LD_I(\boldsymbol{\theta})$. Então, comparam-se os resultados obtidos para cada medida, respectivamente, verificando a sensibilidade do grupo de observações influentes.

Todas as teorias abordadas em revisão de literatura serão usadas para desenvolver a análise de diagnóstico ao modelo de riscos múltiplos em material e métodos.

3 MATERIAL e MÉTODOS

3.1 Material

Os dados utilizados referem-se ao tempo de sobrevivência de peixes da espécie “Notropis Dourado, crysoleucas de Notemigonus”, obtidos dos experimentos de campo realizados no lago Saint Pierre, Quebeque, em 2005 (Laplante et al., dados não publicados). As variáveis envolvidas no estudo foram: y_i , tempo de sobrevivência em anos; δ_i , indicador de censura; x_{i1} , tamanho do peixe em cm; x_{i2} , profundidade do rio em cm e x_{i3} , transparência da água; em que $i = 1, 2, \dots, 106$.

3.2 Métodos

Para a realização deste trabalho é ajustado os modelos Weibull múltiplo e log-logístico múltiplo aos dados de peixes da espécie “Notropis Dourado, crysoleucas de Notemigonus”.

Aos modelos de riscos múltiplos são aplicadas as técnicas de influência local para verificar a sensibilidade dos estimadores a pequenas perturbações nos dados ou mesmo nos modelos, e para finalizar, verifica-se o ajuste do modelo através da metodologia de resíduos.

3.2.1 Modelos de riscos múltiplos

Para aplicar a teoria de riscos múltiplos enunciada na seção (2.2) é necessário estimar os parâmetros dos modelos e realizar testes estatísticos e intervalos de confiança para os mesmos. Então, é desenvolvida a metodologia de inferência estatística descrita na seção (2.2.3) utilizando o procedimento NLMIXED do *software* SAS, e o *software* OX através do procedimento MAXBFGS para estimar os parâmetros dos modelos Weibull múltiplo e log-logístico múltiplo.

Através da distribuição assintótica dos estimadores de máxima verossimilhança

é necessário calcular a matriz $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ dada na seção (2.2.3). Então, para os modelos de riscos múltiplos a matriz de informação observada é representada por:

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} -\ddot{\mathbf{L}}_{\omega\omega} & -\ddot{\mathbf{L}}_{\omega\beta} \\ -\ddot{\mathbf{L}}_{\beta\omega} & -\ddot{\mathbf{L}}_{\beta\beta} \end{pmatrix},$$

tal que os elementos que compõem a matriz, depois de algumas manipulações algébricas, são dados, respectivamente, por:

(i) Modelo Weibull múltiplo

$$\begin{aligned} \ddot{\mathbf{L}}_{\omega\omega} &= \sum_{i:\delta_i=1} \left\{ \frac{g_{ij}t_i^{-1}\log(t_i)[2 + \omega_j\log(t_i)]h_i - \left\{g_{ij}t_i^{-1}[1 + \omega_j\log(t_i)]\right\}^2}{h_i^2} + \right. \\ &\quad \left. g_{ij}\log^2(t_i) \right\} + \sum_{i:\delta_i=0} g_{ij}\log^2(t_i), \\ \ddot{\mathbf{L}}_{\beta\omega} &= \sum_{i:\delta_i=1} \left\{ x_{rj}g_{ij}t_i^{-1}[1 + \omega_j\log(t_i)] \left[\frac{h_i - \omega_jg_{ij}t_i^{-1}}{h_i^2} \right] - x_{rj}g_{ij}\log(t_i) \right\} - \\ &\quad \sum_{i:\delta_i=0} x_{rj}g_{ij}\log(t_i), \\ \ddot{\mathbf{L}}_{\beta\beta} &= \sum_{i:\delta_i=1} \left\{ \frac{\omega_jg_{ij}t_i^{-1}x_{ir}x_{is}[h_i - \omega_jg_{ij}t_i^{-1}]}{h_i^2} - g_{ij}x_{ir}x_{is} \right\} - \\ &\quad \sum_{i:\delta_i=0} g_{ij}x_{ir}x_{is}. \end{aligned}$$

(ii) Modelo log-logístico múltiplo

$$\begin{aligned} \ddot{\mathbf{L}}_{\omega\omega} &= \sum_{i:\delta_i=1} \left\{ -q_i^{-2} \left[\frac{u_{ij}(1 + \omega_j\log(t_i) + g_{ij})}{(1 + g_{ij})^2} \right]^2 + \right. \\ &\quad \left. q_i^{-1} \left[\frac{u_{ij}\log(t_i)[2 + \omega_j\log(t_i) + 2g_{ij} - \omega_j\log(t_i)g_{ij}]}{(1 + g_{ij})^3} \right] - \frac{g_{ij}[\log(t_i)]^2}{(1 + g_{ij})^2} \right\} \\ &\quad - \sum_{i:\delta_i=0} \frac{g_{ij}[\log(t_i)]^2}{(1 + g_{ij})^2}, \end{aligned}$$

$$\begin{aligned}\ddot{\mathbf{L}}_{\beta\omega} = & \sum_{i:\delta_i=1} \left\{ -q_i^{-2} \left[\frac{\omega_j u_{ij} x_{ir}}{(1+g_{ij})^2} \right] \left[\frac{u_{ij}(1+\omega_j \log(t_i) + g_{ij})}{(1+g_{ij})^2} \right] \right. \\ & + q_i^{-1} \left[\frac{u_{ij} x_{ir} (1 + \omega_j \log(t_i) + g_{ij} - \omega_j g_{ij} \log(t_i))}{(1+g_{ij})^3} \right] - \\ & \left. \frac{g_{ij} x_{ir} \log(t_i)}{(1+g_{ij})^3} \right\} - \sum_{i:\delta_i=0} \frac{g_{ij} x_{ir} \log(t_i)}{(1+g_{ij})^3},\end{aligned}$$

$$\begin{aligned}\ddot{\mathbf{L}}_{\beta\beta} = & \sum_{i:\delta_i=1} \left\{ -q_i^{-2} \left[\frac{\omega_j u_{ij} x_{ir}}{(1+g_{ij})^2} \right]^2 + q_i^{-1} \left[\frac{\omega_j u_{ij} x_{ir} x_{is} (1 - g_{ij})}{(1+g_{ij})^3} \right] - \right. \\ & \left. \frac{g_{ij} x_{ir} x_{is}}{(1+g_{ij})^2} \right\} - \sum_{i:\delta_i=0} \frac{g_{ij} x_{ir} x_{is}}{(1+g_{ij})^2},\end{aligned}$$

em que, $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ é de ordem $[k + k(p+1)] \times [k + k(p+1)]$, e $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\beta})^T$.

Em geral, esta teoria tem sido aplicada considerando $k = 2$ motivos que podem causar a falha do indivíduo observado por tornar o modelo mais aplicável.

3.2.2 Influência local

A análise de influência local definida na seção (2.3) é desenvolvida para os modelos Weibull múltiplo e poly-log-logístico, através do cálculo da matriz de informação observada obtido na seção (3.2.1), dos autovalores e autovetores da matriz $F = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$, e obtendo-se as curvaturas para os três tipos de perturbações: ponderação de casos, perturbando um covariável e perturbando a variável resposta, respectivamente.

3.2.2.1 Perturbação de casos

Para esse esquema de perturbação é apresentado o logaritmo da função de verossimilhança perturbada e as curvaturas para os modelos Weibull múltiplo e log-logístico múltiplo. Seja o vetor de pesos $\mathbf{m} = (m_1, m_2, \dots, m_n)^T$, então, o logaritmo da função de verossimilhança do modelo perturbado é representado por:

$$l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) = \sum_{i=1}^n m_i l_i(\boldsymbol{\omega}, \boldsymbol{\beta}),$$

em que $l_i(t_i; \boldsymbol{\omega}, \boldsymbol{\beta})$ pode ser representada pelas equações (16) e (19), e o vetor de parâmetros, neste caso, é dado por $\boldsymbol{\theta} = (\boldsymbol{\omega}^T, \boldsymbol{\beta}^T)^T$.

(i) Modelo Weibull múltiplo

O logaritmo da função de verossimilhança perturbada para esse modelo é dado por:

$$\begin{aligned} l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) &= \sum_{i:\delta_i=1} m_i \log \left[\sum_{j=1}^k \omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] - \sum_{i:\delta_i=1} m_i \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] \\ &\quad - \sum_{i:\delta_i=0} m_i \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right]. \end{aligned}$$

Nesse caso, o vetor correspondente à não perturbação é o vetor $\mathbf{m}_0 = (1, \dots, 1)^T$, n -dimensional. Para a obtenção da direção de maior influência local, dada através da matriz $F = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$, é calculada a matriz de derivadas, $\boldsymbol{\Delta}$, na i -ésima linha representada por:

$$\boldsymbol{\Delta}_i^T = \left[\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_1}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_k}, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{0j}}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{pj}} \right].$$

Logo, os elementos da i -ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$, são expressos por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_j} = \begin{cases} \hat{g}_{ij} t_i^{-1} [1 + \hat{\omega}_j \log(t_i)] \hat{h}_i^{-1} - \hat{g}_{ij} \log(t_i), & \forall \quad i : \delta_i = 1, \\ -\hat{g}_{ij} \log(t_i), & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} x_{ir} \hat{\omega}_j \hat{g}_{ij} t_i^{-1} \hat{h}_i^{-1} + x_{ir} \hat{g}_{ij}, & \forall \quad i : \delta_i = 1, \\ -x_{ir} \hat{g}_{ij}, & \forall \quad i : \delta_i = 0. \end{cases}$$

(ii) Modelo log-logístico múltiplo

O logaritmo da função de verossimilhança perturbada para esse modelo é dado por:

$$\begin{aligned}
l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) &= \sum_{i:\delta_i=1} m_i \log \left[\sum_{j=1}^k \frac{\omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}}{1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}} \right] \\
&\quad - \sum_{i:\delta_i=1} m_i \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right] \\
&\quad - \sum_{i:\delta_i=0} m_i \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right].
\end{aligned}$$

Com o vetor n -dimensional $\mathbf{m}_0 = (1, \dots, 1)^T$ correspondente à não perturbação, o cálculo da matriz $F = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$, é dado através da matriz de informação observada e dos elementos da i -ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$, que são expressos por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_j} = \begin{cases} \hat{q}_i^{-1} \left[\frac{\hat{u}_{ij}(1 + \hat{\omega}_j \log(t_i) + \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right] - \left[\frac{\hat{g}_{ij} \log(t_i)}{(1 + \hat{g}_{ij})} \right], & \forall \quad i : \delta_i = 1, \\ - \left[\frac{\hat{g}_{ij} \log(t_i)}{(1 + \hat{g}_{ij})} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} \hat{q}_i^{-1} \left[\frac{\hat{\omega}_j \hat{u}_{ij} x_{ir}}{(1 + \hat{g}_{ij})^2} \right] - \left[\frac{\hat{g}_{ij} x_{ir}}{(1 + \hat{g}_{ij})} \right], & \forall \quad i : \delta_i = 1, \\ - \left[\frac{\hat{g}_{ij} x_{ir}}{(1 + \hat{g}_{ij})} \right], & \forall \quad i : \delta_i = 0. \end{cases}$$

3.2.2.2 Perturbação de uma covariável

Thomas e Cook (1990) sugerem modificar a r -ésima coluna da matriz $\mathbf{X}$, adicionando o vetor $\mathbf{m}$ de pequenas perturbações multiplicadas por um fator de escala S_x , que pode ser, por exemplo, a estimativa do desvio padrão amostral de $\mathbf{x}_r$. Então, a variável explanatória perturbada fica:

$$x_{irm} = x_{ir} + S_x m_i, \quad i = 1, \dots, n \quad \text{e} \quad r = 0, \dots, p.$$

Para os modelos de riscos múltiplos estudados são apresentados o logaritmo da função de verossimilhança perturbada e as curvaturas, correspondentes a esse esquema de perturbação.

(i) Modelo Weibull múltiplo

Para esse modelo o logaritmo da função de verossimilhança perturbada é dado por:

$$l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) = \sum_{i:\delta_i=1} \log \left[\sum_{j=1}^k \omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right] - \sum_{i:\delta_i=1} \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right] \\ - \sum_{i:\delta_i=0} \left[\sum_{j=1}^k \omega_j t_i^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right].$$

O vetor correspondente à não perturbação é o vetor n -dimensional $\mathbf{m}_0 = (0, \dots, 0)^T$. Para a obtenção da direção de maior influência local, dada através da matriz $F = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$, utiliza a matriz de informação observada e é calculada a matriz de derivadas, $\boldsymbol{\Delta}$, na i -ésima linha representada por:

$$\boldsymbol{\Delta}_i^T = \left[\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_1}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_k}, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{0j}}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{pj}} \right].$$

Nesse caso, os elementos da i -ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$, são representados por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_{ij}} = \begin{cases} \hat{g}_{ij} t_i^{-1} S_x \left[1 + \hat{\omega}_j \log(t_i) \right] \hat{h}_i^{-1} \left[\hat{\beta}_{rj} - \hat{h}_i^{-1} \sum_{i=1}^k \hat{\beta}_{rj} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j \right] - \\ \beta_{rj} \hat{g}_{ij} \log(t_i) S_x, & \forall \quad i : \delta_i = 1, \\ -\hat{\beta}_{rj} \hat{g}_{ij} \log(t_i) S_x, & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} x_{ir} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j S_x \hat{h}_i^{-1} \left[\hat{\beta}_{rj} - \hat{h}_i^{-1} \sum_{j=1}^k \hat{\beta}_{rj} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j \right] - \\ x_{ir} \hat{\beta}_{rj} \hat{g}_{ij} S_x, & \forall \quad i : \delta_i = 1, \\ -x_{ir} \hat{\beta}_{rj} \hat{g}_{ij} S_x, & \forall \quad i : \delta_i = 0, \end{cases}$$

em que $r \neq t$,

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{tj}} = \begin{cases} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j S_x \hat{h}_i^{-1} \left[(x_{it} \hat{\beta}_{tj} + 1) - x_{it} \hat{h}_i^{-1} \sum_{j=1}^k \hat{\beta}_{tj} \hat{g}_{ij} \hat{\omega}_j \right] - \\ \hat{g}_{ij} S_x (x_{it} \hat{\beta}_{tj} + 1), & \forall \quad i : \delta_i = 1, \\ -\hat{g}_{ij} S_x (x_{it} \hat{\beta}_{tj} + 1), & \forall \quad i : \delta_i = 0. \end{cases}$$

(ii) Modelo log-logístico múltiplo

Para esse modelo o logaritmo da função de verossimilhança perturbada é dado por:

$$\begin{aligned}
 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) &= \sum_{i: \delta_i=1} \log \left[\sum_{j=1}^k \frac{\omega_j t_i^{\omega_j-1} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\}}{1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\}} \right] \\
 &\quad - \sum_{i: \delta_i=1} \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right) \right] \\
 &\quad - \sum_{i: \delta_i=0} \left[\sum_{j=1}^k \log \left(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right) \right].
 \end{aligned}$$

Os elementos da i -ésima linha da matriz Δ , avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$, em que o vetor de não perturbação $\mathbf{m}_0 = (0, \dots, 0)^T$, são representados por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_j} = \begin{cases} -\hat{q}_i^{-2} \left[\sum_{j=1}^k \frac{\hat{\omega}_j \hat{u}_{ij} S_x \hat{\beta}_{tj}}{(1+\hat{g}_{ij})^2} \right] \left[\frac{\hat{u}_{ij} (1+\hat{\omega}_j \log(t_i) + \hat{g}_{ij})}{(1+\hat{g}_{ij})^2} \right] + \\ \hat{q}_i^{-1} \left[\frac{\hat{u}_{ij} S_x \hat{\beta}_{tj} (1+\hat{\omega}_j \log(t_i) + \hat{g}_{ij} - \hat{\omega}_j \log(t_i) \hat{g}_{ij})}{(1+\hat{g}_{ij})^3} \right] - \left[\frac{\hat{g}_{ij} S_x \hat{\beta}_{tj} \log(t_i)}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 1, \\ -\left[\frac{\hat{g}_{ij} S_x \hat{\beta}_{tj} \log(t_i)}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} -\hat{q}_i^{-2} \left[\sum_{j=1}^k \frac{\hat{\omega}_j \hat{u}_{ij} S_x \hat{\beta}_{tj}}{(1+\hat{g}_{ij})^2} \right] \left[\frac{\hat{\omega}_j \hat{u}_{ij} x_{ir}}{(1+\hat{g}_{ij})^2} \right] + \hat{q}_i^{-1} \left[\frac{\hat{\omega}_j \hat{u}_{ij} S_x \hat{\beta}_{tj} x_{ir} (1-\hat{g}_{ij})}{(1+\hat{g}_{ij})^3} \right] - \\ \left[\frac{\hat{g}_{ij} S_x \hat{\beta}_{tj} x_{ir}}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 1, \\ -\left[\frac{\hat{g}_{ij} S_x \hat{\beta}_{tj} x_{ir}}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

em que $r \neq t$,

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{tj}} = \begin{cases} -\hat{q}_i^{-2} \left[\sum_{j=1}^k \frac{\hat{\omega}_j \hat{u}_{ij} S_x \hat{\beta}_{tj}}{(1+\hat{g}_{ij})^2} \right] \left[\frac{\hat{\omega}_j \hat{u}_{ij} (x_{it} + S_t \hat{m}_i)}{(1+\hat{g}_{ij})^2} \right] + \\ \hat{q}_i^{-1} \left[\frac{\hat{\omega}_j \hat{u}_{ij} S_x \left\{ \hat{\beta}_{tj} (x_{it} + S_x \hat{m}_i) + 1 + \hat{g}_{ij} - \hat{g}_{ij} (x_{it} + S_x \hat{m}_i) \right\}}{(1+\hat{g}_{ij})^3} \right] - \\ \left[\frac{\hat{g}_{ij} S_x \left\{ \hat{\beta}_{tj} (x_{it} + S_x \hat{m}_i) + 1 + \hat{g}_{ij} \right\}}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 1, \\ -\left[\frac{\hat{g}_{ij} S_x \left\{ \hat{\beta}_{tj} (x_{it} + S_x \hat{m}_i) + 1 + \hat{g}_{ij} \right\}}{(1+\hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

sendo que,

$$\mathbf{x}_i^{*T} \boldsymbol{\beta}_j = \beta_{0j} + x_{i1} \beta_{1j} + x_{i2} \beta_{2j} + \dots + (x_{ir} + S_x m_i) \beta_{rj} + \dots + x_{ip} \beta_{pj}.$$

3.2.2.3 Perturbação da variável resposta

Para perturbar a variável resposta t , reescreve-se adicionando o vetor $\mathbf{m}$ de pequenas perturbações multiplicadas por um fator de escala S_t , que pode ser, por exemplo, a estimativa do desvio padrão da variável T . Assim, a variável resposta perturbada fica:

$$t_{im} = t_i + S_t m_i, \quad i = 1, \dots, n.$$

Adotando os modelos em estudo são apresentados o logaritmo da função de verossimilhança perturbada e as curvaturas, correspondentes a esse esquema de perturbação.

(i) Modelo Weibull múltiplo

Para esse modelo o logaritmo da função de verossimilhança perturbada é dado por:

$$\begin{aligned} l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) &= \sum_{i:\delta_i=1} \log \left[\sum_{j=1}^k \omega_j t_{im}^{\omega_j-1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] - \sum_{i:\delta_i=1} \left[\sum_{j=1}^k \omega_j t_{im}^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right] \\ &\quad - \sum_{i:\delta_i=0} \left[\sum_{j=1}^k \omega_j t_{im}^{\omega_j} \exp\{\mathbf{x}_i^{*T} \boldsymbol{\beta}_j\} \right]. \end{aligned}$$

Nesse caso, o vetor correspondente à não perturbação é $\mathbf{m}_0 = (0, \dots, 0)^T$. E a obtenção da direção de maior influência local, dada através da matriz $F = \boldsymbol{\Delta}^T \ddot{L}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}$, utiliza a matriz de informação observada e é calculada a matriz de derivadas, $\boldsymbol{\Delta}$, na i -ésima linha representada por:

$$\boldsymbol{\Delta}_i^T = \left[\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_1}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_k}, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{0j}}, \dots, \frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{pj}} \right].$$

Logo, os elementos da i -ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$,

em que $\mathbf{m}_0 = (0, \dots, 0)^T$ são representados por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_j} = \begin{cases} \hat{g}_{ij} t_i^{-2} \left\{ (\hat{\omega}_j - 1) S_t + \hat{\omega}_j \left[(\hat{\omega}_j - 1) \log(t_i) + S_t \right] \right\} \hat{h}_i^{-1} - \\ \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_j\} \left[t_i + \hat{\omega}_j t_i^{\hat{\omega}_j - 1} \log(t_i) \right] * \\ S_t \hat{h}_i^{-2} \left[\sum_{i=1}^k \hat{\omega}_j (\hat{\omega}_j - 1) \hat{g}_{ij} t_i^{-2} \right] - \\ \hat{g}_{ij} t_i^{-1} S_t \left[\hat{\omega}_j \log(t_i) + 1 \right], & \forall \quad i : \delta_i = 1, \\ -\hat{g}_{ij} t_i^{-1} S_t \left[\omega_j \log(t_i) + 1 \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} x_{ir} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j S_t \hat{h}_i^{-1} \left[t_i^{-1} (\hat{\omega}_j - 1) - \right. \\ \left. \hat{h}_i^{-1} \sum_{j=1}^k \hat{g}_{ij} t_i^{-2} \hat{\omega}_j (\hat{\omega}_j - 1) \right] - \\ x_{ir} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j S_t, & \forall \quad i : \delta_i = 1, \\ -x_{ir} \hat{g}_{ij} t_i^{-1} \hat{\omega}_j S_t, & \forall \quad i : \delta_i = 0. \end{cases}$$

(ii) Modelo log-logístico múltiplo

Para esse modelo o logaritmo da função de verossimilhança perturbada é dado por:

$$\begin{aligned} l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m}) &= \sum_{i:\delta_i=1} \log \left[\sum_{j=1}^k \frac{\omega_j t_{im}^{\omega_j - 1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}}{1 + t_{im}^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}} \right] \\ &\quad - \sum_{i:\delta_i=1} \left[\sum_{j=1}^k \log \left(1 + t_{im}^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right] \\ &\quad - \sum_{i:\delta_i=0} \left[\sum_{j=1}^k \log \left(1 + t_{im}^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right]. \end{aligned}$$

Assim, os elementos da i -ésima linha da matriz $\boldsymbol{\Delta}$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\mathbf{m} = \mathbf{m}_0$, em que $\mathbf{m}_0 = (0, \dots, 0)^T$ são representados por:

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \omega_j} = \begin{cases} -\hat{q}_i^{-2} \left[\sum_{j=1}^k \frac{\hat{\omega}_j \hat{u}_{ij} t_i^{-1} S_t (\hat{\omega}_j - 1 - \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right] \left[\frac{\hat{u}_{ij} (1 + \hat{\omega}_j \log(t_i) + \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right] + \\ \hat{q}_i^{-1} \left[\frac{\hat{u}_{ij} t_i^{-1} S_t (2\hat{\omega}_j - 1 + \hat{\omega}_j^2 \log(t_i) + \hat{\omega}_j \log(t_i) + 2\hat{\omega}_j \hat{g}_{ij})}{(1 + \hat{g}_{ij})^3} + \right. \\ \left. \frac{\hat{u}_{ij} t_i^{-1} S_t (-2\hat{g}_{ij} + \hat{\omega}_j \hat{g}_{ij} \log(t_i) + \hat{g}_{ij}^2 - \hat{\omega}_j^2 \hat{g}_{ij} \log(t_i))}{(1 + \hat{g}_{ij})^3} \right] \\ \left[\frac{\hat{g}_{ij} t_i^{-1} S_t (\hat{\omega}_j \log(t_i) + 1 + \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 1, \\ -\left[\frac{\hat{g}_{ij} t_i^{-1} S_t (\hat{\omega}_j \log(t_i) + 1 + \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

$$\frac{\partial^2 l(t_i; \boldsymbol{\omega}, \boldsymbol{\beta} | \mathbf{m})}{\partial m_i \partial \beta_{rj}} = \begin{cases} -\hat{q}_i^{-2} \left[\sum_{j=1}^k \frac{\hat{\omega}_j \hat{u}_{ij} t_i^{-1} S_t(\hat{\omega}_j - 1 - \hat{g}_{ij})}{(1 + \hat{g}_{ij})^2} \right] \left[\frac{\hat{\omega}_j \hat{u}_{ij} x_{ir}}{(1 + \hat{g}_{ij})^2} \right] + \\ \hat{q}_i^{-1} \left[\frac{\hat{\omega}_j \hat{u}_{ij} t_i^{-1} S_t x_{ir} (\hat{\omega}_j - 1 - \hat{g}_{ij} - \hat{\omega}_j \hat{g}_{ij})}{(1 + \hat{g}_{ij})^3} \right] - \left[\frac{\hat{\omega}_j \hat{g}_{ij} t_i^{-1} S_t x_{ir}}{(1 + \hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 1, \\ - \left[\frac{\hat{\omega}_j \hat{g}_{ij} t_i^{-1} S_t x_{ir}}{(1 + \hat{g}_{ij})^2} \right], & \forall \quad i : \delta_i = 0, \end{cases}$$

sendo que,

$$\begin{aligned} \hat{u}_{ij} &= t_i^{\hat{\omega}_j - 1} \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_j\}, \quad \hat{g}_{ij} = t_i^{\hat{\omega}_j} \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_j\}, \\ \hat{h}_i &= \sum_{j=1}^k \omega_j t_i^{\omega_j - 1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}, \quad \hat{q}_i = \sum_{j=1}^k \frac{\hat{\omega}_j t_i^{\hat{\omega}_j - 1} \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_j\}}{1 + t_i^{\hat{\omega}_j} \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_j\}}, \end{aligned}$$

$$i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k, \quad \text{e} \quad r = 0, 1, \dots, p.$$

Outras aplicações, resultados e discussões sobre a metodologia de influência local podem ser encontrados em Souza (1999), Paula (2004), Ortega (2006), Souza (2006), entre outros.

3.2.3 Análise de resíduos

Ao ajustar um modelo a um conjunto de dados, é importante avaliar a qualidade do modelo ajustado. Com esse objetivo são desenvolvidas as técnicas dos resíduos martingal e deviance definidos na seção (2.4) para os modelos Weibull múltiplo e log-logístico múltiplo.

3.2.3.1 Resíduo martingal

Para a obtenção do resíduo martingal para os modelos de riscos múltiplos em estudo, utiliza-se a equação (5), obtendo a função risco acumulada referente aos modelos Weibull múltiplo e log-logístico múltiplo. E através da equação (24), tem-se o resíduo martingal para:

(i) Modelo Weibull múltiplo

$$r_{M_i} = \begin{cases} 1 - \sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}, & \forall \quad i : \delta_i = 1, \\ - \sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}, & \forall \quad i : \delta_i = 0, \end{cases}$$

em que δ_i é a variável indicadora da falha e $\sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}$ é a função de risco acumulada do modelo Weibull múltiplo.

(ii) Modelo log-logístico múltiplo

$$r_{M_i} = \begin{cases} 1 - \sum_{j=1}^k \log\left[1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}\right], & \forall \quad i : \delta_i = 1, \\ - \sum_{j=1}^k \log\left[1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}\right], & \forall \quad i : \delta_i = 0. \end{cases}$$

em que δ_i é a variável indicadora da falha e $\sum_{j=1}^k \log\left[1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}\right]$ é a função de risco acumulada do modelo log-logístico múltiplo.

3.2.3.2 Resíduo deviance

O resíduo deviance para os modelos de riscos múltiplos em estudo é obtido substituindo o resíduo de martingal obtido na seção (3.2.3.1) na equação (25) para cada caso considerado, respectivamente, ou seja,

(i) Modelo Weibull múltiplo

$$r_{D_i} = \begin{cases} \sin(\alpha(r_{M_i})) \left[-2 \left\{ 1 - \sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} + \log\left(\sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right) \right\} \right]^{1/2}, & \forall \quad i : \delta_i = 1, \\ \sin(\alpha(r_{M_i})) \left[2 \sum_{j=1}^k t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\} \right]^{1/2}, & \forall \quad i : \delta_i = 0, \end{cases}$$

em que δ_i é a variável indicadora da falha, $\hat{r}_{M_i}$ é o resíduo martingal do modelo Weibull múltiplo e $\sin(\alpha(r_{M_i}))$ é o sinal de r_{M_i} .

(ii) Modelo log-logístico múltiplo

$$r_{D_i} = \begin{cases} \begin{aligned} & \text{sinal}(r_{M_i}) \left[-2 \left\{ 1 - \sum_{j=1}^k \log(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}) + \right. \right. \\ & \left. \left. \log \left[\sum_{j=1}^k \log(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}) \right] \right\} \right]^{1/2} \end{aligned} & \forall \quad i : \delta_i = 1 \\ \text{sinal}(r_{M_i}) \left[2 \sum_{j=1}^k \log(1 + t_i^{\omega_j} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_j\}) \right]^{1/2} & \forall \quad i : \delta_i = 0 \end{cases}$$

sendo que δ_i é a variável indicadora da falha, $\hat{r}_{M_i}$ é o resíduo martingal do modelo log-logístico múltiplo, $\text{sinal}(r_{M_i})$ é o sinal de r_{M_i} , $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, k$.

Outros estudos envolvendo análise de resíduos podem ser encontrados, por exemplo, McCullagh e Nelder (1989), Therneau e Grambsch (1990), Ortega; Bolfarine e Paula (2003), entre outros.

4 RESULTADOS e DISCUSSÃO

Aos dados apresentados na seção (3.1) são ajustados os modelos discutidos na seção (2.2), considerando $k = 2$, devido a tornar o modelo mais aplicável, além de facilitar a obtenção dos cálculos.

4.1 Modelo de riscos múltiplos

Os modelos de riscos múltiplos, para $k = 2$, são conhecidos como modelo bi-Weibull e modelo bi-log-logístico. Então, ao utilizar o procedimento NLMIXED do *software* SAS e o procedimento MAXBFGS do *software* OX, as estimativas de máxima verossimilhança, respectivamente, são dadas por:

(i) Modelo bi-Weibull

Esse modelo é representado por:

$$h(t) = \sum_{j=1}^2 \omega_j t^{\omega_j - 1} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j},$$

em que t indica o tempo de sobrevivência dos peixes, e o vetor de parâmetros é dado por $\boldsymbol{\theta} = (\omega_1, \omega_2, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$.

Conseqüentemente, na Tabela 1, são apresentadas as estimativas de máxima verossimilhança, erro padrão, a estatística z e o intervalo de confiança.

Tabela 1 - Estimativas dos parâmetros para os dados de peixes

Parâmetro	Estimativa	Erro padrão	Estatística z	p-valor	95% C.L.
ω_1	0.5674	0.08025	7.07	<0.0001*	(0.4083, 0.7265)
ω_2	0.5813	0.1942	2.99	0.0034*	(0.1963, 0.9663)
β_{01}	-5.3867	3.0755	-1.75	0.0828	(-11.4842 , 0.7108)
β_{11}	0.07746	0.03972	1.95	0.0538*	(-0.0013 , 0.1562)
β_{21}	-0.01166	0.008155	-1.43	0.1556	(-0.0278 , 0.0045)
β_{31}	-0.7061	0.4322	-1.63	0.1053	(-1.5630 , 0.1508)
β_{02}	5.8282	5.9756	0.98	0.3316	(-6.0190 , 17.6754)
β_{12}	-0.1137	0.1068	-1.07	0.2893	(-0.3255 , 0.0980)
β_{22}	-0.01213	0.01837	-0.66	0.5104	(-0.0485 , 0.0243)
β_{32}	0.9790	0.7425	1.32	0.1901	(-0.4930 , 2.4510)
Estatística	Valor	Estatística	Valor		
AIC	475.7	BIC	502.3		

* significativo à 5%

Ao analisar a Tabela 1, verifica-se que as estimativas dos parâmetros de risco, ω_1 e ω_2 , e a covariável tamanho do peixe para o risco 1 são significativas ao nível de 5% de probabilidade. Mas, observa-se que as estimativas relacionadas aos riscos são aproximadamente iguais, portanto, esse fato indica que o modelo bi-Weibull se reduz ao modelo Weibull simples, conseqüentemente, não é um bom ajuste para os dados de peixes.

Outro método descrito na seção (2.2) para verificar a distribuição adequada, é através da construção do gráfico TTT *plot*, o qual indica a forma que a função taxa de falha de um determinado modelo pode assumir.

Logo, para os dados de peixes, o gráfico TTT *plot* encontra-se na Figura 4. Observa-se uma forma convexa e depois um segmento, o qual indica que o modelo Weibull simples não seria adequado para ajustar os dados. Então, da Tabela 1 e Figura 4, tem-se que o modelo Weibull não seria uma distribuição adequada para modelar os dados de peixes, uma sugestão seria o ajuste do modelo bi-log-logístico.

Figura 4 - Grafico TTT *plot* para tempo de sobrevivência de peixes

(ii) Modelo bi-log-logístico

Esse modelo é expresso por:

$$h(t) = \sum_{j=1}^2 \frac{\omega_j t^{\omega_j-1} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}{1 + t^{\omega_j} e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}},$$

sendo que t indica o tempo de sobrevivência dos peixes, e o vetor de parâmetros é dado por $\boldsymbol{\theta} = (\omega_1, \omega_2, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$.

Na Tabela 2, são apresentadas as estimativas de máxima verossimilhança, erro padrão, a estatística z e o intervalo de confiança. As estimativas dos parâmetros de risco, a covariável profundidade do rio relacionada ao risco 1 e o intercepto para o risco 2, são significativas ao nível de 5% de probabilidade. E verifica-se que as suposições do modelo não são violadas, ou seja, as estimativas para os parâmetros de risco são positivas, sendo essas significativamente diferentes, e $\hat{\omega}_1 < \hat{\omega}_2$. Portanto, o modelo bi-log-logístico ajusta os dados de peixes.

Tabela 2 - Estimativas dos parâmetros para os dados de peixes

Parâmetro	Estimativa	Erro padrão	Estatística z	p-valor	95% I.C.
ω_1	0.7463	0.06923	10.78	<0.0001*	(0.6091 , 0.8836)
ω_2	8.7760	3.1071	2.82	0.0057*	(2.6160 , 14.9361)
β_{01}	-0.8223	1.5835	-0.52	0.6046	(-3.9618 , 2.3171)
β_{11}	0.03581	0.02230	1.61	0.1113	(-0.0084 , 0.0800)
β_{21}	-0.02160	0.008188	-2.64	0.0096*	(-0.0378 , -0.0054)
β_{31}	-0.2598	0.1764	-1.47	0.1438	(-0.6096 , 0.0899)
β_{02}	-28.7130	11.1716	-2.57	0.0116*	(-50.8618 , -6.5642)
β_{12}	-0.1945	0.2808	-0.69	0.4899	(-0.7511 , 0.3621)
β_{22}	0.07480	0.06590	1.14	0.2589	(-0.0559 , 0.2055)
β_{32}	-3.3458	2.7237	-1.23	0.2220	(-8.7459 , 2.0542)
Estatística	Valor	Estatística	Valor		
AIC	463.5	BIC	490.1		

* significativo à 5%

4.2 Influencia local

Ao considerar o modelo bi-log-logístico, a análise de influência local considerando os três esquemas de perturbações apresentados na seção (3.2.2) é realizada a seguir.

4.2.1 Perturbação de casos

Através desse esquema de perturbação para o modelo bi-log-logístico, tem-se que:

$$C_{\mathbf{d}_{max}}(\theta) = 15.221,$$

é o maior autovalor obtido. Então, na Figura 5, é apresentado o gráfico do autovetor correspondente a $C_{\mathbf{d}_{max}}(\theta)$, e na Figura 6, a influência local total do i -ésimo indivíduo. Ambos os gráficos em relação ao vetor de parâmetros θ .

Figura 5 - Gráfico de influência - ponderação de casos

Figura 6 - Gráfico de influência do i -ésimo indivíduo - ponderação de casos

Portanto, das Figuras 5 e 6, verifica-se que os indivíduos 5, 17, 22 e 30 são os que mais se destacam. Então, ao analisar esses indivíduos observa-se que o indivíduo 5 possui o maior tempo de sobrevivência ao compará-lo com os demais, e para os outros não existe característica que explique esse destaque.

4.2.2 Perturbação de uma covariável

Para o modelo bi-log-logístico ajustado aos dados de peixes, perturbam-se as três covariáveis: tamanho do peixe, profundidade do rio e transparência da água, respectivamente.

Ao perturbar a covariável tamanho do peixe obtém-se que,

$$C_{\mathbf{d}_{max}}(\theta) = 0.63471,$$

é o maior autovalor calculado. Logo, na Figura 7, tem-se o gráfico desse autovetor, e na Figura 8, a influência local total do i -ésimo indivíduo, considerando o vetor de parâmetros θ para os dois gráficos.

Para o modelo bi-log-logístico tem-se que,

$$C_{\mathbf{d}_{max}}(\theta) = 0.0074227,$$

o qual é o maior autovalor obtido. Portanto, na Figura 9, tem-se o gráfico desse autovetor, e na Figura 10, a influência local total do i -ésimo indivíduo, considerando o vetor de parâmetros θ para os dois gráficos.

E ao perturbar a covariável transparência da água obtém-se,

$$C_{\mathbf{d}_{max}}(\theta) = 28.566,$$

sendo esse o maior autovalor obtido. Então, na Figura 11, tem-se o gráfico desse autovetor, e na Figura 12, a influência local total do i -ésimo indivíduo, considerando o vetor de parâmetros θ para ambos os gráficos.

Nas Figuras 7, 8 e 9, observa-se que se destacam os indivíduos: 5, 17, 22 e 105, e nas Figuras 10, 11 e 12, os indivíduos 17, 22 e 105. Logo, ao verificá-los tem-se que o indivíduo 5 possui o maior tempo de sobrevivência ao compará-lo com os demais, e para os outros não existe característica que explique esse destaque.

Através desse esquema de perturbação, conclui-se que apenas o indivíduo 5 possui uma explicação que o torna um possível ponto influente.

Figura 7 - Gráfico de influência - perturbação da covariável tamanho do peixe

Figura 8 - Gráfico de influência do i -ésimo indivíduo da covariável tamanho do peixe

Figura 9 - Gráfico de influência - perturbação da covariável profundidade do rio

Figura 10 - Gráfico de influência do i -ésimo indivíduo da covariável profundidade do rio

Figura 11 - Gráfico de influência - perturbação da covariável transparência da água

Figura 12 - Gráfico de influência do i -ésimo indivíduo da covariável transparência da água

4.2.3 Perturbação da variável resposta

Ao considerar esse esquema de perturbação para o modelo bi-log-logístico, tem-se que:

$$C_{\mathbf{d}_{max}}(\theta) = 172.39,$$

é o maior autovalor obtido. Portanto, na Figura 13, tem-se o gráfico do autovetor correspondente ao $C_{\mathbf{d}_{max}}(\theta)$, e na Figura 14, a influência local total do i -ésimo indivíduo, ambos os gráficos considerando o vetor de parâmetros θ .

Nas Figuras 13 e 14, verifica-se que os indivíduos 9, 34 e 94 são os mais influentes para esse caso. Então, ao analisar esses indivíduos observa-se que os indivíduos 34 e 94 possuem o menor tempo de sobrevivência e o indivíduo 9 possui o segundo menor tempo de sobrevivência de todos os indivíduos considerados na análise.

Portanto, de todos os indivíduos considerados possíveis pontos influentes, os que de fato constam uma justificativa ao analisar os dados originais são: 5, 9, 34, e 94.

Figura 13 - Gráfico de influência - perturbação da variável resposta

Figura 14 - Gráfico de influência do i -ésimo indivíduo da variável resposta

4.3 Análise de resíduo

Ao modelo bi-log-logístico é aplicada a teoria de resíduo apresentada na seção (3.2.3) para verificar a presença de pontos infuents, bem como possíveis afastamentos das suposições do modelo. Conseqüentemente, as Figuras 15 e 16, apresentam os gráficos dos resíduos martingal e deviance, respectivamente.

Figura 15 - Gráfico do Resíduo Martingal

Figura 16 - Gráfico do Resíduo Deviance

Da Figura 15, verifica-se que todos os indivíduos estão no intervalo $(-1.5, 1.5)$, logo as suposições do modelo não foram violadas. E da Figura 16, observa-se que os indivíduos 34, 94 e 100 estão acima de 3, e o indivíduo 5 está abaixo de -3. Assim, esses indivíduos são possíveis observações influentes no modelo.

4.4 Impacto das observações influentes detectadas

A teoria apresentada na seção (2.5), é desenvolvida considerando o modelo bi-log-logístico para revelar o impacto das observações influentes detectadas.

Então, para esse modelo tem-se que: $TRC = 220.35$, $MRC = 145.5$ e $LD_I = -26.7$. Para comparar o impacto das observações influentes, repete a análise removendo quatro observações aleatórias selecionadas das observações não-influente. Nesse caso, tem-se que: $TRC = 0.3715$, $MRC = 0.11$ e $LD_I = -7.3$. Assim observa-se uma mudança significativa entre as medidas TRC , MRC e LD_I indicando que as observações: 5, 9, 34 e 94, são influentes para o modelo proposto.

4.5 Reanálise dos dados

A reanálise dos dados é baseada nos resultados obtidos na análise de influência local e na análise dos resíduos, então as seguintes observações influentes: 5, 9, 34 e 94, são retiradas e através do novo conjunto de dados são recalculadas as estimativas de máxima verossimilhança, erro padrão, a estatística z e o intervalo de confiança, que se encontram na Tabela 3.

Tabela 3 - Estimativas dos parâmetros para os dados de peixes

Parâmetro	Estimativa	Erro padrão	Estatística z	p-valor	95% I.C.
ω_1	0.7944	0.07397	10.74	<0.0001*	(0.6477 , 0.9412)
ω_2	35.6855	10.2696	3.47	0.0008*	(15.3158 , 56.0552)
β_{01}	-0.7589	1.6146	-0.47	0.6394	(-3.9614 , 2.4437)
β_{11}	0.03844	0.02267	1.70	0.0930	(-0.0065 , 0.0834)
β_{21}	-0.02493	0.008499	-2.93	0.0041*	(-0.0418 , -0.0081)
β_{31}	-0.2624	0.1802	-1.46	0.1485	(-0.6198 , 0.0951)
β_{02}	-120.47	35.1753	-3.42	0.0009*	(-190.24 , -50.7027)
β_{12}	-0.01876	0.09577	-0.20	0.8451	(-0.2087 , 0.1712)
β_{22}	0.08719	0.03924	2.22	0.0285*	(0.0094 , 0.1650)
β_{32}	-4.4529	1.3677	-3.26	0.0015*	(-7.1657 , -1.7401)
Estatística	Valor	Estatística	Valor		
AIC	436.8	BIC	463.0		

* significativo à 5%

Da Tabela 3, as estimativas dos parâmetros de risco, as covariáveis profundidade do rio e transparência da água relacionadas ao risco 2, a covariável profundidade do rio relacionada ao risco 1 e o intercepto para o risco 2, são significativas ao nível de 5% de probabilidade. Então, comparando os resultados obtidos na Tabela 2 com o da Tabela 3, verifica-se o ganho no ajuste do modelo bi-log-logístico ao retirar as observações influentes.

4.6 Qualidade de ajuste

Através da função de sobrevivência de Kaplan-Meier, da função de sobrevivência estimada do modelo bi-log-logístico completo e da função de sobrevivência estimada do modelo bi-log-logístico sem as observações influentes constroem-se as Figuras 17 e 18.

As Figuras 17 e 18, auxiliam na verificação da melhora do ajuste do modelo bi-log-logístico ao retirar as observações influentes do conjunto de dados de peixes. Pode-se observar como a curva do modelo bi-log-logístico sem observações influentes acompanha melhor a curva de Kaplan-Meier.

Figura 17 - Gráfico de sobrevivência completa

Figura 18 - Gráfico de sobrevivência sem indivíduos influentes

Com base nos resultados apresentados observar-se que o modelo bi-log-logístico

melhor se ajustou aos dados. No entanto, quando aplicou-se a metodologia de influência local e de resíduos nesse modelo, verificou-se a presença de pontos influentes, tal fato é comprovado quando utilizou-se as medidas TRC , MDC e $LD_I(\boldsymbol{\theta})$. Sendo assim, decidiu-se realizar um novo ajuste do modelo sem considerar os possíveis pontos influentes, logo, observou-se que o ganho foi significativo, pois as covariáveis, profundidade do rio e transparência da água, em relação ao risco 2 passaram a ser significativas. Finalmente, o modelo proposto será o bi-log-logístico sem considerar os pontos influentes.

5 CONSIDERAÇÕES FINAIS

O objetivo inicial deste trabalho, visava um estudo de diagnóstico e resíduos nos modelos de riscos múltiplos. Foi apresentado de forma resumida a parte inferencial dos modelos de riscos múltiplos. Para estimar os parâmetros de máxima verossimilhança utilizou-se os procedimentos NLMIXED e MAXBFGS dos *software* SAS e OX, respectivamente.

Com a finalidade de facilitar os cálculos, para realizar a análise de influência local, foram obtidas todas as curvaturas em forma fechada para os três tipos de perturbação adotados no presente trabalho, assim como a matriz de informação observada. E encontrou-se os resíduos para os modelos de riscos múltiplos baseados nos resíduos martingal e deviance.

Os resultados obtidos na aplicação a um conjunto de dados reais de peixes da espécie “Notropis Dourado, crysoleucas de Notemigonus”, foram coerentes com a teoria desenvolvida nesse trabalho.

5.1 PERSPECTIVAS PARA TRABALHOS FUTUROS

Algumas investigações que se acredita ser de interesse para ampliar e dar continuidade aos resultados obtidos são:

1. O desenvolvimento e implementação do modelo de riscos múltiplos em observações, ou indivíduos, de longa duração ou fração de cura, segundo o enfoque de Maller e Zhar (1996), os quais propuseram a construção de uma função de sobrevivência populacional imprópria na forma de mistura dada por:

$$G(t) = p + (1 - p)S(t),$$

em que: $S(t)$ é a função de sobrevivência própria e p indica a proporção de curados.

No caso dos modelos de riscos múltiplos será introduzido uma função de sobrevivência imprópria para cada risco. Assim, a função de sobrevivência global imprópria será dada

por:

$$G(t) = p + (1 - p) \prod_{j=1}^k S_j(t).$$

2. Pode-se utilizar as funções de ligação logito, probito e complemento log-log para saber quais covariáveis influenciam positivo ou negativamente na proporção de curados.
3. O desenvolvimento e implementação computacional de outros métodos de estimação, tais como, estimador Baysiano e o método de Jackknife.
4. Outra possibilidade seria estimar a distribuição dos resíduos propostos na estimação de Monte Carlo e dessa maneira utilizar o gráfico de envelopes como uma boa medida de qualidade de ajuste.

REFERÊNCIAS

- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. **Regression diagnostics**. New York: John Wiley, 1980. 292 p.
- BERGER, J. O.; SUN, D. Bayesian analysis for the poly-Weibull distribution. **Journal of the American Statistical Association**, Alexandria, v.88, p.1412-1418, 1993.
- COLOSIMO, E. A.; GIOLO S. R. **Análise de sobrevivência aplicada**, São Paulo: Blucher, 2006. 392 p.
- COOK, R. D. Detection of influential observations in linear regression. **Technometrics** , Alexandria, v.19, p.15-18, 1977.
- COOK, R. D. Assesment of local influence (with dicussion). **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.
- COOK, R. D.; PEÑA, D.; WEISBERG, S. The likelihood displacement: A unifying principle for influence measures. **Communications in Statistics: Theory and Methods**, Philadelphia, v.17, p.623-640, 1988.
- COX, D. R.; OAKES, D. **Analysis of survival data**. New York: Chapaman e Hall, 1984. 201 p.
- COX, D. R.; SNELL, E. J. A General definition of residuals. **Journal of the Royal Statistical Society, Series B**, London, v.30, p.248-275, 1968.
- DAVISON, A. C.; LOUZADA-NETO, F. Inference for the poly-Weibull model. **The Statistician**, London, v.49, n°2, p.189-196, 2000.
- FUNG, W. K.; KWAN, C. W. A note on local influence based on normal curvature. **Journal of the Royal Statistical Society, Series B**, London, v.59, p.839-843, 1997.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**, 2nd ed., New

York: Wiley, 2003. 439 p.

LEE, E. T. **Statistical models and for survival data analysis**, 2nd ed., New York: Wiley, 1992. 482 p.

LEE, S. Y, LU, B.; SONG, X. Y. Assessing local influence for nonlinear structural equation models with ignorable missing data. **Computational Statistics and Data Analysis**, New York, v.50, p.1356-1377, 2006.

LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, Washington, v.54, p.570-582, 1998.

LOUZADA-NETO, F. Polyhazard models for lifetime data. **Biometrics**, Washington, v.55, p.1281-1285, 1999.

MALLER, R.; ZHOU, X. **Survival analysis with long-term survivors**. New York: Wiley, 1996. 278 p.

MAZUCHELI, J.; LOUZADA-NETO, F.; ACHCAR, J. A. Bayesian Inference for polyhazard models in the presence of covariates. **Computational Statistics and Data Analysis**, New York, v.38, p.1-14, 2001.

MCCULLAGH, P. e NELDER, J. A. **Generalized linear models**, 2nd ed., London: Chapman e Hall, 1989. 511 p.

MUDHOLKAR, G. S; SRIVASTAVA, D. K.; FRIEMER, M. The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. **Technometrics** , Alexandria, v.37, p.436-445, 1995.

ORTEGA, E. M. M. **Análise de influência e resíduos em modelos de regressão log-gama generalizados**. 2001. p.159. Tese (Doutorado em Estatística) - Instituto de Matemática, Estatística da Universidade de São Paulo, 2001.

ORTEGA, E. M. M.; BOLFARINE, H.; PAULA, G. A. Influence diagnostics in generalized log-gamma regression models. **Computational Statistics and Data Analysis**, New

York, v.42, p.165-186, 2003.

ORTEGA, E. M. M., CANCHO, V. G.; BOLFARINE, H. Influence diagnostics in exponentiated-Weibull regression models with censored data. **Statistics and Operations Research Transactions**, Espanha, v.30, (to appear), 2006.

PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: IME-USP, 2004. 245 p.

SOUZA, E. C. **Análise de influência local no modelo de regressão logística**. 2006. 101p. Tese (Mestrado em Estatística e Experimentação Agronomica) - Escola Superior de Agricultura "Luis de Queiroz", Universidade de São Paulo, Piracicaba, 2006.

SOUZA, F. A. M. **Influência local e análise de resíduos em modelos de regressão Von Mises**. 1999. 116p. Tese (Doutorado em Estatística) - Instituto de Matemática, Estatística da Universidade de São Paulo, São Paulo, 1999.

THERNEAU, T. M.; GRAMBSCH, P. M. e FLEMING, T. R. Martingal-based residuals for survival models. **Biometrika**, London, v.77, p.147-160, 1990.

THOMAS, W. ; COOK, R. D. Assessing influence on regression coefficients in generalized linear models. **Biometrika**, London, v.76, p.741-749, 1989.

THOMAS, W. ; COOK, R. D. Assessing influence on predictions from generalized linear models. **Technometrics**, Alexandria, v.32, p.59-65, 1990.

ANEXOS

Programa no SAS para estimar os parâmetros do modelo bi-Weibull

```

data lista3;
input pred exp_pred t lntiempo censor inter orilla1 axl longf prof1 arc_irve
tran_edw horainic obs;
cards;
*** Dados ***
proc sort;
by censor;
run;

*** Achar o valor inicial das covariadas ***
proc reg;
model t= longf prof1 tran_edw;
by censor;
run;

*** Achar o valor inicial para o risco ***
proc lifereg;
    model t*censor(0)= / dist=weibull;
run;

*** Ajustando o modelo bi-Weibull aos dados ***

proc nlmixed data=lista3;
parms beta01=7.505 beta02=7.505 beta11=-0.154 beta12=-0.154 beta21=0.063
beta22=0.063 beta31=-0.595 beta32=-0.595 risco1=0.25 risco2=0.28;
bounds risco1 > 0, risco2 > 0;
BETA1= beta01+ beta11*longf+ beta21*prof1+ beta31*tran_edw;
BETA2= beta02+ beta12*longf+ beta22*prof1+ beta32*tran_edw;

```

```

a= risco1*(t**((risco1)-1))*exp(BETA1);
b= risco2*(t**((risco2)-1))*exp(BETA2);
c=(t**(risco1))*exp(BETA1);
d=(t**(risco2))*exp(BETA2);
lv=((censor*log(a+b))-(c+d));
model t ~ general(lv);
run;

```

Programa no SAS para estimar os parâmetros do modelo bi-log-logístico

```

data lista3;
input pred exp_pred t lntiempo censor inter orilla1 axl longf prof1 arc_irve
tran_edw horainic obs;
cards;
*** Dados ***
proc sort;
by censor;
run;

*** Achar o valor inicial das covariadas ***
proc reg;
model t= longf prof1 tran_edw;
by censor;
run;

*** Achar o valor inicial do risco ***
proc lifereg;
    model t*censor(0)= / dist=llogistic;
run;

```


*** Ajustando o modelo bi-log-logístico aos dados ***

```
proc nlmixed data=lista3;
parms beta01=7.505 beta02=7.505 beta11=-0.154 beta12=-0.154 beta21=0.063
beta22=0.063 beta31=-0.595 beta32=-0.595 risco1=0.25 risco2=0.374;
bounds risco1 > 0, risco2 > 0; *alfa1 > 0, alfa2 > 0;
BETA1x= beta01+ beta11*longf+ beta21*prof1+ beta31*tran_edw;
BETA2x= beta02+ beta12*longf+ beta22*prof1+ beta32*tran_edw;
a=(risco1*(t**(risco1-1))*exp(BETA1x))/(1+((t**risco1)*exp(BETA1x)));
b=(risco2*(t**(risco2-1))*exp(BETA2x))/(1+((t**risco2)*exp(BETA2x)));
c=log(1+((t**risco1)*exp(BETA1x)));
d=log(1+((t**risco2)*exp(BETA2x)));
lv=((censor*log(a+b))-(c+d));
model t ~ general(lv);
run;
```

Programa no OX para estimar os parâmetros do modelo bi-Weibull

```
#include<oxstd.h>
#include<oxdraw.h>
#include<oxfloat.h>
#include<maximize.h>
#include<simula.h>
#pragma link("maximize.oxo")
static decl g_mY;
static decl g_mX;
log_vero(const vP,const adFunc, const avScore,const amHessian)
{
decl n=rows(g_mY);
decl cont,t,xx0, xx1, xx2, xx3, xx4,xx5, xx6, xx7;
decl uns=ones(n,1);
```

```

decl vero=zeros(1,n);
decl a1=fabs(vP[0][0]); //w1
decl a2=fabs(vP[1][0]); //w2
decl a3=vP[2][0];      //beta 01
decl a4=vP[3][0];      //beta 11
decl a5=vP[4][0];      //beta 21
decl a6=vP[5][0];      //beta 31
decl a7=vP[6][0];      //beta 02
decl a8=vP[7][0];      //beta 12
decl a9=vP[8][0];      //beta 22
decl a10=vP[9][0];     //beta 32

for(cont=0;cont<n;++cont)
{
//t=log(g_mY[cont][0]);
t=g_mY[cont][0];
xx0=g_mY[cont][2];
xx1=g_mY[cont][3];
xx2=g_mY[cont][4];
xx3=g_mY[cont][5];

decl beta1=a3*xx0+a4*xx1+a5*xx2+a6*xx3;
decl beta2=a7*xx0+a8*xx1+a9*xx2+a10*xx3;

if(g_mY[cont][1]==1)
vero[0][cont]=log(a1*(t^(a1-1))*exp(beta1)+a2*(t^(a2-1))*exp(beta2))
-((t^(a1))*exp(beta1)+(t^(a2))*exp(beta2));
if(g_mY[cont][1]==0) vero[0][cont]=-((t^(a1))*exp(beta1)
+(t^(a2))*exp(beta2));
}

```

```

adFunc[0]=double(vero*uns);
return 1;
}
main()
{
    g_mY = loadmat("a:juliana.txt") ;

    print(" Dados",g_mY);
    decl nc=rows(g_mY);
    decl dfunc;
    //decl vP=<1.25;1.28;7.505;-0.154;0.63;-0.595;7.505;-0.154;0.63;-0.595>;
    decl vP=<0.56;2;-5.38; 0.077;-0.011;-0.706;5.82;-0.11;-0.012;0.9>;

    log_vero(vP,&dfunc,0,0);
    println("vero=" ,dfunc);

    MaxControl(-1,20);
    decl mhess=0.00001*unit(4);
    decl ir, var, var1;
    ir=MaxBFGS(log_vero,&vP,&dfunc,&mhess,1);
    //decl zz=(g_mY[] [0]-vP[3] [0]-g_mY[] [2] .*vP[3] [0]) ./vP[1] [0];
    //print(zz);
    Num2Derivative(log_vero,vP,&mhess);
    var=-1/mhess;
    var1=invertsym((-1)*mhess);
    print(" os valores da mhess  \n",mhess);
    print(" os valores das variâncias  \n",var);
    print(" os valores das variâncias 1  \n",var1);
    print(" os valores dos parametros \n", vP);
}

```

Programa no OX para estimar os parâmetros do modelo bi-log-logístico

```
#include<oxstd.h>
#include<oxdraw.h>
#include<oxfloat.h>
#include<maximize.h>
#include<simula.h>
#pragma link("maximize.oxo")
static decl g_mY;
static decl g_mX;
log_vero(const vP,const adFunc, const avScore,const amHessian)
{
decl n=rows(g_mY);
decl cont,t,xx0, xx1, xx2, xx3, xx4,xx5, xx6, xx7;
decl uns=ones(n,1);
decl vero=zeros(1,n);
decl a1=fabs(vP[0][0]); //w1
decl a2=fabs(vP[1][0]); //w2
decl a3=vP[2][0];      //beta 01
decl a4=vP[3][0];      //beta 11
decl a5=vP[4][0];      //beta 21
decl a6=vP[5][0];      //beta 31
decl a7=vP[6][0];      //beta 02
decl a8=vP[7][0];      //beta 12
decl a9=vP[8][0];      //beta 22
decl a10=vP[9][0];     //beta 32

for(cont=0;cont<n;++cont)
{
```

```

//t=log(g_mY[cont][0]);
t=g_mY[cont][0];
xx0=g_mY[cont][2];
xx1=g_mY[cont][3];
xx2=g_mY[cont][4];
xx3=g_mY[cont][5];

decl beta1=a3*xx0+a4*xx1+a5*xx2+a6*xx3;
decl beta2=a7*xx0+a8*xx1+a9*xx2+a10*xx3;

if(g_mY[cont][1]==1)
vero[0][cont]=log(((a1*(t^(a1-1))*exp(beta1))*((1+((t^(a1))*exp(beta1)))^(-1)))
+((a2*(t^(a2-1))*exp(beta2))*((1+((t^(a2))*exp(beta2)))^(-1))))
-(log(1+( t^(a1)*exp(beta1)      ))      +log(1+(t^(a2)*exp(beta2))))
;
if(g_mY[cont][1]==0)
vero[0][cont]=-(log(1+( t^(a1)*exp(beta1)))+log(1+(t^(a2)*exp(beta2))))
;
}
adFunc[0]=double(vero*uns);
return 1;
}
main()
{
  g_mY = loadmat("a:juliana.txt") ;

  //print(" Dados",g_mY);
  decl nc=rows(g_mY);
  decl dfunc;

```

```
//decl vP=<0.374;0.25;7.505;7.505;-0.154;-0.154;0.063;0.063;-0.595;-0.595>;
```

```
decl vP=<0.7463;8.7760;
-0.8223;0.03581;-0.02160; -0.2598;
-28.7129; -0.1945 ;0.07480;-3.3458;
>;
```

```
log_vero(vP,&dfunc,0,0);
println("vero=" ,dfunc);
```

```
MaxControl(-1,20);
decl mhess=0.00001*unit(4);
decl ir, var, var1;
ir=MaxBFGS(log_vero,&vP,&dfunc,&mhess,1);
//decl zz=(g_mY[] [0]-vP[3] [0]-g_mY[] [2] .*vP[3] [0]) ./vP[1] [0];
//print(zz);
Num2Derivative(log_vero,vP,&mhess);
var=-1/mhess;
var1=invertsym((-1)*mhess);
print(" os valores da mhess \n",mhess);
print(" os valores das variâncias \n",var);
print(" os valores das variâncias 1 \n",var1);
print(" os valores dos parametros \n", vP);
}
```