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

**Modelo de regressão para dados com censura intervalar e dados
de sobrevivência grupados**

Elizabeth Mie Hashimoto

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

Piracicaba
2008

Elizabeth Mie Hashimoto
Bacharel em Estatística

**Modelo de regressão para dados com censura intervalar e dados
de sobrevivência grupados**

Orientador:

Prof. Dr. **EDWIN MOISES MARCOS ORTEGA**

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

Piracicaba
2008

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Hashimoto, Elizabeth Mie

Modelo de regressão para dados com censura intervalar e dados de sobrevivência
grupados / Elizabeth Mie Hashimoto. - - Piracicaba, 2008.
121 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Análise de regressão e correlação 2. Análise de sobrevivência 3. Bootstrap Jackknife
re-amostragem 4. Dados censurados 5. Distribuições (Probabilidade) 6. Verosimilhança I.
Título

CDD 519.536
H348m

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

DEDICATÓRIA

Aos meus pais,
Agostinho Hashimoto e Eiko Kurihara Hashimoto,
pelo incentivo e o esforço pela minha formação.

Aos meus "jiichann's",
Tadashi Hashimoto(in memorian) e Kazuji Kurihara
(in memorian), o exemplo de vida que foram e por sem-
pre acreditarem em mim.

Ao meu irmão,
Fábio Jun Iti Hashimoto, pela amizade e apoio.

*A eles, dedico este trabalho
com amor e carinho.*

AGRADECIMENTOS

Ao Prof. Dr. Edwin, a orientação, as sugestões, as discussões, a paciência, a amizade que contribuíram no meu crescimento e na minha formação acadêmica.

A Deus por tudo, pela minha saúde, pela minha família, e por ter me dado a oportunidade desta conquista.

Ao Prof. Dr. Barreto por ceder os dados de vitamina A.

A toda a minha família, em especial aos meus pais, a minha avó, ao meu irmão Fábio que de alguma forma contribui para prosseguir os meus estudos e aos meus tios Alécio, Isa, Nelson e Shirley pelo apoio e o todo o suporte neste período de mestrado.

À Andréia e à Juliana pela amizade, pelos estudos e pelo conforto de família que muito me ajudou no desenvolvimento deste trabalho.

Aos professores do curso de Pós-graduação em Estatística e Experimentação Agronômica, Dr. César Gonçalves de Lima, Dra. Clarice Garcia Borges Demétrio, Dr. Décio Barbin, Dr. Gerson Barreto Mourão, Dra. Roseli Aparecida Leandro, Dr. Silvio Sandoval Zocchi e Dra. Sônia Maria Stefano Piedade, pela formação.

Aos amigos, Cláudia, Giovana, Mariana, Marisol, Michele, Sérgio e Vanderly, um agradecimento especial pela ajuda no desenvolvimento deste trabalho, pela amizade e pelos conselhos.

Aos colegas de estudo do mestrado e do doutorado, pela amizade e companheirismo, em particular, ao Afrânio, à Angela, à Fernanda e à Lucimary por verificarem a tradução do trabalho e por toda a amizade.

Aos funcionários do Departamento de Ciências Exatas da ESALQ / USP, em especial, às secretárias Luciane Brajão e Solange de Assis Paes Sabadin, o técnico em informática Eduardo e a funcionária Terezinha pelos auxílios permanentes.

Aos meus amigos de graduação que sempre me incentivaram e aos meus amigos do seinen-kai, principalmente, André Endo, Júlio Wada, Koiti Miyazaki, Marcio Watanabe, Marcia Miyashita, Massumi Miyazaki, Missuzu Fukui, Sônia Wada, Yukie Hirata, Wellington Mizobuchi e William Takata pela amizade e por sempre estarem me ajudando.

Às amigas Flávia e Terezinha pelos momentos de diversão e amizade.

À CNPq pelo apoio financeiro.

Às pessoas que contribuíram de alguma forma para a realização do presente trabalho.

SUMÁRIO

RESUMO	7
ABSTRACT	8
1 INTRODUÇÃO	9
Referências	12
2 MODELOS DE REGRESSÃO WEIBULL-EXPONENCIADA PARA DADOS COM CEN-	
SURA INTERVALAR	15
Resumo	15
Abstract	15
2.1 Introdução	16
2.2 Modelo de regressão Weibull-exponenciada para dados de censura intervalar	18
2.2.1 Estimador de máxima verossimilhança para o modelo com censura intervalar	22
2.2.2 Testes de hipóteses	24
2.2.3 Estimador Jackknife	25
2.3 Análise de sensibilidade	26
2.3.1 Influência global	27
2.3.2 Influência local	27
2.3.3 Cálculo das curvaturas máximas	28
2.4 Análise de resíduos	36
2.4.1 Resíduo de Cox-Snell ajustado	36
2.4.2 Resíduo de martingale	36
2.4.3 Resíduo deviance modificado	37
2.4.4 Estudo de simulação	38
2.5 Aplicação	44
2.5.1 Estimador de máxima verossimilhança e estimador Jackknife	45
2.5.2 Influência global	46
2.5.3 Influência local	47
2.5.4 Análise de resíduos	50
2.5.5 Impacto da deleção de observações influentes	50
2.5.6 Qualidade de ajuste	51
2.6 Conclusão	52
2.6.1 Trabalhos futuros	52
Referências	54
3 MODELO DE REGRESSÃO PARA DADOS DE SOBREVIVÊNCIA GRUPADOS	60
Resumo	60

Abstract	60
3.1 Introdução	60
3.2 Especificação do modelo de regressão para dados grupados	62
3.3 Estimador de máxima verossimilhança	65
3.4 Estimador Jackknife	69
3.5 Critérios seleção de modelos	70
3.6 Análise de sensibilidade	70
3.6.1 Influência Global	71
3.6.2 Influência Local	72
3.6.3 Cálculo das curvaturas máximas	73
3.7 Análise de resíduos	82
3.8 Aplicação	83
3.8.1 Seleção de modelos	85
3.8.2 Estimação de máxima verossimilhança e Jackknife	86
3.8.3 Análise de influência global	88
3.8.4 Análise de influência local e total	89
3.8.5 Análise de resíduo	91
3.8.6 Impacto da deleção de observações influentes	92
3.9 Conclusão	94
3.9.1 Trabalhos futuros	94
Referências	96
APÊNDICES	102

RESUMO

Modelo de regressão para dados com censura intervalar e dados de sobrevivência grupados

Neste trabalho foi proposto um modelo de regressão para dados com censura intervalar utilizando a distribuição Weibull-exponenciada, que possui como característica principal a função de taxa de falha que assume diferentes formas (unimodal, forma de banheira, crescente e decrescente). O atrativo desse modelo de regressão é a sua utilização para discriminar modelos, uma vez que o mesmo possui como casos particulares os modelos de regressão Exponencial, Weibull, Exponencial-exponenciada, entre outros. Também foi estudado um modelo de regressão para dados de sobrevivência grupados na qual a abordagem é fundamentada em modelos de tempo discreto e em tabelas de vida. A estrutura de regressão representada por uma probabilidade é modelada adotando-se diferentes funções de ligação, tais como, logito, complemento log-log, log-log e probito. Em ambas as pesquisas, métodos de validação dos modelos estatísticos propostos são descritos e fundamentados na análise de sensibilidade. Para detectar observações influentes nos modelos propostos, foram utilizadas medidas de diagnóstico baseadas na deleção de casos, denominadas de influência global e medidas baseadas em pequenas perturbações nos dados ou no modelo proposto, denominada de influência local. Para verificar a qualidade de ajuste do modelo e detectar pontos discrepantes foi realizada uma análise de resíduos nos modelos propostos. Os resultados desenvolvidos foram aplicados a dois conjuntos de dados reais.

Palavras-chave: Distribuição Weibull-Exponenciada; Censura intervalar; Modelo de regressão; Dados de sobrevivência grupados; Funções de ligação; Análise de sensibilidade; Análise de resíduos

ABSTRACT

Regression model for interval-censored data and grouped survival data

In this study, a regression model for interval-censored data were developed, using the Exponentiated-Weibull distribution, that has as main characteristic the hazard function which assumes different forms (unimodal, bathtub shape, increase, decrease). A good feature of that regression model is their use to discriminate models, that have as particular cases, the models of regression: Exponential, Weibull, Exponential-exponentiated, amongst others. Also a regression model were studied for grouped survival data in which the approach is based in models of discrete time and in life tables, the regression structure represented by a probability is modeled through the use of different link function, logit, complementary log-log, log-log or probit. In both studies, validation methods for the statistical models studied are described and based on the sensitivity analysis. To find influential observations in the studied models, diagnostic measures were used based on case deletion, denominated as global influence and measures based on small perturbations on the data or in the studied model, denominated as local influence. To verify the goodness of fitting of the model and to detect outliers it was performed residual analysis for the proposed models. The developed results were applied to two real data sets.

Keywords: Exponentiated-Weibull distribution; Interval-censored data; Regression model; Grouped survival data; Link function; Sensitivity analysis; Residual analysis

1 INTRODUÇÃO

A análise estatística de tempos de vida é um tópico importante utilizada em diferentes áreas tais como, por exemplo, medicina, biologia, epidemiologia, engenharia, entre outros. Tempo de falha se refere ao tempo até a ocorrência de um evento de interesse, podendo ser a morte, o aparecimento de um tumor, o desenvolvimento de uma doença, a quebra de um componente eletrônico e assim por diante.

O tempo de falha, denotado por T , é uma variável aleatória não negativa, caracterizada através da função de sobrevivência, definida como

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(t)dt,$$

que é a probabilidade de um indivíduo sobreviver até o tempo t e $F(t)$ é a função de distribuição acumulada da variável T .

Outra função importante é a função de taxa de falha que é definida por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)},$$

essa função é útil para descrever a distribuição do tempo de vida das observações em estudo.

A principal característica dos dados de tempo de vida é a presença de censura, que é a observação parcial da resposta. Essa característica diferencia análise de sobrevivência das outras áreas, tais como, planejamento de experimentos, modelos lineares generalizados, entre outros.

Em dados com censura intervalar, o tempo do evento de interesse não é observado, mas é conhecido o intervalo de tempo onde ocorreu o evento de interesse. Para modelar esse tipo de dados existem várias abordagens, por exemplo, Finkelstein e Wolfe (1985) propuseram uma técnica semi-paramétrica para análise de regressão para dados com censura intervalar. Finkelstein (1986) propôs um método para ajustar o modelo de riscos proporcionais quando os dados apresentam censura à esquerda, à direita e intervalar. Odell, Anderson e D'Agostino (1992) utilizaram um modelo de tempo de falha acelerado baseado na distribuição Weibull em dados com censura intervalar. Lindsey e Ryan (1998) apresentaram um tutorial sobre modelagem de dados com censura intervalar. Sinha, Chen e Ghosh (1999) utilizaram métodos bayesianos para analisar dados com censura intervalar. Farrington (2000) apresentou uma análise de resíduos para modelos de riscos proporcionais com censura intervalar. Langohr, Gómez e Muga (2004) consideraram um modelo de regressão Weibull para dados com dupla censura. Zhang et al. (2007) propuseram um modelo assumindo censura intervalar informativa e utilizaram o algoritmo EM para estimar os parâmetros.

Neste trabalho, propõe-se um modelo de regressão log-linear para modelar dados com censura intervalar utilizando a distribuição Weibull-exponenciada. A vantagem dessa distribuição é que ela possui como casos particulares outras distribuições usuais em análise de sobrevivência. Outra

característica da distribuição Weibull-exponenciada é em relação a flexibilidade da função de taxa de falha para modelar diferentes formas, tais como, crescente, decrescente, unimodal e forma de banheira.

Por outro lado, dados grupados podem ser considerados um caso particular de dados de censura intervalar quando todas as unidades amostrais são avaliados nos mesmos tempos. Muitas vezes, este tipo de dado é identificado por um número excessivo de empates (COLOSIMO; GIOLO, 2006). Na literatura existem muitas formas de abordar esse problema. Prentice e Gloeckler (1978) propuseram um modelo de riscos proporcionais modificado para acomodar dados de sobrevivência grupados. Allison (1982) descreveu métodos para análise de dados de sobrevivência grupados. Baker, Wax e Patterson (1993) estudaram a modelagem de dados de sobrevivência grupados sujeitos a censura informativa. Chalita (1997) apresentou critérios de seleção entre modelos discretos e contínuos. Heitjan (1989) apresentou um tutorial sobre modelagem de dados de sobrevivência grupados.

Outro enfoque desta pesquisa é propor um modelo de regressão para dados grupados construindo a função de verossimilhança da mesma forma que o trabalho de Thompson (1977) e Colosimo, Chalita e Demétrio (2000). Além disso, foram utilizadas quatro funções de ligação tais como, logito, complemento log-log, log-log e probito para modelar a estrutura de regressão.

Em ambos os modelos de regressão propostos, modelo de regressão Weibull-exponenciada com censura intervalar e modelos de regressão para dados grupados, os procedimentos inferenciais para estimação dos parâmetros dos modelos são estudados.

Para estimar os parâmetros desses modelos de regressão é utilizado o método de máxima verossimilhança, em que o procedimento numérico do tipo Quase-Newton foi necessário. Também é realizado testes de hipóteses a partir da distribuição assintótica dos estimadores de máxima verossimilhança. Entretanto, usar a distribuição assintótica dos estimadores de máxima verossimilhança em situações cuja amostra é pequena, situação comum na prática, ou que as condições de regularidade da função da verossimilhança não sejam satisfeitas pode ser inadequado e conduzir a resultados difíceis de serem justificados. Assim, como alternativa ao método de máxima verossimilhança, o método de Jackknife é utilizado.

Numa etapa posterior ao ajuste de um modelo a um conjunto de dados, é necessário verificar as suposições do modelo. Caso o modelo não seja adequado pode levar a conclusões errôneas. É importante ainda verificar a presença de observações discrepantes ou observações influentes, que podem causar distorções nos resultados das estimativas dos parâmetros do modelo. Para detectar observações influentes nos modelos propostos, medidas de diagnóstico baseadas no caso deleção (COOK, 1977), denominada de influência global e medidas de influência baseadas em pequenas perturbações nos dados ou no modelo, chamada de influência local (COOK, 1986) são empregadas. Além dessas medidas, também é utilizada a medida de influência local total desenvolvida por Lesaffre e Verbeke (1998). Para verificar se existe afastamento das suposições feitas ao modelo

e detectar pontos atípicos, uma análise de resíduos baseados nos trabalhos de Collett (2003), Klein e Moeschberger (1997) e Lawless (2003) que descreveram os resíduos de Cox-Snell, Martingale e Deviance, Collett (1991) e Ortega, Souza e Cancho (2006) que descreveram o resíduo de Person e Farrington (2000) que apresentou resíduos modificados para acomodar dados com censura intervalar é realizada. Finalmente, os resultados são aplicados à dois conjuntos de dados reais.

O objetivo principal deste trabalho é apresentar o modelo de regressão Weibull-exponenciada para dados com censura intervalar e o modelo de regressão para dados de sobrevivência grupados utilizando diferentes funções de ligação para modelar a estrutura de regressão. Os outros objetivos estão relacionados à diagnóstico e qualidade de ajuste dos modelos. Assim, o trabalho está organizado da seguinte maneira. No capítulo 2 é proposto o modelo de regressão Weibull-exponenciada para dados com censura intervalar. Na seção 2.1 é apresentada uma introdução do trabalho, na seção 2.2, o modelo de regressão Weibull-exponenciada com censura intervalar é proposto e utiliza-se o método de máxima verossimilhança e o método de Jackknife na estimação dos parâmetros. Também é discutido os possíveis testes de hipóteses para discriminar modelos. Na seção 2.3, calcula-se as medidas de influência global e as curvaturas normais para a influência local sob cinco esquemas de perturbação para o modelo proposto são obtidas. Na seção 2.4, são apresentados os resíduos de Cox-Snell, martingale e deviance modificados para acomodar dados com censura intervalar com a finalidade de avaliar possíveis afastamentos da suposição da distribuição dos erros e encontrar possíveis observações discrepantes. Estudos de simulação de Monte Carlo também são realizados para comparar a distribuição empírica dos resíduos com a distribuição normal padrão. Na seção 2.5, a metodologia proposta é aplicada a um conjunto de dados reais obtido da literatura e por fim, na seção 2.6, encontra-se as principais conclusões e descreve-se o direcionamento da continuidade deste trabalho.

No capítulo 3 é discutido o modelo de regressão para dados de sobrevivência grupados utilizando funções de ligação para modelar a estrutura de regressão. Na seção 3.1 é feita uma introdução do trabalho, na seção 3.2 é descrito o modelo de regressão para dados de sobrevivência grupados bem como as funções de ligação utilizadas no trabalho. Na seção 3.3, o método de estimação de máxima verossimilhança é estudado. Na seção 3.4 é abordado o método de Jackknife. Na seção 3.5 alguns critérios para seleção de modelos são descritos, na seção 3.6 uma análise de sensibilidade baseadas nas técnicas de influência global e influência local é realizada. No caso das medidas de influência local, os esquemas de perturbação de casos e de uma covariável contínua são considerados e as matrizes de curvatura máxima são calculadas. Na seção 3.7 com o objetivo de verificar observações discrepantes propôs-se o resíduo de Pearson. Na seção 3.8, a metodologia desenvolvida é aplicada a um conjunto de dados reais e por fim na seção 3.9 encontra-se as principais conclusões do trabalho e assuntos que podem dar continuidade aos estudos envolvendo dados de sobrevivência grupados.

Referências

- ALLISON, P.D. Discrete-time methods for the analysis of event histories. **Sociological Methodology**, Washington, v.13, p.61-98, 1982.
- BAKER, S.G.; WAX, Y.; PATTERSON, B.H. Regression analysis of grouped survival data: informative censoring and double sampling. **Biometrics**, Washington, v.49, p.379-389, 1993.
- CHALITA, L.V.A.S. **Modelos para dados agrupados e censurados**. 1997.135p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1997.
- COLLETT, D. **Modelling survival data in medical Research**. 2nd ed. London: Chapman & Hall, 2003. 391p.
- _____. **Modelling binary data**. London: Chapman & Hall, 1991. 369p.
- COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 374p.
- COLOSIMO, E.A.; CHALITA, L.V.A.S.; DEMÉTRIO, G.B. Tests of proportional hazards and proportional odds models for grouped survival data. **Biometrics**, Washington, v.56, p.1233-1240, 2000.
- COOK, R. D. Detection of influential observations in linear regression. **Technometrics**, Alexandria, v.19, p.15-18, 1977.
- _____. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.
- FARRINGTON, C. P. Residuals for proportional hazards models with interval-censored survival data. **Biometrics**, Washington, v.56, p.473-482, 2000.

FINKELSTEIN, D.M. A proportional hazards model for interval-censored failure time data. **Biometrics**, Washington, v.42, p.845-854, 1986.

FINKELSTEIN, D.M.; WOLFE, R.A. Semi-parametric model for regression analysis of interval censored failure time data. **Biometrics**, Washington, v.41, p.933-945, 1985.

GIOLO, S.R.; COLOSIMO, E.A.; DEMÉTRIO, C.G.B. Different approaches for modelling grouped survival data: a mango tree study. **Journal of Agricultural Biological and Environmental Statistics**, Alexandria, v.42, p.165-186, 2008.

HEITJAN, D.F. Inference from grouped continuous data: a review. **Statistical Science**, Hayward, v.4, p.164-179, 1989.

KLEIN, J.P.; MOESCHBERGER, M.L. **Survival analysis techniques for censored and truncated data**. New York: Springer-Verlang, 1997. 357p.

LANGOHR, K.; GÓMEZ, G.; MUGA, R. A parametric survival model with an interval-censored covariate. **Statistics in Medicine**, Chichester, v.23, p.3159-3175, 2004.

LAWLESS, J.F. **Statistical models and methods for lifetime data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 637p.

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

LINDSEY, J.C.; RYAN, L.M. Tutorial in biostatistics methods for interval-censored data. **Statistics in Medicine**, Chichester, v.17, p.129-238, 1998.

MUDHOLKAR, G.S.; SRIVASTAVA, D.K.; KOLLIA, G.D. A Generalization of the Weibull Distribution with Application to the Analysis of Survival Data. **Journal of the American Statistical Association**, Boston, v.91, p.1575-1583, 1996.

ODELL, P.M.; ANDERSON, K.M.; D'AGOSTINO, R.B. Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. **Biometrics**, Washington, v.48, p.951-959, 1992.

ORTEGA, E.M.M.; SOUZA, E.C.; CANCHO, V.G. Aplicação da análise de influência local em modelos de regressão logística. **Revista de Matemática e Estatística**, Jaboticabal, v.24, p.120-149, 2006.

PRENTICE, R.L.; GLOECKLER, L.A. Regression analysis of grouped survival data with application to breast cancer data. **Biometrics**, Washington, v.34, p.57-67, 1978.

SINHA, D.; CHEN, M.; GHOSH, S.K. Bayesian analysis and model selection for interval-censored survival data. **Biometrics**, Washington, v.55, p.585-590, 1999.

THOMPSON JR, W.A. On the treatment of grouped observation in life studies. **Biometrics**, Washington, v.33, p.463-470, 1977.

ZHANG, Z.; SUN, L.; SUN, J.; FINKELSTEIN, D.M. Regression analysis of failure time data with informative interval censoring. **Statistics in Medicine**, Chichester, v.26, p.2533-2546, 2007.

2 MODELOS DE REGRESSÃO WEIBULL-EXPONENCIADA PARA DADOS COM CENSURA INTERVALAR

Resumo

Dados com censura intervalar resultam de estudos no qual o evento de interesse não é diretamente observado, mas conhecido somente o intervalo de tempo aleatório. Neste trabalho, um modelo de regressão baseado na distribuição Weibull-exponenciada foi proposto para modelar dados com censura intervalar. A vantagem desse modelo consiste no fato de representar uma família paramétrica que possui como casos particulares outros modelos de regressão utilizados na análise de tempos de vida. Os parâmetros do modelo foram estimados por meio dos métodos de máxima verossimilhança e Jackknife. Alguns métodos de análise de sensibilidade como influência global, influência local e total foram introduzidas para encontrar pontos influentes. Além dos esquemas usuais de perturbação descreveu-se três esquemas de perturbação quando os dados são observados na presença de censura intervalar. Uma análise de resíduos também foi realizada para verificar possíveis observações discrepantes bem como uma medida de qualidade de ajuste do modelo proposto. Nesse contexto, os resíduos usuais em análise de sobrevivência foram modificados para incorporar as censuras do tipo intervalar e simulações de Monte Carlo foram realizadas para comparar a distribuição empírica desses resíduos com a distribuição normal padrão. A relevância dos métodos de análise de regressão para dados com censura intervalar é ilustrada em um conjunto de dados reais.

Palavras-chave: Distribuição Weibull-Exponenciada; Censura intervalar; Modelo de regressão; Análise de sensibilidade; Análise de resíduo

Abstract

Interval-censored data results of studies on which the event of interest is not observed directly, but only known for a random time interval. In this study, a regression model based on the Exponentiated-Weibull distribution was proposed to model interval-censored data. The advantage of this model consists on the fact that it represents a parametric family that has, as particular cases, other regression models used in the lifetime data analysis. The model's parameters were estimated through the maximum likelihood and Jackknife methods. Some sensitivity analysis methods as global influence, local and total influence were introduced to find influential points and outliers. In addition to the usual perturbation schemes, it were described three perturbation schemes when the data were observed in the presence of interval censoring. Residual analysis was also accomplished to verify goodness of fitting of the proposed model. In that context, the usual residuals in survival analysis were modified to incorporate the censorships of the interval type and Monte Carlo simulations were performed to compare the empirical distribution of those residuals with the standard normal distribution. The relevance of the methods of regression analysis for interval-censored data is illustrated with a real data set.

Keywords: Exponentiated-Weibull distribution; Interval-censored data; Regression model; Sensitivity analysis; Residual analysis

2.1 Introdução

O que distingue a análise de sobrevivência de outras áreas da estatística é a presença de censura que é uma característica de dados de tempo de vida. Por censura, entende-se que o tempo de falha de um indivíduo não é exatamente observado devido a limitações do estudo. Ao invés disso, somente uma informação parcial sobre o tempo de falha é observado. Há vários tipos de mecanismos de censura incluindo censura à direita, à esquerda e intervalar.

Neste trabalho, o interesse são os dados com censura intervalar, que ocorrem naturalmente em estudos epidemiológicos e médicos em que há uma avaliação periódica. Por exemplo, o tempo de falha associado a uma variável aleatória T é definido como sendo o tempo de infecção por HIV em um estudo de AIDS. Uma vez que a infecção por HIV somente é determinada por exames de sangue durante visitas periódicas, o tempo de ocorrência exato de T não é observado, nesse caso, o tempo de infecção é conhecido somente dentro de um intervalo, denotado por $(U, V]$, em que U representa a última avaliação com o teste negativo e V representa a primeira avaliação com teste positivo. Assim temos que, a variável T apresenta censura intervalar.

De acordo com Lindsey e Ryan (1998), os que tempos exatos de falha, bem como censuras à direita e à esquerda são casos particulares de censura intervalar. Quando $U = V$ tem-se tempos exatos de falha observados, quando $V = \infty$ tem-se censura à direita e por último quando $U = 0$ tem-se censura à esquerda, sendo este último pouco abordado na literatura.

Várias abordagens e soluções para problemas envolvendo dados com censura intervalar foram propostos na literatura, principalmente com relação à análise não paramétrica e semi-paramétrica. Por exemplo, Rabinowitz, Tsiatis e Aragon (1995) propuseram uma classe de estatísticas score utilizadas nos procedimentos de estimação e construção de intervalos de confiança. Para modelos com censura intervalar, Rosingana (1996) propôs uma análise de dados com censura intervalar sobre o enfoque não paramétrico e bayesiano, Kim (1997) modificou as técnicas de tabela de vida e utiliza o modelo log-linear para incorporar as censuras. Goetghebeur e Ryan (2000) propuseram uma extensão dos modelos de riscos proporcionais para análise de dados com censura intervalar, Pan (2000) apresentou uma comparação entre estimadores de suavização em modelos semi-paramétricos, Fang, Sun e Lee (2002) modificaram os métodos de testes de comparação de curvas de sobrevivência para dados com censura intervalar, Finkelstein, Goggins e Schoenfeld (2002) apresentaram métodos para analisar dados com censura intervalar informativa e Wang (2006) considerou métodos estatísticos para tratar de censura intervalar no caso multivariado com o objetivo de avaliar a associação entre os eventos.

Recentemente, tem crescido a modelagem paramétrica, Odell, Anderson e D'Agostino (1992) utilizaram a distribuição Weibull para modelos de regressão com censura intervalar, Sun (1997) propôs um método baseado em modelos logísticos discretos, Lindsey (1998) comparou alguns

modelos de regressão paramétrica para modelar dados com censura intervalar, Langohr, Gómez e Muga (2004) apresentaram um modelo para dados duplamente censurados e Sparling, Younes e Lachin (2006) consideraram uma família paramétrica que acomoda efeitos fixos e covariáveis tempo dependentes.

A abordagem paramétrica é baseada em métodos de máxima verossimilhança e uma distribuição específica para os tempos de vida é considerada. As distribuições usuais para o tempo de falha são as distribuições Exponencial, Weibull e Gama generalizada que acomodam diferentes formas da função de taxa de falha. Entretanto, é comum encontrar na prática dados de sobrevivência com função de taxa de falha na forma de banheira e unimodal. Mudholkar, Srivastava, Kollia (1996) apresentaram a distribuição Weibull-exponenciada que tem como atrativo a função de taxa de falha que acomoda não apenas função risco crescente, decrescente ou constante, como também, a forma de banheira e unimodal. A distribuição Weibull-exponenciada também é utilizada para discriminar modelos, pois possui como casos particulares, a distribuição Exponencial, Weibull e a Exponencial-exponenciada, Rayleigh e Rayleigh generalizada.

Neste trabalho propõe-se um modelo de regressão log-linear para modelar dados com censura intervalar utilizando a distribuição Weibull-exponenciada. Esse novo modelo de regressão apresenta como casos particulares outros modelos de regressão, tais como regressão Weibull, Exponencial-exponenciada, Rayleigh generalizada, entre outros. Assim uma vantagem desse modelo de regressão Weibull-exponenciado com censura intervalar é que ela pode ser utilizada para discriminar modelos de regressão.

Segundo Fachini (2007), quando se ajusta um modelo a um conjunto de dados é imprescindível que as estimativas obtidas a partir do modelo proposto sejam 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.

Assim, é necessário verificar as suposições dos erros, a presença de possíveis pontos influentes e a validação do ajuste do modelo adotado. Para essa etapa, métodos de análise de sensibilidade e resíduos são ferramentas essenciais para detectar anomalias do modelo ajustado. Há várias metodologias que descrevem métodos de diagnóstico para dados com censura à direita, por exemplo, Cox e Snell (1968) e Collett (2003) apresentaram resíduos para dados sob o mecanismo de censura à direita. No contexto de análise de sensibilidade, Pettitt e Bin Daud (1989), Escobar e Meeker (1992) descreveram medidas de influência em dados de análise de sobrevivência, Ortega, Cancho e Bolfarine (2006) e Carrasco, Ortega e Paula (2008) utilizaram métodos de influência local e global em distribuições que modelam diferentes formas da função de taxa de falha. Silva et al.(2008) utilizaram as técnicas de análise de sensibilidade em um modelo com função de taxa de falha unimodal. Tais métodos podem ser estendidos à dados com censura intervalar, assim como Farrington (2000) modificou os resíduos usuais em análise de sobrevivência para dados com censura

intervalar.

Dessa forma, o objetivo desse trabalho é apresentar um modelo de regressão utilizando a distribuição Weibull-exponenciada e descrever os métodos de diagnósticos e resíduos para dados com censura intervalar. Para apresentar um roteiro que permita a um pesquisador aplicar o modelo de regressão com a distribuição Weibull-exponenciada para dados com censura intervalar articulada com a análise de diagnósticos e resíduos, a estrutura do presente trabalho é da seguinte maneira. Na seção 2.2, o modelo de regressão Weibull-exponenciada modificada para acomodar censura intervalar e a parte inferencial do modelo proposto são descritos. Na seção 2.3, descreve-se a análise de sensibilidade do modelo proposto fundamentadas nas teorias de influência global, influência local e influência local total. Na seção 2.4, os resíduos modificados para dados com censura intervalar do modelo proposto são estudados e um estudo de simulação é realizado. Na seção 2.5, uma aplicação da metodologia proposta neste estudo é realizada e finalmente, na seção 2.6 encontra-se as principais conclusões deste trabalho.

2.2 Modelo de regressão Weibull-exponenciada para dados de censura intervalar

A distribuição Weibull-exponenciada introduzida por Mudholkar e Srivastava (1993) é uma simples generalização da distribuição Weibull e é obtida por meio de um acréscimo de um parâmetro de forma na função de sobrevivência da distribuição Weibull. Uma das características principais do modelo consiste no fato de representar um família paramétrica que possui como casos particulares distribuições conhecidas e utilizadas na análise de dados, em que a variável resposta é o tempo até a ocorrência de um evento de interesse.

Assim, seja T uma variável aleatória contínua e não negativa com distribuição Weibull - exponenciada com vetor de parâmetros $\theta = (\alpha, \gamma, \lambda)^T$, na qual a função densidade de probabilidade é dada por

$$f(t; \alpha, \gamma, \lambda) = \frac{\alpha\lambda}{\gamma} \left\{ 1 - \exp \left[- \left(\frac{t}{\gamma} \right)^\alpha \right] \right\}^{\lambda-1} \exp \left[- \left(\frac{t}{\gamma} \right)^\alpha \right] \left(\frac{t}{\gamma} \right)^{\alpha-1}, \quad -\infty < y < \infty, \quad (2.1)$$

e a função de sobrevivência é dada por

$$S(t; \alpha, \gamma, \lambda) = 1 - \left\{ 1 - \exp \left[- \left(\frac{t}{\gamma} \right)^\alpha \right] \right\}^\lambda, \quad (2.2)$$

em que $\alpha > 0$ e $\lambda > 0$ são os parâmetros de forma e $\gamma > 0$ é o parâmetro de escala. A notação utilizada é $T \sim \text{WE}(\alpha, \gamma, \lambda)$.

Observa-se que a função de taxa de falha é dada por $h(t; \alpha, \gamma, \lambda) = f(t; \alpha, \gamma, \lambda)/S(t; \alpha, \gamma, \lambda)$. A grande flexibilidade desse modelo para ajustar dados de tempo de vida é dada através de diferentes formas que a função de taxa de falha assume, por exemplo, (i) se $\alpha \geq 1$ e $\alpha\lambda \geq 1$, então a função

de taxa de falha é monótona crescente; (ii) se $\alpha \leq 1$ e $\alpha\lambda \leq 1$, então a função de falha é monótona decrescente; (iii) se $\alpha > 1$ e $\alpha\lambda < 1$, então a função de taxa de falha é da forma de banheira e (iv) se $\alpha < 1$ e $\alpha\lambda > 1$, então tem-se a função de taxa de falha unimodal. A Figura 2.1 mostra o comportamento da função de taxa de falha da distribuição Weibull-exponenciada para diferentes valores dos parâmetros.

Figura 2.1 – Função de taxa de falha para a distribuição Weibull-exponenciada para diferentes valores de α , γ e λ

Entretanto, na prática, os tempos de falha observados são influenciados por covariáveis que, por exemplo, na área médica são: a idade, a contagem de célula, o efeito de tratamento, a altura, etc.

A forma mais comum de acomodar os efeitos dessas covariáveis nas análises é utilizar modelos de regressão apropriados para dados censurados. No contexto de sobrevivência, uma forma de incluir as covariáveis na análise é feita por meio da classe de modelos locação e escala. A abordagem dessa classe de modelos consiste em utilizar a transformação logarítmica nos tempos de falha de tal forma que para dado um vetor de covariáveis o logaritmo do tempo de falha tem uma distribuição com um parâmetro de locação μ e um parâmetro de escala σ .

Seja T uma variável aleatória com distribuição Weibull-exponenciada definida na expressão (2.1) e considerando $Y = \log(T)$ e as reparametrizações $\alpha = 1/\sigma$ e $\gamma = \exp(\mu)$, sugerida por Cancho (1999), a função densidade de probabilidade da variável aleatória Y é escrita como

$$f(y) = \frac{\lambda}{\sigma} \left\{ 1 - \exp \left[- \exp \left(\frac{y - \mu}{\sigma} \right) \right] \right\}^{\lambda-1} \exp \left\{ \left(\frac{y - \mu}{\sigma} \right) - \exp \left[\frac{y - \mu}{\sigma} \right] \right\}, \quad (2.3)$$

em que $-\infty < y < \infty$, $\sigma > 0$ e $\lambda > 0$. Assim, Y tem distribuição log-Weibull-exponenciada com notação $Y \sim (\mu, \sigma, \lambda)$.

Logo, considerando o modelo de regressão baseado na distribuição log-Weibull-exponenciada descrita em (2.3) e relacionando a variável Y a um vetor de covariáveis $\mathbf{x}$, tem-se

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n, \quad (2.4)$$

em que y_i é o logaritmo dos tempos de sobrevivência para o i -ésimo indivíduo, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ é o vetor de covariáveis associado ao i -ésimo indivíduo, $\sigma > 0$ é o parâmetro de escala desconhecido e z_i é uma variável aleatória que possui uma distribuição log-Weibull-exponenciada com a função densidade de probabilidade

$$f(z) = \lambda \left[1 - \exp \{ - \exp(z) \} \right]^{\lambda-1} \exp \{ z - \exp(z) \}, \quad -\infty < z < \infty, \quad (2.5)$$

em que $\lambda > 0$ é o parâmetro de forma desconhecido.

A Figura (2.2) mostra a função densidade de probabilidade da variável aleatória Z para diferentes escolhas do parâmetro de forma λ .

Figura 2.2 – Gráfico da função densidade de probabilidade da variável aleatória Z para diferentes valores de λ

Teorema

Para a variável Z , o k -ésimo momento ordinário é dado por

$$\mu'_k = E(Z^k) = \lambda \Gamma(\lambda) \sum_{j=0}^{\infty} \frac{(-1)^j}{\Gamma(\lambda - j)j!} \times \left. \frac{\partial^k [(j+1)^{-\lambda} \Gamma(\lambda)]}{\partial \lambda^k} \right|_{\lambda=1},$$

em que $\Gamma(\cdot)$ é a função gama. A demonstração é dada no apêndice A1.

A função de densidade de probabilidade e sobrevivência de $Y_i|\mathbf{x}$ assume as formas

$$f(y_i|\mathbf{x}) = \frac{\lambda}{\sigma} \left\{ 1 - \exp \left[- \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}^{\lambda-1} \\ \times \exp \left\{ \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) - \exp \left[\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right] \right\}, \quad -\infty < y < \infty,$$

e

$$S(y_i|\mathbf{x}_i) = 1 - \left\{ 1 - \exp \left[- \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}^{\lambda}, \quad (2.6)$$

respectivamente.

Para alguns valores dos parâmetros de $\boldsymbol{\theta} = (\lambda, \sigma, \boldsymbol{\beta})^T$, o modelo de regressão Weibull-exponenciada, reduz-se aos seguintes modelos

Modelo de regressão Exponencial

Quando $\lambda = 1$ e $\sigma = 1$ são substituídos na expressão (2.6), tem-se o modelo de regressão Exponencial, cuja função de sobrevivência é dada por

$$S(y|\mathbf{x}_i) = \exp \left\{ - \exp (y - \mathbf{x}_i^T \boldsymbol{\beta}) \right\}.$$

Modelo de regressão Weibull

Substituindo $\lambda = 1$ na expressão (2.6), tem-se o modelo regressão Weibull, com função de sobrevivência dada por

$$S(y|\mathbf{x}_i) = \exp \left\{ - \exp \left(\frac{y - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right\}.$$

Modelo de regressão Exponencial-exponenciada

Substituindo $\sigma = 1$ na expressão (2.6), tem-se o modelo de regressão Exponencial-exponenciada, cuja função de sobrevivência tem a forma

$$S(y|\mathbf{x}_i) = 1 - \left[1 - \exp \left\{ - \exp(y - \mathbf{x}_i^T \boldsymbol{\beta}) \right\} \right]^{\lambda}.$$

Modelo de regressão Rayleigh

Substituindo $\lambda = 1$ e $\sigma = 1/2$ na expressão (2.6), tem-se o modelo de regressão Rayleigh, na qual a função de sobrevivência tem a forma

$$S(y|\mathbf{x}_i) = 1 - \left[1 - \exp \left\{ - \exp(2(y - \mathbf{x}_i^T \boldsymbol{\beta})) \right\} \right]^{\lambda}.$$

Modelo de regressão Rayleigh generalizada

Substituindo $\sigma = 1/2$ na expressão (2.6), tem-se o modelo de regressão Rayleigh generalizada, com função de sobrevivência escrita como

$$S(y|\mathbf{x}_i) = 1 - \left[1 - \exp \left\{ - \exp(2(y - \mathbf{x}_i^T \boldsymbol{\beta})) \right\} \right]^\lambda.$$

Na presença de censura intervalar, os dados observados consistem de um intervalo $(\log(u_i), \log(v_i))$ para cada indivíduo, no qual esses intervalos são conhecidos e inclui y_i com probabilidade um, ou seja, $P(\log(u_i) \leq y_i \leq \log(v_i)) = 1$, e se $\log(v_i) = \infty$, então o logaritmo do tempo é censurado à direita para y_i . Este modelo será referido como modelo de regressão Weibull-exponenciada com censura intervalar, que é uma extensão de modelos de tempo de falha acelerado utilizando a distribuição Weibull-exponenciada para dados de censura intervalar.

2.2.1 Estimador de máxima verossimilhança para o modelo com censura intervalar

Sejam $(\log(u_1), \log(v_1), \mathbf{x}_1), \dots, (\log(u_n), \log(v_n), \mathbf{x}_n)$, n observações independentes de uma amostra aleatória, na qual $(\log(u_i), \log(v_i))$ representa o intervalo de tempo em que o logaritmo da falha ocorreu e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ é o vetor de covariáveis associado ao i -ésimo indivíduo. O logaritmo da função de verossimilhança para o modelo de regressão Weibull-exponenciada com vetor de parâmetros $\boldsymbol{\theta} = (\lambda, \sigma, \boldsymbol{\beta}^T)^T$ é dado da seguinte forma

$$l(\boldsymbol{\theta}) = \sum_{i \in F} l_1(\lambda, zu_i, zv_i) + \sum_{i \in C} l_2(\lambda, zu_i), \quad (2.7)$$

em que

$$l_1(\lambda, zu_i, zv_i) = \log \left[\left\{ 1 - \exp \left[- \exp(zv_i) \right] \right\}^\lambda - \left\{ 1 - \exp \left[- \exp(zu_i) \right] \right\}^\lambda \right],$$

$$l_2(\lambda, zu_i) = \log \left[1 - \left\{ 1 - \exp \left[- \exp(zu_i) \right] \right\}^\lambda \right],$$

$zu_i = (\log(u_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $zv_i = (\log(v_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, F é o conjunto de indivíduos que apresentam censura intervalar e C é o conjunto de indivíduos censurados à direita.

Os estimadores de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta} = (\lambda, \sigma, \boldsymbol{\beta}^T)^T$ são obtidos maximizando-se a expressão (2.7). O que equivale a resolver o seguinte sistema de equações

não lineares,

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} &= \sum_{i \in F} \left[\frac{(b_i)_\lambda - (a_i)_\lambda}{(1 - b_i)^\lambda - (1 - a_i)^\lambda} \right] - \sum_{i \in C} \left[\frac{(a_i)_\lambda}{1 - (1 - a_i)^\lambda} \right] = 0, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} &= \sum_{i \in F} \left[\frac{\lambda \sigma^{-1} (zu_i g_i - zv_i h_i)}{(1 - b_i)^\lambda - (1 - a_i)^\lambda} \right] + \sum_{i \in C} \left[\frac{\lambda \sigma^{-1} zu_i g_i}{1 - (1 - a_i)^\lambda} \right] = 0, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i \in F} \left[\frac{\lambda \sigma^{-1} x_{ij} (g_i - h_i)}{(1 - b_i)^\lambda - (1 - a_i)^\lambda} \right] + \sum_{i \in C} \left[\frac{\lambda \sigma^{-1} x_{ij} g_i}{1 - (1 - a_i)^\lambda} \right] = 0, \quad j = 1, 2, \dots, p,\end{aligned}$$

em que $zu_i = (\log(u_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $zv_i = (\log(v_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $a_i = \exp[-\exp(zu_i)]$,

$$b_i = \exp[-\exp(zv_i)], \quad (a_i)_\lambda = (1 - a_i)^\lambda \log(1 - a_i), \quad (b_i)_\lambda = (1 - b_i)^\lambda \log(1 - b_i),$$

$$g_i = (1 - a_i)^{\lambda-1} a_i \exp(zu_i) \quad \text{e} \quad h_i = (1 - b_i)^{\lambda-1} b_i \exp(zv_i).$$

Porém, as equações $\frac{\partial l(\boldsymbol{\theta})}{\partial \lambda} = 0$, $\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = 0$ e $\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} = 0$ são não lineares e, em geral, não possuem soluções exatas. Soluções aproximadas são encontradas por métodos iterativos, por exemplo, Newton-Raphson ou um algoritmo de Quase-Newton. Os métodos iterativos para iniciar o algoritmo especificam uma estimativa inicial $\boldsymbol{\theta}^{(0)}$, iterá-la sucessivamente até obter convergência.

Neste trabalho, a linguagem matricial de programação Ox (DOORNIK, 2007) é utilizada para calcular as estimativas de máxima verossimilhança por meio da sub-rotina MAXBFGS (ver apêndice A2). Estimativas da matriz de covariância do estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ são obtidas usando a matriz Hessiana. Intervalos de confiança e testes de hipóteses são conduzidas utilizando a distribuição assintótica dos estimadores de máxima verossimilhança, que é uma distribuição normal com a matriz de covariância obtida como a inversa da matriz de informação de Fisher sob certas condições de regularidade satisfeitas. Mais especificamente, a matriz de covariância assintótica é dada por $\mathbf{I}^{-1}(\boldsymbol{\theta})$ em que $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ tal que $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$.

Como não é possível determinar a matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\theta})$ devido a presença de observações censuradas, uma forma alternativa é utilizar a matriz de informação observada $\ddot{\mathbf{L}}(\boldsymbol{\theta})$, avaliada no estimador de máxima verossimilhança $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. A aproximação normal assintótica de $\hat{\boldsymbol{\theta}}$ é expressa como

$$\hat{\boldsymbol{\theta}}^T \sim N_{(p+2)}\{\boldsymbol{\theta}^T; \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\},$$

em que $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ é a matriz de informação observada $(p+2)(p+2)$ dada por

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{\lambda\lambda} & \mathbf{L}_{\lambda\sigma} & \mathbf{L}_{\lambda\beta_j} \\ \cdot & \mathbf{L}_{\sigma\sigma} & \mathbf{L}_{\sigma\beta_j} \\ \cdot & \cdot & \mathbf{L}_{\beta_j\beta_s} \end{pmatrix}$$

com as correspondentes submatrizes descrita no apêndice A3.

Assim, o intervalo de confiança assintótico para β_j , com $j = 1, 2, \dots, p$, considerando um nível de confiança de $(1 - \alpha)100\%$, é expresso por

$$\hat{\beta}_j \pm z_{\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\beta}_j)},$$

em que $z_{\alpha/2}$ é o valor que excede com probabilidade $\alpha/2$ a distribuição normal padrão.

2.2.2 Testes de hipóteses

A discriminação de modelos por meio de testes de hipóteses baseados em modelos gerais são bem aceitos na literatura, uma vez que permitem a conclusão direta e não envolvem qualquer componente subjetivo na sua interpretação.

Para testar a adequabilidade de um modelo, as hipóteses de interesse são da seguinte forma

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \quad (2.8)$$

em que $\boldsymbol{\theta}_0$ é o vetor de parâmetros a ser testado.

O teste de hipóteses é feito utilizando a estatística da razão de verossimilhanças em modelos encaixados (COX; HINKLEY, 1974). O teste da razão de verossimilhanças consiste em comparar os valores do logaritmo da função de verossimilhança maximizada do modelo geral e do modelo reduzido sob a hipótese nula H_0 . A partir desses valores, a estatística da razão de verossimilhanças é calculada por

$$\Lambda_n = -2[\ell(\tilde{\lambda}, \tilde{\sigma}, \tilde{\boldsymbol{\beta}}) - \ell(\hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}})] \quad (2.9)$$

em que $\ell(\hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}})$ denota o logaritmo da função de verossimilhança do modelo geral e $\ell(\tilde{\lambda}, \tilde{\sigma}, \tilde{\boldsymbol{\beta}})$ é o logaritmo da função de verossimilhança restrita calculada sob H_0 .

A estatística Λ_n sob H_0 tem aproximadamente uma distribuição qui-quadrado com k graus de liberdade que corresponde a diferença do número de parâmetros dos modelos a serem comparados. Rejeita-se a hipótese nula (H_0) a um nível de $100\alpha\%$ de significância quando $\Lambda_n > \chi_{1-\alpha}(k)$, em que $\chi_{1-\alpha}(k)$ é o quantil $(1 - \alpha)$ da distribuição qui-quadrado com k graus de liberdade.

Nesse contexto, percebe-se a importância do modelo de regressão Weibull-exponenciada com censura intervalar, uma vez que o modelo pode ser utilizado para discriminar modelos. Por exemplo, para testar a adequabilidade do modelo de regressão Weibull para dados de censura intervalar, as hipóteses de interesse (2.8) são dadas por

$$H_0 : \lambda = 1 \quad \text{versus} \quad H_1 : \lambda \neq 1,$$

e a estatística da razão de verossimilhanças, no qual, neste caso, é dado por

$$\Lambda_n = -2[\ell(1, \tilde{\sigma}, \tilde{\boldsymbol{\beta}}) - \ell(\hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}})]$$

em que $\hat{\lambda}$, $\hat{\sigma}$ e $\hat{\beta}$ são os estimadores de máxima verossimilhança que segue através da maximização do logaritmo da função de verossimilhança em (2.7) e $\tilde{\sigma}$ e $\tilde{\beta}$ são os estimadores de máxima verossimilhança restrita calculada sob H_0 , ou seja, com $\lambda = 1$.

Para testar a adequabilidade do modelo de regressão Exponencial para dados de censura intervalar, ou seja, $H_0 : (\sigma, \lambda)^\top = (1, 1)^\top$, segue o procedimento análogo ao teste do modelo de regressão Weibull para dados de censura intervalar.

2.2.3 Estimador Jackknife

A idéia do estimador Jackknife é transformar o problema de estimação de qualquer parâmetro populacional dentro de um problema de estimar uma média populacional. Assim, o que é feito quando estima-se o valor da média é realizado nesse método, mas de um ponto de vista não usual. Um importante trabalho de implementação do método Jackknife é dado por Lipsitz, Laird e Harrington (1990) que sugere um estimador robusto como alternativa da matriz de covariância baseada no Jackknife para análise de dados provenientes de estudos de medidas repetidas. Neste trabalho, esses métodos são utilizados como uma alternativa para estimar os parâmetros populacionais. Sejam $Y_1, \dots, Y_n$ uma amostra aleatória de tamanho n e a média amostral dada por

$$\bar{Y} = \sum_{i=1}^n \frac{Y_i}{n},$$

utilizada para estimar a média populacional.

Para, o método Jackknife, calcula-se a média amostral sem a l -ésima observação,

$$\bar{Y}_{-l} = \frac{\sum_{i=1}^n Y_i - Y_l}{n - 1}. \quad (2.10)$$

Então, através da equação (2.10) obtém-se

$$Y_l = n\hat{Y} - (n - 1)\bar{Y}_{-l}. \quad (2.11)$$

Em uma situação geral, considere que θ é um parâmetro estimado por meio da $\hat{E}(Y_1, \dots, Y_n)$ e para simplificar a notação foi omitido $(Y_1, \dots, Y_n)$. Finalmente, remove-se Y_l e obtém-se o estimador parcial $\hat{E}_{-l}$. Através da analogia com a equação (2.11) existe um conjunto de pseudo-valores que são calculados por

$$\hat{E}_l^* = n\hat{E} - (n - 1)\hat{E}_{-l}, \quad l = 1, \dots, n.$$

A média dos pseudo-valores dada por

$$\hat{E}^* = \frac{\sum_{l=1}^n \hat{E}_l^*}{n}$$

é o estimador Jackknife de θ .

Manly (1997) sugere que um intervalo de confiança aproximado ao nível de $100(1 - \alpha)\%$ para θ é dado por

$$\hat{E}^* \pm t_{\alpha/2, n-1} s / \sqrt{n},$$

em que $t_{\alpha/2, n-1}$ é o valor que excede com probabilidade $\alpha/2$ a distribuição t com $(n - 1)$ graus de liberdade e o estimador Jackknife tem a vantagem de remover o vício de ordem $1/n$.

As estimativas Jackknife calculadas para o modelo de regressão Weibull-exponenciada para dados de censura intervalar são feitos para λ , σ e β_j ($j = 1, \dots, p$) e intervalos de confiança são obtidos separadamente para cada parâmetro.

2.3 Análise de sensibilidade

Após a modelagem, é importante verificar as suposições feitas para o modelo e conduzir um estudo da robustez para encontrar possíveis observações influentes que podem causar distorções nos resultados da análise. Cook (1977) propôs medidas de diagnóstico de influência baseadas na deleção de casos, na qual a influência do i -ésimo indivíduo sobre as estimativas dos parâmetros é estudada por meio da retirada de casos da análise. Esta metodologia tem sido aplicada em vários modelos estatísticos, por exemplo, Chirstensen, Pearson e Johnson (1992), Davison e Tsai (1992), Xie e Wei(2007), entre outros.

Entretanto, quando a deleção de casos é utilizada todas as informações de um único indivíduo é deletado imediatamente e portanto é difícil dizer se aquele indivíduo tem alguma influência sobre um aspecto específico do modelo. Uma solução para o problema é encontrada na abordagem de influência local, a qual verifica como os resultados das análises são alterados quando pequenas perturbações são introduzidas no modelo ou no conjunto de dados. Cook (1986) propôs um trabalho para detectar observações influentes que indicam a sensibilidade no modelo quando pequenas perturbações são provocadas. Alguns autores tem investigado a avaliação de influencia local em análise de sobrevivência. Pettitt e Bin Daud (1989) investigaram influência local em modelos de regressão de riscos proporcionais, Escobar e Meeker (1992) adaptaram os métodos de influência local para análise de regressão com censura, Ortega, Bolfarine e Paula (2003) consideraram o problema de influencia local em modelos de regressão log-gama generalizada com observações censuradas, mais recentemente, Ortega, Cancho e Paula (2008) derivaram os cálculos de curvatura sob vários esquemas de perturbação de casos em modelos de regressão com fração de cura, Fachini, Ortega e Louzada-Neto(2008) adaptaram os métodos de influência local para modelos de riscos múltiplos.

Nesta seção, a metodologia de influência global é apresentada considerando a distância de Cook generalizada e o afastamento da verossimilhança como medidas de influência. Também é estudada

a metodologia de influência local total e influência local para verificar possíveis observações influentes para o modelo de regressão Weibull-exponenciada com censura intervalar.

2.3.1 Influência global

A primeira ferramenta para avaliar a análise de sensibilidade do modelo proposto são as medidas de influência global partindo da idéia de deleção de casos proposta por Cook (1977), a qual o enfoque é estudar o efeito de retirar o i -ésimo indivíduo, denominada de caso, de um conjunto de dados. A deleção de casos para o modelo (2.4) é dado por

$$y_{(i)} = \mathbf{x}_{(i)}^T \boldsymbol{\beta} + \sigma z_{(i)}, \quad i = 1, \dots, n, \quad (2.12)$$

em que o subscrito “(i)” significa que o i -ésimo caso foi retirado da amostra. Para o modelo (2.12), o logaritmo da função de verossimilhança é denotado por $l_{(i)}(\boldsymbol{\theta})$.

Para $Y = \log(T)$ uma variável aleatória contínua com distribuição log-Weibull-exponenciada (2.6), o logaritmo da função de verossimilhança com parâmetro $\boldsymbol{\theta}$ é denotado por $l(\boldsymbol{\theta})$. Seja $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\lambda}_{(i)}, \hat{\sigma}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}^T)^T$ o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ para $l_{(i)}(\boldsymbol{\theta})$. A influência do i -ésimo caso nas estimativas de máxima verossimilhança de $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}}^T)^T$ é avaliada por meio da diferença entre $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}$.

Se a deleção de um caso provoca variações nas estimativas, maiores atenções devem ser dadas para aquele caso. Conseqüentemente, se $\hat{\boldsymbol{\theta}}_{(i)}$ está distante de $\hat{\boldsymbol{\theta}}$, então o caso é observado como uma observação influente. A primeira medida de influência global é definida como a norma padronizada de $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$, conhecida como distância de Cook generalizada, expressa por

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T \{ -\ddot{\mathbf{L}}(\boldsymbol{\theta}) \} (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Outra medida popular para verificar a existência de pontos influentes no modelo é definida em termos da diferença entre $l_{(i)}(\boldsymbol{\theta})$ e $l(\hat{\boldsymbol{\theta}})$. A medida conhecida como afastamento da verossimilhança é dada por

$$LD_i(\boldsymbol{\theta}) = 2 \left\{ l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(i)}) \right\}.$$

2.3.2 Influência local

O método de influência local introduzida por Cook(1986), objetiva estudar a parte de diagnóstico de um modelo, por meio de pequenas perturbações no modelo ou nos dados, ao invés de remover uma observação. Segundo Paula(2004) a idéia básica do método consiste em verificar a existência de pontos que sob modestas modificações no modelo causam variações desproporcionais nos resultados.

Seja $l(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança correspondente ao modelo postulado, em que $\boldsymbol{\theta}$ é o vetor de parâmetros desconhecidos $p \times 1$ e $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ o logaritmo da função de verossimilhança do modelo perturbado, em que $\boldsymbol{\omega}$ é o vetor de perturbação $q \times 1$. Suponha que existe um vetor $\boldsymbol{\omega}_0 \in \Omega$ tal que $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$ para todo o $\boldsymbol{\theta}$. Então, a medida sugerida por Cook(1986) é definida por

$$LD(\boldsymbol{\omega}) = 2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})], \quad (2.13)$$

em que $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denota a estimativa de máxima verossimilhança sob o modelo perturbado e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_q)^T$ é o vetor de perturbações aplicados ao modelo proposto. A expressão (2.13) é conhecida como afastamento da verossimilhança.

A proposta de influência local é estudar o comportamento da expressão (2.13) em torno do vetor de não perturbação $\boldsymbol{\omega}_0$ e selecionar uma direção unitária $\mathbf{d}$ que determina o plano normal de Ω e contém o eixo da função $LD(\boldsymbol{\omega})$. Logo, desde que $LD(\boldsymbol{\omega}) \geq 0$, $\boldsymbol{\omega}_0$ é um ponto de mínimo da expressão (2.13). A sugestão de Cook (1986) é investigar a curvatura normal da linha projetada $LD(\boldsymbol{\omega}_0 + a\mathbf{d})$, em que $a \in \mathbb{R}$, em torno de $a = 0$ para alguma direção arbitrária $\mathbf{d}$. A curvatura normal de $\boldsymbol{\theta}$ na direção de $\mathbf{d}$ para o modelo proposto é dado por

$$C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^T \Delta [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \Delta \mathbf{d}|,$$

em que Δ é uma matriz $(p+2 \times n)$ dependendo do esquema de perturbação e seus elementos são dados por $\Delta_{ij} = \partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \theta_j \partial \omega_i$, $i = 1, \dots, n$ e $j = 1, \dots, p+2$, avaliada em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ e $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0$.

Cook (1986) sugere tomar a direção correspondente à maior curvatura, denotada por $\mathbf{d}_{\max}$, que é o maior autovetor e o correspondente maior autovalor $C_{\mathbf{d}_{\max}}$, da matriz $\ddot{F} = \Delta^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \Delta$.

Por outro lado, Lesaffre e Verbeke (1998) sugerem considerar a direção do i -ésimo indivíduo que corresponde ao vetor $\mathbf{d}_i = (0, \dots, 1, \dots, 0)^T$, tal que o i -ésimo elemento recebe o valor um. Nesse caso, a curvatura normal chamada de influência local total do i -ésimo indivíduo é dada por

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

em que Δ_i denota a i -ésima linha da matriz Δ . Os possíveis pontos influentes são aqueles na qual $C_i > 2\bar{C}$, em que $\bar{C} = \sum_{i=1}^n C_i / n$.

Assim, os gráficos dos elementos de $\mathbf{d}_{\max}$ *versus* o índice das observações bem como os gráficos dos elementos C_i *versus* o índice das observações são utilizadas para detectar possíveis pontos influentes.

2.3.3 Cálculo das curvaturas máximas

As medidas de influência local e total são calculadas após um esquema de perturbação no modelo ou nos dados. Diferentes esquemas podem ser considerados de acordo com o objetivo principal

da análise, neste trabalho, considerou-se cinco esquemas de perturbação para os quais calcula-se a matriz de curvaturas máximas definida por

$$\Delta = (\Delta_{vi})_{(p+2) \times n} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_v \partial \omega_i} \right)_{(p+2) \times n},$$

em que $v = 1, \dots, p+2$ e $i = 1, \dots, n$.

Considerando o modelo definido em (2.4) e o logaritmo da função de verossimilhança expresso em (2.7), a seguir apresenta-se os esquemas de perturbação no modelo de regressão Weibull-exponenciada com censura intervalar. Assume-se também o vetor de perturbação denotado por $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$, os pesos que perturbam o modelo ou os dados.

Perturbação de casos

O logaritmo da função de verossimilhança perturbada incorpora diferentes pesos para cada elemento do conjunto de dados e é definido por

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i \in F} \omega_i l_1(\lambda, zu_i, zv_i) + \sum_{i \in C} \omega_i l_2(\lambda, zu_i),$$

em que $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ é o vetor de não perturbação, $l_1(\cdot)$ e $l_2(\cdot)$ são definidos na expressão (2.7).

Os elementos da matriz $\Delta = (\Delta_\lambda^T, \Delta_\sigma^T, \Delta_\beta^T)^T$ são expressos da seguinte maneira. Os elementos do vetor Δ_λ são dados por

$$\Delta_i = \begin{cases} \frac{(\hat{b}_i)_\lambda - (\hat{a}_i)_\lambda}{(1-\hat{b}_i)^\lambda - (1-\hat{a}_i)^\lambda} & \text{se } i \in F, \\ \frac{(\hat{a}_i)_\lambda}{1-(1-\hat{a}_i)^\lambda} & \text{se } i \in C, \end{cases}$$

em que

$$(\hat{a}_i)_\lambda = (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma},$$

$$(\hat{b}_i)_\lambda = (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}.$$

Os elementos do vetor Δ_σ são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-1} (\hat{z}u_i \hat{g}_i - \hat{z}v_i \hat{h}_i)}{(1-\hat{b}_i)^\lambda - (1-\hat{a}_i)^\lambda} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-1} \hat{z}u_i \hat{g}_i}{1-(1-\hat{a}_i)^\lambda} & \text{se } i \in C, \end{cases}$$

em que

$$\hat{g}_i = (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma},$$

$$\hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (v_i - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}.$$

Os elementos da matriz Δ_{β} , para $j = 1, \dots, p$, são dados por

$$\Delta_{ji} = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-1} x_{ij} (\hat{g}_i - \hat{h}_i)}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-1} x_{ij} \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned} \hat{g}_i &= (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}, \\ \hat{h}_i &= (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}. \end{aligned}$$

Perturbação do limite inferior da variável resposta ($\log(u_i)$)

O interesse aqui é detectar a sensibilidade do modelo quando o limite inferior ($\log(u_i)$) da variável resposta é submetida a uma perturbação aditiva $u_{iw} = \log(u_i) + \omega_i S_u$, em que S_u é um fator escala que pode ser a estimativa do desvio padrão da variável $\log(U)$ e $\omega_i \in \mathbf{R}$.

Neste caso, o logaritmo da função de verossimilhança perturbada é da forma

$$l(\theta|\omega) = \sum_{i \in F} l_1(\lambda, zu_i^*, zv_i) + \sum_{i \in C} l_2(\lambda, zu_i^*),$$

em que $(zu_i^*) = (u_{iw} - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}$, $\omega_0 = (0, \dots, 0)^T$ é o vetor de não perturbação, $l_1(\cdot)$ e $l_2(\cdot)$ são definidos em (2.7).

Os elementos da matriz $\Delta = (\Delta_{\lambda}^T, \Delta_{\sigma}^T, \Delta_{\beta}^T)^T$ são expressos da seguinte maneira. Os elementos do vetor Δ_{λ} são dados por

$$\Delta_i = \begin{cases} \frac{-S_u \hat{\sigma}^{-1} \hat{g}_i \hat{p}_i}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} + \frac{[(\hat{b}_i)_{\lambda} - (\hat{a}_i)_{\lambda}](\hat{\lambda} \hat{\sigma}^{-1} S_u \hat{g}_i)}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ \frac{-\hat{\sigma}^{-1} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \left[\hat{p}_i + \frac{\hat{\lambda} (\hat{a}_i)_{\lambda}}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned} (\hat{a}_i)_{\lambda} &= (1 - \hat{a}_i)^{\hat{\lambda}} \log(1 - \hat{a}_i), \quad \hat{g}_i = (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{p}_i = \hat{\lambda} \log(1 - \hat{a}_i) + 1, \\ \hat{a}_i &= \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}, \quad (\hat{b}_i)_{\lambda} = (1 - \hat{b}_i)^{\hat{\lambda}} \log(1 - \hat{b}_i), \\ \hat{b}_i &= \exp[-\exp(\hat{z}v_i)], \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\beta})/\hat{\sigma}. \end{aligned}$$

Os elementos do vetor Δ_{σ} são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} S_u \hat{g}_i (\hat{a}_i)_{\sigma\beta}}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} + \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} S_u \hat{g}_i (\hat{z}u_i \hat{g}_i - \hat{z}v_i \hat{h}_i)}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-2} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \left[(\hat{a}_i)_{\sigma\beta} + \frac{\hat{\lambda} \hat{z}u_i \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned}
 (\hat{a}_i)_{\sigma\beta} &= 1 + (1 - \hat{a}_i)^{-1} \hat{z}u_i \hat{a}_i \exp(\hat{z}u_i)[\hat{\lambda} - 1] + \hat{z}u_i[1 - \exp(\hat{z}u_i)], \quad \hat{g}_i = (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i), \\
 \hat{a}_i &= \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}, \\
 (\hat{b}_i)_{\sigma\beta} &= 1 + (1 - \hat{b}_i)^{-1} \hat{z}v_i \hat{b}_i \exp(\hat{z}v_i)[\hat{\lambda} - 1] + \hat{z}v_i[1 - \exp(\hat{z}v_i)], \quad \hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i), \\
 \hat{b}_i &= \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}.
 \end{aligned}$$

Os elementos da matriz $\Delta_{\boldsymbol{\beta}}$, para $j = 1, \dots, p$, são dados por

$$\Delta_{ji} = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} x_{ij} S_u \hat{g}_i (\hat{a}_i)_{\beta}}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} + \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} x_{ij} S_u \hat{g}_i (\hat{g}_i - \hat{h}_i)}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-2} x_{ij} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \left[(\hat{a}_i)_{\beta} + \frac{\hat{\lambda} \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned}
 (\hat{a}_i)_{\beta} &= 1 + (1 - \hat{a}_i)^{-1} \hat{a}_i \exp(\hat{z}u_i)[\hat{\lambda} - 1] - \exp(\hat{z}u_i), \quad \hat{g}_i = (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i), \\
 \hat{a}_i &= \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}, \\
 (\hat{b}_i)_{\beta} &= 1 + (1 - \hat{b}_i)^{-1} \hat{b}_i \exp(\hat{z}v_i)[\hat{\lambda} - 1] - \exp(\hat{z}v_i), \quad \hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i), \\
 \hat{b}_i &= \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}.
 \end{aligned}$$

Perturbação do limite superior da variável resposta ($\log(v_i)$)

Considere que cada limite superior ($\log(v_i)$) do intervalo é perturbado como $v_{iw} = \log(v_i) + \omega_i S_v$, em que S_v é um fator escala que pode ser a estimativa do desvio padrão de $\log(V)$ e $\omega_i \in \mathbf{R}$.

O logaritmo da função de verossimilhança perturbada é expressado por

$$l(\boldsymbol{\theta} | \boldsymbol{\omega}) = \sum_{i \in F} l_1(\lambda, zu_i, zv_i^*) + \sum_{i \in C} l_2(\lambda, zu_i),$$

em que $zv_i^* = (v_{iw} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) / \hat{\sigma}$, $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$ é o vetor de não perturbação, $l_1(\cdot)$ e $l_2(\cdot)$ são definidas na expressão (2.7).

Os elementos da matriz $\Delta = (\Delta_{\lambda}^T, \Delta_{\sigma}^T, \Delta_{\boldsymbol{\beta}}^T)^T$ são expressos da seguinte maneira. Os elementos do vetor Δ_{λ} são dados por

$$\Delta_i = \begin{cases} \frac{S_v \hat{\sigma}^{-1} \hat{h}_i \hat{g}_i}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} - \frac{\hat{\lambda} \hat{\sigma}^{-1} S_v \hat{h}_i [(\hat{b}_i)_{\lambda} - (\hat{a}_i)_{\lambda}]}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ 0 & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned}(\hat{a}_i)_\lambda &= (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \\ (\hat{b}_i)_\lambda &= (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i), \quad \hat{h}_i = (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{q}_i = \hat{\lambda} \log(1 - \hat{b}_i) + 1, \\ \hat{b}_i &= \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.\end{aligned}$$

Os elementos do vetor Δ_σ são dados por

$$\Delta_i = \begin{cases} \frac{-\hat{\lambda} \hat{\sigma}^{-2} S_v \hat{h}_i (\hat{b}_i)_{\sigma\beta}}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} S_v \hat{h}_i (\hat{z}u_i \hat{g}_i - \hat{z}v_i \hat{h}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ 0 & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned}\hat{g}_i &= (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \\ (\hat{b}_i)_{\sigma\beta} &= 1 + (1 - \hat{b}_i)^{-1} \hat{z}v_i \hat{b}_i \exp(\hat{z}v_i) [\hat{\lambda} - 1] + \hat{z}v_i [1 - \exp(\hat{z}v_i)], \quad \hat{h}_i = (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \\ \hat{b}_i &= \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.\end{aligned}$$

Os elementos da matriz Δ_β , para $j = 1, \dots, p$, são dados por

$$\Delta_{ji} = \begin{cases} \frac{-\hat{\lambda} \hat{\sigma}^{-2} x_{ij} S_v \hat{h}_i (\hat{b}_i)_\beta}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} x_{ij} S_v \hat{h}_i (\hat{g}_i - \hat{h}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ 0 & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned}\hat{g}_i &= (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \\ (\hat{b}_i)_\beta &= 1 + (1 - \hat{b}_i)^{-1} \hat{b}_i \exp(\hat{z}v_i) [\hat{\lambda} - 1] - \exp(\hat{z}v_i), \quad \hat{h}_i = (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \\ \hat{b}_i &= \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.\end{aligned}$$

Perturbação da variável resposta intervalar $(\log(u_i), \log(v_i))$

Com o objetivo de detectar a sensibilidade do modelo quando cada intervalo, $\log(u_i)$, $\log(v_i)$ que contém a variável resposta y_i , é submetida a uma perturbação aditiva $u_{iw} = \log(u_i) + \omega_i S_u$ e $v_{iw} = \log(v_i) + \omega_i S_v$, respectivamente, em que S_u é um fator escala que pode ser a estimativa do desvio padrão da variável $\log(U)$, S_v é um fator escala que pode ser a estimativa do desvio padrão de $\log(V)$ e $\omega_i \in \mathbf{R}$.

Assim, o logaritmo da função de verossimilhança perturbada é definido como

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i \in F} l_1(\lambda, zu_i^*, zv_i^*) + \sum_{i \in C} l_2(\lambda, zu_i^*),$$

em que $zu_i^* = (u_{iw} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}$, $zv_i^* = (v_{iw} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}$, $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$ é o vetor de não perturbação, $l_1(\cdot)$ e $l_2(\cdot)$ são definidos na expressão (2.7).

Os elementos da matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\lambda^T, \boldsymbol{\Delta}_\sigma^T, \boldsymbol{\Delta}_\beta^T)^T$ são expressos da seguinte maneira. Os elementos do vetor $\boldsymbol{\Delta}_\lambda$ são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\sigma}^{-1}(S_v \hat{h}_i \hat{q}_i - S_u \hat{g}_i \hat{p}_i)}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda} \hat{\sigma}^{-1}[(\hat{b}_i)_\lambda - (\hat{a}_i)_\lambda](S_v \hat{h}_i - S_u \hat{g}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ \frac{-\hat{\sigma}^{-1} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \left[\hat{p}_i + \frac{\hat{\lambda} (\hat{a}_i)_\lambda}{1 - (1 - \hat{a}_i)^\lambda} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned} (\hat{a}_i)_\lambda &= (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \quad \hat{g}_i = (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{p}_i = \hat{\lambda} \log(1 - \hat{a}_i) + 1, \\ \hat{a}_i &= \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \quad (\hat{b}_i)_\lambda = (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i), \\ \hat{h}_i &= (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{q}_i = \hat{\lambda} \log(1 - \hat{b}_i) + 1, \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \\ \hat{z}v_i &= (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}. \end{aligned}$$

Os elementos do vetor $\boldsymbol{\Delta}_\sigma$ são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} [S_u \hat{g}_i (\hat{a}_i)_{\sigma\beta} - S_v \hat{h}_i (\hat{b}_i)_{\sigma\beta}]}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} (\hat{z}u_i \hat{g}_i - \hat{z}v_i \hat{h}_i) (S_v \hat{h}_i - S_u \hat{g}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-2} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \left[(\hat{a}_i)_{\sigma\beta} + \frac{\hat{\lambda} \hat{z}u_i \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned} (\hat{a}_i)_{\sigma\beta} &= 1 + (1 - \hat{a}_i)^{-1} \hat{z}u_i \hat{a}_i \exp(\hat{z}u_i) [\hat{\lambda} - 1] + \hat{z}u_i [1 - \exp(\hat{z}u_i)], \quad (\hat{a}_i)_\lambda = (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \\ \hat{g}_i &= (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \\ (\hat{b}_i)_{\sigma\beta} &= 1 + (1 - \hat{b}_i)^{-1} \hat{z}v_i \hat{b}_i \exp(\hat{z}v_i) [\hat{\lambda} - 1] + \hat{z}v_i [1 - \exp(\hat{z}v_i)], \quad (\hat{b}_i)_\lambda = (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i), \\ \hat{h}_i &= (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}. \end{aligned}$$

Os elementos da matriz $\boldsymbol{\Delta}_\beta$, para $j = 1, \dots, p$, são dados por

$$\Delta_{ji} = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} x_{ij} [S_u \hat{g}_i (\hat{a}_i)_\beta - S_v \hat{h}_i (\hat{b}_i)_\beta]}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} x_{ij} (\hat{g}_i - \hat{h}_i) (S_v \hat{h}_i - S_u \hat{g}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ \frac{\hat{\lambda} \hat{\sigma}^{-2} x_{ij} S_u \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \left[(\hat{a}_i)_\beta + \frac{\hat{\lambda} \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \right] & \text{se } i \in C, \end{cases}$$

em que

$$\begin{aligned} (\hat{a}_i)_\beta &= 1 + (1 - \hat{a}_i)^{-1} \hat{a}_i \exp(\hat{z}u_i) [\hat{\lambda} - 1] - \exp(\hat{z}u_i), \quad (\hat{a}_i)_\lambda = (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \\ \hat{g}_i &= (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \\ (\hat{b}_i)_\beta &= 1 + (1 - \hat{b}_i)^{-1} \hat{b}_i \exp(\hat{z}v_i) [\hat{\lambda} - 1] - \exp(\hat{z}v_i), \quad (\hat{b}_i)_\lambda = (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i), \\ \hat{h}_i &= (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}. \end{aligned}$$

Perturbação de uma covariável

O objetivo é avaliar a sensibilidade do modelo a pequenas perturbações em uma particular covariável contínua, denotada por X_t . Considere uma perturbação aditiva para a covariável, definida por $x_{it\omega} = x_{it} + \omega_i S_t$, em que S_t é um fator escala que pode ser a estimativa do desvio padrão da variável X e $\omega_i \in \mathbf{R}$.

Nesse caso, o logaritmo da função de verossimilhança perturbada é dada por

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i \in F} l_1(\lambda, zu_i^{**}, zv_i^{**}) + \sum_{i \in C} l_2(\lambda, zu_i^{**}),$$

em que $zu_i^{**} = (u_i - \mathbf{x}_i^{*T} \boldsymbol{\beta})/\sigma$, $zv_i^{**} = (v_i - \mathbf{x}_i^{*T} \boldsymbol{\beta})/\sigma$, em que $\mathbf{x}_i^{*T} \boldsymbol{\beta} = \beta_1 + \beta_2 x_{i2} + \dots + \beta_t x_{it\omega} + \dots + \beta_p x_{ip}$, $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$ é o vetor de não perturbação, $l_1(\cdot)$ e $l_2(\cdot)$ são definidos na expressão (2.7).

Os elementos da matriz $\Delta = (\Delta_\lambda^T, \Delta_\sigma^T, \Delta_\beta^T)^T$ são expressos da seguinte maneira. Os elementos do vetor Δ_λ são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\sigma}^{-1} \hat{\beta}_t S_x [\hat{g}_i \hat{p}_i - \hat{h}_i \hat{q}_i]}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda} \hat{\sigma}^{-1} \hat{\beta}_t S_x [(\hat{b}_i)_\lambda - (\hat{a}_i)_\lambda] (\hat{g}_i - \hat{h}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ \frac{\hat{\sigma}^{-1} \hat{\beta}_t S_x \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \left[\hat{p}_i + \frac{\hat{\lambda} (\hat{a}_i)_\lambda}{1 - (1 - \hat{a}_i)^\lambda} \right] & \text{se } i \in C, \end{cases}$$

em que

$$(\hat{a}_i)_\lambda = (1 - \hat{a}_i)^\lambda \log(1 - \hat{a}_i), \quad \hat{g}_i = (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{p}_i = \hat{\lambda} \log(1 - \hat{a}_i) + 1,$$

$$\hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}, \quad (\hat{b}_i)_\lambda = (1 - \hat{b}_i)^\lambda \log(1 - \hat{b}_i),$$

$$\hat{h}_i = (1 - \hat{b}_i)^{\lambda-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{q}_i = \hat{\lambda} \log(1 - \hat{b}_i) + 1, \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e}$$

$$\hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.$$

Os elementos do vetor Δ_σ são dados por

$$\Delta_i = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} \hat{\beta}_t S_x [\hat{h}_i (\hat{b}_i)_{\sigma\beta} - \hat{g}_i (\hat{a}_i)_{\sigma\beta}]}{(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} \hat{\beta}_t S_x (\hat{z}u_i \hat{g}_i - \hat{z}v_i \hat{h}_i) (\hat{g}_i - \hat{h}_i)}{[(1 - \hat{b}_i)^\lambda - (1 - \hat{a}_i)^\lambda]^2} & \text{se } i \in F, \\ \frac{-\hat{\lambda} \hat{\sigma}^{-2} \hat{\beta}_t S_x \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \left[(\hat{a}_i)_{\sigma\beta} + \frac{\hat{z}u_i \hat{g}_i}{1 - (1 - \hat{a}_i)^\lambda} \right] & \text{se } i \in C, \end{cases}$$

em que

$$(\hat{a}_i)_{\sigma\beta} = 1 + (1 - \hat{a}_i)^{-1} \hat{z}u_i \hat{a}_i \exp(\hat{z}u_i) [\hat{\lambda} - 1] + \hat{z}u_i [1 - \exp(\hat{z}u_i)], \quad \hat{g}_i = (1 - \hat{a}_i)^{\lambda-1} \hat{a}_i \exp(\hat{z}u_i)$$

$$\hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma},$$

$$(\hat{b}_i)_{\sigma\beta} = 1 + (1 - \hat{b}_i)^{-1} \hat{z}v_i \hat{b}_i \exp(\hat{z}v_i)[\hat{\lambda} - 1] + \hat{z}v_i[1 - \exp(\hat{z}v_i)], \quad \hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i),$$

$$\hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.$$

Os elementos da matriz $\Delta_{\boldsymbol{\beta}}$, para $j = 1, \dots, p$ e $t \neq j$ são dados por

$$\Delta_{ji} = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-2} x_{ij} S_x \hat{\beta}_t [\hat{h}_i(\hat{b}_i)_{\beta} - \hat{g}_i(\hat{a}_i)_{\beta}]}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} x_{ij} \hat{\beta}_t S_x (\hat{g}_i - \hat{h}_i)^2}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ \frac{-\hat{\lambda} \hat{\sigma}^{-2} x_{ij} \hat{\beta}_t S_x \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \left[(\hat{a}_i)_{\beta} + \frac{\hat{\lambda} \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \right] & \text{se } i \in C, \end{cases}$$

em que

$$(\hat{a}_i)_{\beta} = 1 + (1 - \hat{a}_i)^{-1} \hat{a}_i \exp(\hat{z}u_i)[\hat{\lambda} - 1] - \exp(\hat{z}u_i), \quad \hat{g}_i = (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i),$$

$$\hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma},$$

$$(\hat{b}_i)_{\beta} = 1 + (1 - \hat{b}_i)^{-1} \hat{b}_i \exp(\hat{z}v_i)[\hat{\lambda} - 1] - \exp(\hat{z}v_i), \quad \hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i),$$

$$\hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.$$

Para $t = j$

$$\Delta_{ti} = \begin{cases} \frac{\hat{\lambda} \hat{\sigma}^{-1} S_x [\hat{h}_i(\hat{b}_i)_{\omega} - \hat{g}_i(\hat{a}_i)_{\omega}]}{(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}} - \frac{\hat{\lambda}^2 \hat{\sigma}^{-2} x_{it} \hat{\beta}_t S_x (\hat{g}_i - \hat{h}_i)^2}{[(1 - \hat{b}_i)^{\hat{\lambda}} - (1 - \hat{a}_i)^{\hat{\lambda}}]^2} & \text{se } i \in F, \\ \frac{-\hat{\lambda} \hat{\sigma}^{-1} \hat{g}_i S_x}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \left[(\hat{a}_i)_{\omega} + \frac{\hat{\lambda} \hat{g}_i}{1 - (1 - \hat{a}_i)^{\hat{\lambda}}} \right] & \text{se } i \in C, \end{cases}$$

em que

$$(\hat{a}_i)_{\omega} = (1 - \hat{a}_i)^{-1} \hat{\sigma}^{-1} x_{it} \exp(\hat{z}u_i) \hat{a}_i \hat{\beta}_t (\hat{\lambda} - 1) + \hat{\sigma}^{-1} x_{it} \hat{\beta}_t [1 - \exp(\hat{z}u_i)],$$

$$\hat{g}_i = (1 - \hat{a}_i)^{\hat{\lambda}-1} \hat{a}_i \exp(\hat{z}u_i), \quad \hat{a}_i = \exp[-\exp(\hat{z}u_i)], \quad \hat{z}u_i = (\log(u_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma},$$

$$(\hat{b}_i)_{\omega} = (1 - \hat{b}_i)^{-1} \hat{\sigma}^{-1} x_{it} \exp(\hat{z}v_i) \hat{b}_i \hat{\beta}_t (\hat{\lambda} - 1) + \hat{\sigma}^{-1} x_{it} \hat{\beta}_t [1 - \exp(\hat{z}v_i)],$$

$$\hat{h}_i = (1 - \hat{b}_i)^{\hat{\lambda}-1} \hat{b}_i \exp(\hat{z}v_i), \quad \hat{b}_i = \exp[-\exp(\hat{z}v_i)] \quad \text{e} \quad \hat{z}v_i = (\log(v_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/\hat{\sigma}.$$

2.4 Análise de resíduos

A fim de estudar as suposições dos erros bem como a presença de observações discrepantes, na literatura são propostos várias análises de resíduos sob o mecanismo de censura à direita, por exemplo, Cox e Snell(1968), Collett (2003) e Colosimo e Giolo (2006) descreveram os resíduos de Cox-Snell, martingale e deviance. Carrasco(2007) apresentou os resíduos martingale e deviance para duas distribuições que modelam função de taxa de falha em forma de banheira e Ortega, Paula e Bolfarine (2007) apresentaram o resíduo deviance para o modelo de regressão Gama generalizada.

De maneira geral, define-se o resíduo referente a i -ésima observação por meio de uma função, $r_i = r(y_i, \hat{\theta})$, que depende da variável resposta, das estimativas dos parâmetros e do vetor de parâmetros estimado, que objetiva identificar a discrepância entre o modelo ajustado e o conjunto de dados.

Farrington (2000) propõe resíduos no contexto de censura intervalar, mas para modelos semi-paramétricos. Nesta seção estão revisados os resíduos que são aplicados na análise de sobrevivência quando tempos exatos são observados em intervalos.

2.4.1 Resíduo de Cox-Snell ajustado

Observações censuradas conduzem a resíduos que não podem ser considerados no mesmo caminho como os resíduos obtidos de observações não censuradas. Deve-se, portanto, modificar o resíduo de Cox-Snell de forma que se possa levar em consideração as censuras intervalares. Assim, define-se o resíduo de Cox-Snell ajustado para dados com censura intervalar como

$$r_{C_i} = \frac{\hat{S}[\log(u_i)] \left\{ 1 - \log \hat{S}[\log(u_i)] \right\} - \hat{S}[\log(v_i)] \left\{ 1 - \log \hat{S}[\log(v_i)] \right\}}{\hat{S}[\log(u_i)] - \hat{S}[\log(v_i)]}.$$

Para o modelo de regressão Weibull-exponenciada para dados de censura intervalar, o resíduo Cox-Snell ajustado é expresso como

$$r_{C_i} = \begin{cases} \frac{(1-\hat{a}_i)^{\hat{\lambda}} [1-\hat{\lambda} \log(1-\hat{a}_i)] - (1-\hat{b}_i)^{\hat{\lambda}} [1-\hat{\lambda} \log(1-\hat{b}_i)]}{(1-\hat{b}_i)^{\hat{\lambda}} - (1-\hat{a}_i)^{\hat{\lambda}}} & \text{se } i \in F, \\ 1 - \hat{\lambda} \log(1 - \hat{a}_i) & \text{se } i \in C, \end{cases}$$

em que $\hat{a}_i = \exp[-\exp(\hat{z}u_i)]$, $\hat{b}_i = \exp[-\exp(\hat{z}v_i)]$, $\hat{z}u_i = \frac{\log(ui) - \mathbf{x}_i^T \hat{\beta}}{\hat{\sigma}}$ e $\hat{z}v_i = \frac{\log(vi) - \mathbf{x}_i^T \hat{\beta}}{\hat{\sigma}}$.

2.4.2 Resíduo de martingale

Uma adaptação do resíduo de Cox-Snell para os dados de censura intervalar é uma versão do resíduo discutido por Lagakos (1980), também chamado de resíduo martingale (BARLOW; PRENTICE, 1988). Os resíduos de martingale são assimétricos e assumem valores máximos em +1 e

valores mínimos em $-\infty$. Assim, o resíduo martingale para dados com censura intervalar é expresso da seguinte forma

$$r_{M_i} = \frac{\hat{S}[\log(u_i)] \log \hat{S}[\log(u_i)] - \hat{S}[\log(v_i)] \log \hat{S}[\log(v_i)]}{\hat{S}[\log(u_i)] - \hat{S}[\log(v_i)]}.$$

Considerando o modelo de regressão Weibull-exponenciada para dados com censura intervalar, o resíduo martingale é definido por

$$r_{M_i} = \begin{cases} \frac{\hat{\lambda}[(1-\hat{a}_i)^{\hat{\lambda}} \log(1-\hat{a}_i) - (1-\hat{b}_i)^{\hat{\lambda}} \log(1-\hat{b}_i)]}{(1-\hat{b}_i)^{\hat{\lambda}} - (1-\hat{a}_i)^{\hat{\lambda}}} & \text{se } i \in F, \\ \hat{\lambda} \log(1 - \hat{a}_i) & \text{se } i \in C, \end{cases}$$

em que $\hat{a}_i = \exp[-\exp(\hat{z}u_i)]$, $\hat{b}_i = \exp[-\exp(\hat{z}v_i)]$, $\hat{z}u_i = \frac{\log(ui) - \mathbf{x}_i^\top \hat{\beta}}{\hat{\sigma}}$ e $\hat{z}v_i = \frac{\log(vi) - \mathbf{x}_i^\top \hat{\beta}}{\hat{\sigma}}$.

2.4.3 Resíduo deviance modificado

Therneau, Grambsch e Fleming (1990) introduziram o resíduo deviance em processos de contagem e esse resíduo para os modelos de riscos proporcionais com covariáveis não dependendo do tempo é definido como

$$r_{D_i} = \text{sinal}(r_{M_i}) [-2 \{r_{M_i} + \delta_i \log(\delta_i - r_{M_i})\}]^{1/2}.$$

em que r_{M_i} é o resíduo martingale. O resíduo deviance pode ser considerado como uma modificação feita no resíduo martingale com a finalidade de torná-lo mais simétrico ao redor do zero (COLLETT, 2003).

No modelo de regressão Weibull-exponenciada para dados de censura intervalar, o resíduo deviance é definido por

$$r_{D_i} = \begin{cases} \text{sinal}(\hat{r}_{M_i}) \left\{ -2 \left[\frac{\hat{\lambda}[(\hat{a}_i)_\lambda - (\hat{b}_i)_\lambda]}{(1-\hat{b}_i)^{\hat{\lambda}} - (1-\hat{a}_i)^{\hat{\lambda}}} + \log \left(1 - \left\{ \frac{\hat{\lambda}[(\hat{a}_i)_\lambda - (\hat{b}_i)_\lambda]}{(1-\hat{b}_i)^{\hat{\lambda}} - (1-\hat{a}_i)^{\hat{\lambda}}} \right\} \right) \right] \right\}^{1/2} & \text{se } i \in F, \\ \text{sinal}(\hat{r}_{M_i}) [-2\hat{\lambda} \log(1 - \hat{a}_i)]^{1/2} & \text{se } i \in C, \end{cases}$$

em que $(\hat{a}_i)_\lambda = (1 - \hat{a}_i)^{\hat{\lambda}} \log(1 - \hat{a}_i)$, $(\hat{b}_i)_\lambda = (1 - \hat{b}_i)^{\hat{\lambda}} \log(1 - \hat{b}_i)$, $\hat{a}_i = \exp[-\exp(\hat{z}u_i)]$, $\hat{b}_i = \exp[-\exp(\hat{z}v_i)]$, $\hat{z}u_i = \frac{\log(ui) - \mathbf{x}_i^\top \hat{\beta}}{\hat{\sigma}}$ e $\hat{z}v_i = \frac{\log(vi) - \mathbf{x}_i^\top \hat{\beta}}{\hat{\sigma}}$.

Os gráficos dos resíduos deviance modificado bem como dos resíduos Cox-Snell ajustado e martingale *versus* os tempos observados fornecem uma forma de verificar a adequação do modelo ajustado e auxiliam a encontrar observações atípicas.

2.4.4 Estudo de simulação

Um estudo de simulação foi realizado para investigar o comportamento da distribuição empírica dos resíduos r_{C_i} , r_{M_i} e r_{D_i} do modelo de regressão Weibull-exponenciada com censura intervalar proposto na seção 2.2.

O logaritmo dos tempos de sobrevivência $y_1, \dots, y_n$ foram gerados por meio de simulações de Monte Carlo da distribuição log-Weibull-exponenciada considerando o modelo de regressão (2.4), isto é, fazendo $y_i = \beta_0 + \beta_1 x_{i1} + \sigma [\log\{-\log(1 - w_i^{1/\lambda})\}]$ em que $w_i \sim U(0, 1)$ e os valores fixados foram $\beta_0 = 4$, $\beta_1 = 2$, $\sigma = 0,5$ e $\lambda = 1$ e $0,5$ e x_i sendo gerado de uma distribuição uniforme no intervalo $(0, 1)$. Simulou-se amostras de tamanho $n = 50$, $n = 100$ e $n = 200$. Os tempos de censura à direita $c_1, \dots, c_n$ foram gerados da distribuição uniforme $(0, \tau)$, em que τ foi ajustado até obter as porcentagens de censura à direita 10% e 30%. Para os logaritmos dos tempos com censura intervalar foram gerados $\log(u_i)$ e $\log(v_i)$ de tal forma que $y_i \in (\log(u_i), \log(v_i))$ com probabilidade um, seguindo o mesmo enfoque adotado por Zhao(2004).

Para cada combinação de n , λ , σ e porcentagens de censuras, 1000 amostras aleatórias foram geradas e para cada uma ajustou-se o modelo de regressão Weibull-exponenciada com censura intervalar. Então para cada modelo ajustado, os resíduos r_{C_i} , r_{M_i} e r_{D_i} foram calculados e armazenados. O programa da simulação é descrito no apêndice A4. As Figuras 2.3-2.8 apresentam o gráfico dos resíduos *versus* os valores esperados das estatísticas de ordem da distribuição normal padrão, esse é o gráfico normal de probabilidades que serve para avaliar o afastamento da suposição de normalidade para esses resíduos (WEISBERG, 1985).

Desses gráficos as seguintes interpretações são obtidas:

- Observou-se claramente que a distribuição empírica do resíduo deviance modificado concorda com a distribuição normal padrão
- A distribuição empírica do resíduo de Cox-Snell ajustado e o resíduo martingale apresentam comportamento semelhante.
- A distribuição empírica do resíduo de Cox-Snell ajustado e o resíduo martingale se apresentam em geral com uma acentuada assimetria.
- Quando a censura decresce, a distribuição empírica do resíduo deviance modificado parece chegar próximo a distribuição normal padrão mais rapidamente do que os outros dois resíduos considerados na análise.
- Quando o tamanho da amostra aumenta, a distribuição empírica do resíduo deviance modificado parece apresentar melhor acordo com a distribuição normal.

Figura 2.3 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=0 e $\lambda = 1$

Figura 2.4 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=10 e $\lambda = 1$

Figura 2.5 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=30 e $\lambda = 1$

Figura 2.6 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=30 e $\lambda = 0,5$

Figura 2.7 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=30 e $\lambda = 0,5$

Figura 2.8 – Gráficos normais de probabilidade para o resíduo Cox-Snell ajustado (r_{C_i}), resíduo martingale (r_{M_i}) e resíduo deviance modificado (r_{D_i}). Tamanho de amostra $n=50$, $n=100$ e $n=200$, porcentagem de censura à direita=30 e $\lambda = 0,5$

2.5 Aplicação

Os dados desse exemplo referem-se a um estudo retrospectivo, envolvendo câncer de mama, extraído do trabalho de Colosimo e Giolo (2006). Esses dados têm sido utilizados na literatura como exemplo para abordagem de estudos de censura intervalar por meio de modelos de riscos proporcionais. O estudo envolveu 94 mulheres com diagnóstico precoce de câncer de mama, cujo objetivo principal é pesquisar duas terapias no tratamento da doença, um somente com radioterapia e outra

com radioterapia em conjunto com quimioterapia. As pacientes foram acompanhadas a cada 4 ou 6 meses, e como as visitas foram realizadas em tempos aleatórios, não se sabe com exatidão quando a primeira retração da mama ocorreu, somente que esta ocorreu entre duas das visitas realizadas. A covariável associada a cada paciente é o tipo de tratamento: 0 se recebeu somente radioterapia e 1 se recebeu radioterapia em conjunto com quimioterapia. Detalhes sobre esses dados e respectiva documentação estão disponíveis em Finkelstein e Wolfe (1985) e em Klein e Moeschberger (1997). A porcentagem de censura à direita foi de 40%.

2.5.1 Estimador de máxima verossimilhança e estimador Jackknife

As estimativas de máxima verossimilhança (EMV) dos parâmetros do modelo de regressão Weibull-exponenciada para dados de censura intervalar foram obtidas por meio da sub-rotina MAXBFGS da linguagem de programação Ox. Os resultados são apresentados na Tabela 2.1.

Tabela 2.1 – Estimativas de máxima verossimilhança dos parâmetros do modelo de regressão Weibull-exponenciada para censura intervalar ajustado para os dados de câncer

Parâmetro	Estimativa	Erro padrão	<i>p</i> -valor	Intervalo de confiança(95%)
λ	0,451	0,349	-	(0,000 ;1,136)
σ	0,341	0,216	-	(0,000 ;0,765)
β_0	4,187	0,233	<0,001	(3,731;4,644)
β_1	-0,569	0,169	<0,001	(-0,899;-0,238)

Na Tabela 2.2 são apresentados as estimativas de Jackknife para os parâmetros do modelo de regressão Weibull-exponenciada com censura intervalar.

Tabela 2.2 – Estimativas Jackknife dos parâmetros do modelo de regressão Weibull-exponenciada para censura intervalar ajustado para os dados de câncer

Parâmetro	Estimativa	Erro padrão	Intervalo de confiança(95%)
λ	0,383	0,355	(0,000 ; 1,080)
σ	0,341	0,220	(0,000 ; 0,773)
β_0	4,245	0,226	(3,802;4,688)
β_1	-0,577	0,160	(-0,891;-0,263)

Observou-se que a covariável X_1 (tratamento) é significativa no modelo de regressão Weibull-exponenciada com censura intervalar ao nível de 5% de significância quando o estimador de máxima verossimilhança, bem como o estimador Jackknife é utilizado.

As estimativas por meio dos dois métodos parecem muito similares, porém as EMV parecem mais conservativas (erros padrões maiores).

Além disso, o teste da razão de verossimilhanças foi realizado para testar as hipóteses $H_0 : \lambda = 1$ ou o modelo de regressão Weibull-exponenciada é adequado contra $H_1 : \lambda \neq 1$ ou o modelo de regressão Weibull é adequado. O valor da estatística do teste de (2.9) é dado por $\Lambda_n = 0,94$; na qual é não significativo ao nível de significância de 5%. Logo, o modelo de regressão Weibull também é adequado para este conjunto de dados, entretanto, continuou-se a análise utilizando as EMV e considerando o modelo de regressão Weibull-exponenciada.

2.5.2 Influência global

Nesta sub-seção, utilizou-se a linguagem de programação Ox para calcular medidas de deleção de casos, distância de Cook generalizada ($GD_i(\theta)$) e afastamento da verossimilhança ($LD_i(\theta)$) apresentadas na sub-seção 3.6.1. Os resultados das medidas de influência global são apresentadas na Figura 3.2. A Figura 3.2a indica que as observações #48 e #81 são possíveis observações influentes. A Figura 3.2b destaca somente a observação #81.

Figura 2.9 – Gráficos de medidas de influência global para o modelo de regressão Weibull-exponenciada ajustado aos dados de câncer. (a) Gráfico da distância de Cook generalizada e (b) gráfico do afastamento da verossimilhança

A observação #48 representa o indivíduo no grupo de tratamento com radioterapia conjuntamente com quimioterapia e apresenta o maior tempo dentro do conjunto dos indivíduos censurados

à direita. A observação #81 representa o indivíduo com menor intervalo de tempo no grupo de indivíduos que receberam o tratamento somente com radioterapia.

2.5.3 Influência local

Nesta sub-bseção, realizou-se uma análise de influência local apresentada na sub-seção 2.3.2 para o conjunto de dados de câncer utilizando o modelo de regressão Weibull-exponenciada para dados com censura intervalar.

De acordo com o esquema de perturbação de casos, o valor $C_{d_{max}} = 1,17$ foi obtido como a curvatura máxima. Na Figura 3.4a é apresentado o gráfico do autovetor correspondente a direção máxima d_{max} . A medida de influência local total C_i é apresentada na Figura 3.4b. As observações #48 e #81 novamente se mostram distantes em relação as outras observações.

Figura 2.10 – Gráficos de medidas de influência para θ do modelo de regressão Weibull-exponenciada sob o esquema de perturbação de casos nos dados de câncer. (a) Gráfico de influência local d_{max} e (b) gráfico de influência local total C_i

Ao considerar o esquema de perturbação do intervalo, o valor para curvatura máxima obtido foi $C_{d_{max}} = 137,51$. A Figura 2.11a mostra o gráfico para a direção máxima d_{max} versus os índices das observações e a Figura 2.11b mostra a influência local total C_i , onde as observações #15 e #48 são possíveis pontos influentes.

A observação #15 representa o indivíduo com o maior intervalo de tempo e com menor comprimento de intervalo no grupo de tratamento somente com radioterapia.

Figura 2.11 – Gráficos de medidas de influência para θ do modelo de regressão Weibull-exponenciada sob o esquema de perturbação no intervalo nos dados de câncer. (a) Gráfico de influência local d_{max} e (b) gráfico de influência local total C_i

Assumindo o esquema de perturbação do limite inferior do intervalo, o valor $C_{d_{max}} = 131,19$ foi obtido como a curvatura máxima. Na Figura 2.12a é apresentado o gráfico do autovetor correspondente ao $C_{d_{max}}$. A medida de influência local total C_i é apresentada na Figura 2.12b. A Figura 2.12 indica que a observação #48 destaca-se como possível ponto influente. Sob o enfoque do limite inferior do intervalo, a observação #48 é o maior tempo de vida do indivíduo que foi censura intervalar e pertencente ao grupo de tratamento com radioterapia conjuntamente com quimioterapia.

Agora, perturbando o limite superior do intervalo, o valor para curvatura máxima obtido foi $C_{d_{max}} = 1,12$. A Figura 2.13a mostra o gráfico para direção máxima d_{max} versus os índices das observações e a Figura 2.13b mostra a influência local total C_i , onde apenas a observação a #15 destaca-se como possível ponto influente.

Analisando somente o limite superior do intervalo, a observação #15 representa o maior tempo de vida do grupo de indivíduos que recebeu o tratamento com radioterapia conjuntamente com quimioterapia.

Figura 2.12 – Gráficos de medidas de influência para θ do modelo de regressão Weibull-exponenciada o esquema de perturbação no limite inferior resposta nos dados de câncer. (a) Gráfico de influência local d_{max} e (b) gráfico de influência local total C_i

Figura 2.13 – Gráficos de medidas de influência para θ do modelo de regressão Weibull-exponenciada sob o esquema de perturbação no limite superior da resposta nos dados de câncer. (a) Gráfico de influência local d_{max} e (b) gráfico de influência local total C_i

2.5.4 Análise de resíduos

Com o objetivo de verificar possíveis observações discrepantes bem como o afastamento das suposições do modelo, a Figura 2.14a apresenta o gráfico do resíduo de Cox-Snell ajustado, a Figura 2.14b mostra o gráfico do resíduo de martingale e a 2.14c tem-se o gráfico do resíduo deviance modificado.

Por meio da análise gráfica dos resíduos da Figura 3.6, observou-se o comportamento aleatório dos dados e indicam como possíveis pontos discrepantes as observações #48 e #81. O gráfico da Figura 2.14c mostra que os modelos de regressão Weibull-exponenciada para dados de censura intervalar não parece ser inadequado para ajustar os dados de câncer.

Figura 2.14 – Gráfico de resíduo para o modelo de regressão Weibull-exponenciada para dados de censura intervalar nos dados de câncer. (a)Resíduo Cox-Snell ajustado (r_{C_i}), (b)resíduo martingale (r_{M_i}) e (c)resíduo deviance modificado (r_{D_i})

2.5.5 Impacto da deleção de observações influentes

Assim, a análise de sensibilidade (influência global e influência local) e a análise de resíduo detectaram como possíveis pontos influentes as observações #48 e #81. A fim de analisar o impacto dessas observações nas estimativas dos parâmetros, reajustou-se o modelo eliminando individualmente cada observação, em seguida, eliminando as duas observações.

Na Tabela 2.3 são apresentados as mudanças relativas (em porcentagens) de cada parâmetro estimado, definido por $\mathbf{RC}_{\theta_j} = \left[(\hat{\theta}_j - \hat{\theta}_{j(i)}) / \hat{\theta}_j \right] \times 100$, em que $\theta_{j(i)}$ é o estimador de máxima verossimilhança sem a i -ésima observação.

Tabela 2.3 – Mudanças relativas [-RC-em %], estimativas dos parâmetros e os correspondentes p -valores em parênteses, excluindo as observações #48 e #81

Exclusão	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\beta}_0$	$\hat{\beta}_1$
Nenhuma	-	-	-	-
	0,451	0,341	4,187	-0,569
	(-)	(-)	(0,000)	(0,001)
#48	[48]	[46]	[-1]	[-3]
	0,234	0,185	4,261	-0,586
	(-)	(-)	(0,000)	(0,000)
#81	[-30]	[-17]	[2]	[-2]
	0,584	0,400	4,122	-0,582
	(-)	(-)	(0,000)	(0,000)
#48 e #81	[32]	[34]	[-1]	[-6]
	0,305	0,224	4,229	-0,600
	(-)	(-)	(0,000)	(0,000)

Os resultados da Tabela 2.3 mostraram que as estimativas de máxima verossimilhança do modelo de regressão Weibull-exponenciada para dados de censura intervalar são robustos sobre a deleção das observações influentes. Em geral, a significância dos parâmetros estimados não alterou-se (ao nível de 1% de significância) após a remoção dos casos. Portanto, não houve mudanças inferências após a remoção das observações consideradas influentes nos gráficos de diagnósticos.

2.5.6 Qualidade de ajuste

Com o objetivo de verificar a qualidade do ajuste do modelo de regressão Weibull-exponenciada com censura intervalar, a Figura 2.15 representa o gráfico de probabilidade normal para o resíduo deviance modificado com envelope simulado (ATKINSON, 1985), adotado como uma medida de qualidade de ajuste.

Por meio da Figura 2.15, observou-se que há indícios de um bom ajuste do modelo de regressão Weibull-exponenciada para os dados de câncer de mama com censura intervalar.

Figura 2.15 – Gráfico normal de probabilidade para o resíduo deviance modificado (r_{D_i}) com envelope para o modelo de regressão Weibull-exponenciada nos dados de câncer

2.6 Conclusão

Neste trabalho, um modelo de regressão Weibull-exponenciada para dados de censura intervalar foi proposto. Utilizou-se o algoritmo de Quase-Newton para obter as estimativas de máxima verossimilhança e foi realizado testes para os parâmetros baseados nas distribuições assintóticas dos estimadores de máxima verossimilhança. Além disso, vários estudos de simulações desenvolvidas neste trabalho indicaram que a distribuição do resíduo deviance modificado mostra uma alta concordância com a distribuição normal padrão. As matrizes necessárias para aplicações das técnicas de influência local foram obtidas levando em conta perturbações para dados com censura intervalar no modelo ou nos dados. Por meio da aplicação dos procedimentos em um conjunto de dados da área médica, avaliou-se os aspectos sensíveis das estimativas de máxima verossimilhança sobre esquemas de perturbações bem como verificar a qualidade de ajuste do modelo postulado. Embora os gráficos de diagnósticos detectaram algumas possíveis observações influentes, a deleção dos mesmos não causaram mudanças inferenciais nos resultados.

2.6.1 Trabalhos futuros

Como possíveis trabalhos futuros pode-se considerar os seguintes temas de pesquisa:

1. Considerar modelos com fração de cura e censura intervalar.

2. Propor um modelo de regressão com censura intervalar e efeito aleatório.
3. Propor um modelo de regressão locação-escala no caso de dados de dupla censura, caso particular de dados de censura intervalar.
4. Realizar um estudo de análise de sensibilidade sob o enfoque de influência local conformal proposta por Poon e Poon (1999).
5. Desenvolver a metodologia de dados de censura sob o enfoque bayesiano.

Referências

ATKINSON, A.C. **Plots, transformations, and regression**. Oxford: University Press, 1985. 280p.

BARLOW, W.E.; PRENTICE, R.L. Residuals for relative risk regression. **Biometrika**, Cambridge, v.75, p.65-74, 1988.

BETENSKY, R.A.; RABINOWITZ, D.; TSIATIS, A.A. Computationally simple accelerated failure time regression for interval censored data. **Biometrika**, Cambridge, v.88, p.703-711, 2001.

CANCHO, V.G. **Métodos de monte carlo em análise de sobrevivência**. 1999. 207p. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1999.

CARRASCO, J.M.F. **Modelo de regressão log-Weibull modificado e a nova distribuição Weibull modificada generalizada**. 2007. 128p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

CARRASCO, J.M.F.; ORTEGA, E.M.M.; PAULA, G.A. Log-Modified Weibull regression models with censored data: sensitivity and residual analysis. **Computational Statistics and Data Analysis**, Amsterdam, v.52, p.4021-4029, 2008.

CHRISTENSEN, R.; PEARSON, L.; JOHNSON, W. Case-deletion diagnostics for mixed models. **Technometrics**, Alexandria, v.34, p.38-45, 1992

COLLETT, D. **Modelling survival data in medical research**. 2nd ed. London: Chapman & Hall, 2003. 391p.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 374p.

COOK, R.D. Detection of influential observations in linear regression. **Technometrics**, Alexandria, v.19, p.15-18, 1977

_____. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society B**, London, v.48, p.133-169, 1986.

COOK, R.D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman & Hall, 1982. 280p.

COOK, R.D.; PEÑA, D.; WEISBERG, S. The likelihood displacement: a unifying principle for influence. **Communications in Statistics: Theory and Methods**, New York, v.17, p.623-640, 1988.

COX, D.R.; HINKLEY, D.V. **Theoretical statistics**. London: Chapman & Hall, 1974, 511p.

COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society B**, Oxford, v.30, p.248-275, 1968.

DAVISON, A.C.; TSAI, C.L. Regression model diagnostics. **International Statistical Reviews**, Hague, v.60, p.337-355, 1992.

DOORNIK, J.A. **Ox**: an object-oriented matrix programming language. Disponível em: <<http://www.nuff.ox.ac.uk/User/Doonik/>>. Acesso em: 15 ago. 2007.

ESCOBAR, L.A.; MEEKER, W.Q. Assessing influence in regression analysis with censored data. **Biometrics**, Washington, v.48, p.507-528, 1992.

FACHINI, J.B. **Análise de influência local nos modelos de riscos múltiplos**. 2007. 78p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

FACHINI, J.B.; ORTEGA, E.M.M.; LOUZADA-NETO, F. Influence diagnostics for polyhazard models in the presence of covariates. **Statistical Methods and Applications**, New York, v.17, p.413-433, 2008.

FANG, H.; SUN, J.; LEE, M.T. Nonparametric survival comparisons for interval-censored continuous data. **Statistica Sinica**, Taipei, v.12, p.1073-1083, 2002.

FARRINGTON, C.P. Residuals for proportional hazards models with interval-censored survival data. **Biometrics**, Washington, v.56, p.473-482, 2000.

FINKELSTEIN, D.M. A proportional hazards model for interval-censored failure time data. **Biometrics**, Washington, v.42, p.845-854, 1986.

FINKELSTEIN, D.M.; WOLFE, R.A. Semi-parametric model for regression analysis of interval censored failure time data. **Biometrics**, Washington, v.41, p.933-945, 1985.

FINKELSTEIN, D.M.; GOGGINS, W.B.; SCHOENFELD, D.A. Analysis of failure time data with dependent interval censoring. **Biometrics**, Washington, v.58, p.298-304, 2002.

GALEA, M.; SÁNCHEZ, V.L.; PAULA, G.A. Influence diagnostics in log-Birnbaum-Saunders regression models. **Journal of Applied Statistics**, Abingdon, v.31, p.1049-1064, 2004.

GOETGHEBEUR, E.; RYAN, L.; Semiparametric regression analysis of interval-censored data. **Biometrics**, Washington, v.56, p.1134-1144, 2000.

HAREZLAK, J.; TU, W. Estimation of survival functions in interval and right censored data using STD behavioural diaries. **Statistics in Medicine**, Chichester, v.25, p.4053-4064, 2006

KIM, D.K. Regression analysis of interval-censored survival data with covariates using log-linear models. **Biometrics**, Washington, v.53, p.1274-1283, 1997.

KLEIN, J.P.; MOESCHBERGER, M.L. **Survival analysis techniques for censored and truncated data**. New York: Springer-Verlang, 1997. 357p.

KOOPERBERG, C.; CLARKSON, D.B. Hazard regression with interval-censored data. **Biometrics**, Washington, v.53, p.1485-1494, 1997.

KUNDU, D.; RAQAB, M.Z. Generalized Rayleigh distribution: different methods of estimations. **Computational Statistics and Data Analysis**, Amsterdam, v.49, p.187-200, 2005.

LAGAKOS, S.W. The graphical evaluation of explanatory variables in proportional hazards regression. **Biometrika**, Cambridge, v.56, p.93-98, 1980.

LANGOHR, K.; GÓMEZ, G.; MUGA, R. A parametric survival model with an interval-censored covariate. **Statistics in Medicine**, Chichester, v.23, p.3159 - 3175, 2004.

LAWLESS, J.F. **Statistical models and methods for lifetime data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 637p.

LEIVA-SANCHEZ, V.; BARROS, M.; PAULA, G.A.; GALEA, M. Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. **Computational Statistics and Data Analysis**, Amsterdam, v.51, p.5694-5707, 2006.

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

LINDSEY, J.K. A study of interval censoring in parametric regression models. **Lifetime Data Analysis**, Dordrecht, v.4, p.329-354, 1998.

LINDSEY, J.C.; RYAN, L.M. Tutorial in biostatistics methods for interval-censored data. **Statistics in Medicine**, Chichester, v.17, p.129-238, 1998.

LIPSITZ, S.R.; LAIRD, N.M.; HARRINGTON, D.P. Using the Jackknife to estimate the variance of regression estimators from repeated measures studies. **Communications in Statistics: Theory Methods**, New York, v.19, p.821-845, 1990.

MAGNUS, J.R.; VASNEV, A.L. Local sensitivity and diagnostic tests. **Econometrics Journal**, Cambridge, v.10, p.166-192, 2007.

MANLY, B.F.J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2nd ed. London: Chapman & Hall, 1997. 399p.

MUDHOLKAR, G.S.; SRIVASTAVA, D.K.; KOLLIA, G.D. A Generalization of the Weibull Distribution with Application to the Analysis of Survival Data. **Journal of the American Statistical Association**, Boston, v.91, p.1575-1583, 1996.

NADARAJAH, S. The exponentiated Gumbel distribution with climate application. **Environmetrics**, Elmont, v.17, p.13-23, 2006.

ODELL, P.M.; ANDERSON, K.M.; D'AGOSTINO, R.B. Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. **Biometrics**, Washington, v.48, p.951-959, 1992.

ORTEGA, E.M.M.; BOLFARINE, H.; PAULA G.A. Influence diagnostics in generalized log-gamma regression models. **Computational Statistics and Data Analysis**, Amsterdam, 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**, Catalunya, v.30, p.172-192, 2006.

ORTEGA, E.M.M.; CANCHO, V.G.; PAULA, G.A. Generalized log-gamma regression models with cure. **Lifetime Data Analysis**, 2008, In Press.

ORTEGA, E.M.M.; PAULA, G.A.; BOLFARINE, H. Deviance residuals in generalised log-gamma regression models with censored observations. **Journal of Statistical Computation and Simulation**, New York, v. 78, p.747-764, 2008.

PAN, W. Smooth estimation of the survival function for interval-censored data. **Statistics in Medicine**, Chichester, v.19, p.2611-2624, 2000.

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

PETTITT, A.N.; BIN DAUD, I. Case-weight measures of influence for proportional hazards regression. **Applied Statistics**, London, v.38, p.51-67, 1989.

POON, W.Y.; POON, Y.S. Conformal normal curvature and assessment of local influence. **Journal of the Royal Statistical Society Series B**, London, v.61, p.51-61, 1999.

PRUDNIKOV, A.P.; BRYCHKOV, Y.A.; MARICHEV, O.I. **Integrals and series**. New York: Gordon and Breach Science Publishers, 1986. 758p.

RABINOWITZ, D.; TSIATIS, A.; ARAGON, J. Regression with interval-censored data. **Biometrika**, Cambridge, v.82, p.501-513, 1995.

ROSINGANA, M.L. **The analysis of interval-censored survival data. From a nonparametric perspective to a nonparametric bayesian approach.** 1996. 135p. Memorial(Doutorado em Matemática) - Dept. d'Estadística I Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, 1996.

SILVA, G.O.; ORTEGA, E.M M.; CANCHO, V.G.; BARRETO, M.L. Log-Burr XII regression models with censored Data. **Computational Statistics and Data Analysis**, Amsterdam, v.52, p.3820-3842, 2008.

SINHA, D.; CHEN, M.; GHOSH, S.K. Bayesian analysis and model selection for interval-censored survival data. **Biometrics**, Washington, v.55, p.585-590, 1999.

SPARLING, Y.H.; YOUNES, N.; LACHIN, M.J. Parametric survival models for interval-censored data with time dependent covariates. **Biostatistics**, Oxford, v.7, p.599-614, 2006.

SUN, J. Regression analysis of interval-censored failure time data. **Statistics in Medicine**, Chichester, v.16, p.497-504, 1997.

THERNEAU, T.M.; GRAMBSCH, P.M; FLEMING, T.R. Martingale-based residuals for survival models. **Biometrika**, Cambridge, v.77, p.147-60, 1990.

XIE, F.; WEI, B. Diagnostics analysis for log-Birnbaum-Saunders regression models. **Computational Statistics and Data Analysis**, Amsterdam, v.51, p.4692-4706, 2007.

XIE, F.C.; WEI, B.C. Diagnostics analysis in censored generalized Poisson regression model. **Journal of Statistical Computation and Simulation**, New York, v.77, p.695-708, 2007.

WANG, L. **Statistical analysis of multivariate interval-censored failure time data.** 2006, 109 p. Thesis(Philosophy Doctor) - Faculty of the Graduate School, University of Missouri, Columbia, 2006.

WEISBERG, S. **Applied linear regression.** 2nd ed. New York: John Wiley & Sons, 1985, 324p.

ZHAO, Q. **Nonparametric treatment comparisons for interval-censored failure time data.** 2004. 130p. Thesis (Phylosophy Doctor) - Faculty of the Graduate School, University of Missouri, Columbia, 2004.

3 MODELO DE REGRESSÃO PARA DADOS DE SOBREVIVÊNCIA GRUPADOS

Resumo

Neste trabalho, foi estudado um modelo regressão para dados de sobrevivência grupados quando o efeito de tempo de censura é considerado na modelagem e a estrutura de regressão modelada por meio de quatro funções de ligação. A metodologia de dados de sobrevivência grupados baseia-se em tabelas de vida, os tempos são grupados em k intervalos de modo que os empates sejam eliminados e assim, a modelagem dos dados é feita considerando os modelos de regressão de tempos de vida discreto. Os parâmetros do modelo foram estimados utilizando os métodos de máxima verossimilhança e Jackknife. Para detectar observações influentes no modelo proposto foram usadas medidas de diagnósticos baseadas no caso deleção, denominada influência global, e medidas de influência baseadas em pequenas perturbações nos dados ou no modelo, chamada de influência local. Além dessas medidas foi também utilizada a medidas de influência local total e estimativa influente. Adicionalmente, desenvolveu-se uma análise de resíduos baseada no resíduo de Pearson. Por fim, um conjunto de dados reais foi analisado, utilizando o modelo de regressão para dados grupados proposto como ilustração.

Palavras-chave: Dados de sobrevivência grupados; Funções de ligação; Análise de sensibilidade

Abstract

In this work, it was studied a model regression for grouped survival data when the effect of time of censorship is considered in the modeling and the regression structure modeled through four link functions. The methodology of grouped survival data is based in life tables, the times are grouped in k intervals so that the ties are eliminated and like this, the modeling of the data is made considering the discrete models of regression of lifetime. The parameters of the model were estimated using the maximum likelihood and Jackknife methods. To detect influential observations in the proposed model based on diagnostic measures were used based on the case deletion, denominated of global influence, and influence measures based on small perturbations in the data or in the model, called of local influence. Besides those measures it was also used the total local influence and influential estimate. Additionally, it was developed residual analysis based in the Pearson residual. Finally, a data set is analyzed by using the regression model for grouped data proposed as illustration.

Keywords: Survival grouped data; Link function; Sensitivity analysis;

3.1 Introdução

Dados de sobrevivência são comuns em estudos de ensaios clínicos, biológicos, químicos bem como em estudos agrônômicos, na qual a variável resposta é o tempo até a ocorrência de um evento de interesse. A característica principal destes dados é a presença de censuras, que é a variável resposta observada parcialmente.

Entretanto, dependendo do objetivo do experimento e como é conduzido, diferentes unidades amostrais podem falhar ou até mesmo censurar nas mesmas unidades de tempo, tais tempos de vida repetidos são denominadas de empates e são características de dados de sobrevivência grupados. Dados desse tipo são freqüentes em estudos de medidas repetidas longitudinais, entre outros, por exemplo, o experimento envolvendo mangueiras (CHALITA, 1997). No estudo, as mangueiras foram avaliadas nos anos de 1973 – 1992, ocasionando uma quantidade de empates.

Assim, considerar para esta situação modelos de sobrevivência usuais, não são adequados, pois segundo Thompson(1977) a presença de empates causa complicações na modelagem dos dados, principalmente nos resíduos que não apresentam aleatoriedade. A importância dos estudos de dados de sobrevivência grupados se deve a compreensão da natureza dos dados e a forma adequada de tratar o tempo de vida quando há a presença de censura e empates.

A metodologia de dados grupados se baseia em métodos de tabela de vida. Os dados são grupados de tal forma que tem-se apenas a informação sobre em quais intervalos os indivíduos falharam ou foram censurados e não se sabe os tempos exatos de sobrevivência ou censura, o que elimina a presença de empates. Portanto, os dados grupados são as freqüências de mortes ou censuras em cada um dos intervalos (BOLFARINE;RODRIGUES;ACHCAR, 1991).

Assim, o enfoque desta pesquisa é propor um modelo de regressão para dados grupados construindo a função de verossimilhança da mesma forma que propõem os trabalhos de Thompson (1977) e Colosimo, Chalita e Demétrio (2000). Além disso foram utilizadas quatro funções de ligação tais como, logito, complemento log-log, log-log e probito, para modelar a estrutura de regressão.

Algumas aplicações mais recentes em dados de sobrevivência grupados podem ser encontrados na literatura. Hertz-Piccioto e Rockhill (1997) estudaram a eficiência das aproximações da verossimilhança parcial na presença de empates, Lam e Ip (2003) apresentaram uma modelagem baseada em grupos e Yu et al. (2004) propuseram modelos com fração de cura para dados de sobrevivência grupados sob a abordagem paramétrica na ausência de covariáveis.

Para estimar os parâmetros desse modelo de regressão utilizou-se o método de máxima verossimilhança na qual o método numérico do tipo Quase-Newton foi necessário. Também realizou-se testes de hipóteses e intervalos de confiança a partir da distribuição assintótica dos estimadores de máxima verossimilhança. Entretanto, utilizar a distribuição assintótica dos estimadores e máxima verossimilhança em situações que a amostra é pequena, comum na prática, ou que as condições de regularidade não sejam satisfeitas pode ser inadequado e apresentar resultados difíceis de serem justificados. Como alternativa ao método de máxima verossimilhança foi utilizado o método de Jackknife. Por outro lado, é necessário avaliar a sensibilidade do modelo quando pequenas perturbações são inseridas no modelo ou nos dados. As técnicas que permitem estudar a sensibilidade das estimativas são conhecidas como métodos de diagnósticos e são importantes ferramentas em análise de regressão para detectar anomalias com o modelo ajustado (ORTEGA; BOLFARINE;

PAULA, 2003).

Dentre os métodos de diagnósticos, destaca-se a influência global que consiste na deleção de casos e tem por objetivo estudar o efeito da exclusão do i -ésimo indivíduo de um conjunto de dados e influência local que baseia-se na análise do afastamento da verossimilhança e tem por finalidade avaliar como os resultados da análise são modificados sob pequenas perturbações feitas no modelo. Aplicações de influência global e local para especificar problemas na área de análise de sobrevivência tem sido descritos em vários trabalhos, por exemplo, Escobar e Meeker (1992) utilizaram as técnicas de influência em análise de sobrevivência, Parker e Delong (2000) apresentaram métodos de diagnósticos para modelos discreto baseados na verossimilhança parcial, Resche-Rigon e Chevret (2006) utilizaram medidas de influência em modelos de riscos competitivos, Carrasco (2007) apresentou métodos de diagnóstico para distribuições que modelam taxa de falha em forma de banheira, Ortega, Cancho e Paula (2008) descreveram técnicas de influência em modelos de regressão log-Gama generalizada com fração de cura, Silva et al. (2008) descreveram técnicas de diagnóstico para modelos de regressão Burr-XII e Gomes (2008) apresentou análise de sensibilidade em modelos de sobrevivência bivariados utilizando a metodologia de cópulas.

Dessa forma, o presente trabalho se propõe a apresentar uma abordagem na qual as censuras não ocorrem no final do intervalo, bem como a utilização de quatro funções de ligação para modelar a estrutura de regressão de modelos de dados de sobrevivência grupados. Na seção 3.2 descreveu-se a proposta do modelo de regressão para dados de sobrevivência grupados e as quatro funções de ligação, logito, complemento log-log, log-log e probito. Na seção 3.3 e 3.4 descreveu-se os métodos de máxima verossimilhança e Jackknife para estimar os parâmetros do modelo proposto, respectivamente. Na seção 3.5 apresentou-se alguns critérios de seleção de modelos, com a finalidade de selecionar uma das funções de ligação propostas. A seção 3.6 abrange descrição da análise de sensibilidade do modelo proposto baseado em técnicas de influência global, local e local total. Na seção 3.7 descreveu-se o resíduo de Pearson para dados de sobrevivência grupados. Na seção 3.8 realizou-se uma aplicação da metodologia proposta e finalmente, na seção 3.9 encontra-se as principais conclusões deste trabalho.

3.2 Especificação do modelo de regressão para dados grupados

Define-se T como uma variável aleatória representando o tempo de falha de um indivíduo e C , uma outra variável aleatória independente de T , representando o tempo de censura associado a este indivíduo. O tempo observado para esse indivíduo é $t = \min(T, C)$ e δ é um indicador de censura associado ao indivíduo com $\delta = 1$ se $T \leq C$ e $\delta = 0$ se $T > C$.

O mecanismo de censura definido acima é conhecido como censura aleatória, a característica desse tipo de censura é a ocorrência dos tempos de censura durante o estudo bem como ao final do estudo.

Se o conjunto de dados observados, além da presença censura, apresentar um excessivo número de empates, os tempos observados $t = \min(T, C)$ são agrupados em um certo número de intervalos com a finalidade de eliminar os empates e considerar os intervalos como tempos de vida discreto de modo que as técnicas de análise de sobrevivência discreto possam ser utilizadas. Na literatura a decisão do número de intervalos é um tanto imprecisa, segundo Bolfarine, Rodrigues e Achcar (1991) a metodologia de dados agrupados se baseia na tabela de vida, dessa forma, o número de intervalos é arbitrário, de acordo com o interesse do pesquisador, sabendo que em cada intervalo haja pelo menos uma falha. Para experimentos realizados em ciclos, os tempos de vida são discretos e os tempos são agrupados de modo que os tempos iguais estejam no mesmo no intervalo (CHALITA, 1997), Aranda-Ordaz (1983) leva em consideração o tamanho da amostra para o número de intervalos. Porém Chalita (1997) ressalta que número grande de intervalos causa vícios na estimação.

Os intervalos são construídos de tal forma que o eixo do tempo é dividido em k intervalos definidos através dos pontos de corte $a_1, \dots, a_k$, assim o j -ésimo intervalo é denotado pela expressão $I_j = [a_{j-1}, a_j)$, para $j = 1, \dots, k$.

Em muitas aplicações, tem-se o interesse em estudar o efeito de outras variáveis que são referidas na literatura como covariáveis. Tais covariáveis podem ser, por exemplo, o peso de uma pessoa, a idade, um tratamento, entre outras possibilidades. Dessa forma, é importante explorar o relacionamento entre o tempo de vida e suas covariáveis associadas. Uma abordagem que pode ser utilizada para estudar tal relacionamento é baseada nos modelos de regressão.

Considere $(T_i : 1 \leq i \leq n)$ uma amostra aleatória de tamanho n e seja $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ um vetor p -dimensional de covariáveis associados ao i -ésimo indivíduo. Considere que os tempos de vida T_i sejam agrupados em k intervalos mutuamente exclusivos, $I_j = [a_{j-1}, a_j)$, $j = 1, \dots, k$, com $0 = a_0 < a_1 < \dots < a_k = \infty$. A probabilidade de falha do i -ésimo indivíduo, condicionada a covariável $\mathbf{x}_i$ e a sobrevivência no início do j -ésimo intervalo é dada por

$$p_j(\mathbf{x}_i) = P(T_i < a_j | T_i \geq a_{j-1}, \mathbf{x}_i), \quad (3.1)$$

em termos da função de sobrevivência $S(\cdot)$, a expressão (3.1) é escrita como

$$p_j(\mathbf{x}_i) = P(a_{j-1} \leq T_i < a_j | T_i \geq a_{j-1}, \mathbf{x}_i) = 1 - \frac{S(a_j)}{S(a_{j-1})}. \quad (3.2)$$

Por outro lado, a probabilidade de um indivíduo não falhar no j -ésimo intervalo é definido por

$$P(T_i \geq a_j | T_i \geq a_{j-1}) = 1 - P(T_i < a_j | T_i \geq a_{j-1}, \mathbf{x}_i) = 1 - p_j(\mathbf{x}_i), \quad (3.3)$$

em termos da função de sobrevivência, tem-se

$$1 - p_j(\mathbf{x}_i) = \frac{S(a_j)}{S(a_{j-1})}.$$

A estrutura de regressão representada através da probabilidade $p_j(\mathbf{x}_i)$ pode ser modelada utilizando diferentes funções de ligação, em geral é dada por

$$p_j(\mathbf{x}_i) = g(\eta_{ij}) \quad \text{para } i = 1, \dots, n \quad \text{e } j = 1, \dots, k, \quad (3.4)$$

em que $\eta_{ij} = \gamma_j + \mathbf{x}_i^T \boldsymbol{\beta}$ e $g(\cdot)$ é uma função estritamente monótona e duplamente diferenciável que relaciona as variáveis independentes com um preditor linear, γ_j é o efeito do intervalo no modelo e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros associados a cada covariável.

As funções de ligação mais usadas em modelos lineares generalizados são, a função logito, complemento log-log, log-log e probito. Aranda-Ordaz (1983) assumiu uma função de ligação mais geral, conhecida com o nome de ligação de Aranda-Ordaz, na qual a ligação logito e a ligação complemento log-log são casos particulares dessa ligação. Collett (2003) exemplificou a metodologia de dados agrupados para por meio da ligação complemento log-log, utilizada para linearizar o modelo de riscos proporcionais modificado para situações em que os tempos de vida são agrupados ou discretos. Muthén e Masyn (2005) assumiram a ligação logito para modelar os dados de tempo discreto. Singer e Willett (1993), Moreira (2001), Colosimo e Giolo (2006), entre outros, utilizaram as ligações logito e complemento log-log com a finalidade de comparar as duas funções de ligação. Observa-se por esses estudos que as estimativas dos parâmetros por meio dessas funções de ligação são semelhantes, exceto quando observa-se o erro padrão das estimativas.

Dessa forma, pode-se observar que os trabalhos envolvendo dados de sobrevivência agrupados utilizam com frequência as ligações logito e complemento log-log. A razão pelo qual essas ligações são constantemente utilizadas se deve a simples interpretação dos parâmetros que as mesmas produzem. Entretanto, McCullagh e Nelder (1989) e Paula (2004) citaram outras duas ligações no contexto de modelos lineares generalizados, a ligação log-log e a ligação probito. Para Andrade (2007) e Demétrio (2002) a especificação incorreta de uma função de ligação pode causar problemas na interpretação dos resultados.

Por essa razão, neste trabalho foi proposto utilizar, além das ligação logito e complemento log-log, usuais na análise de dados de sobrevivência agrupados, a ligação log-log e ligação probito. Logo, as escolhas da função $g(\cdot)$ neste trabalho são de particular interesse. A função $g(\cdot)$ assume as seguintes formas

- **Ligação logito**

A ligação logito foi introduzido por Berkson (1944) e é muito utilizado. Segundo Souza (2006) o modelo logito é comumente utilizado, uma vez que permite simples interpretação de seus parâmetros através da razão de chances. A ligação é definida como

$$g(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}.$$

- **Ligação complemento log-log**

A ligação complemento log-log é definida como

$$g(\eta_{ij}) = 1 - \exp[-\exp(\eta_{ij})].$$

- **Ligação log-log**

A ligação log-log escrita como

$$g(\eta_{ij}) = \exp[-\exp(-\eta_{ij})].$$

- **Ligação probito**

O modelo probito é uma especificação popular do modelo linear generalizado, usando a função de ligação probito, o modelo foi introduzido por Bliss (1935) e é amplamente utilizado em diversos campos, principalmente em ensaios clínicos. O modelo é denotado por

$$g(\eta_{ij}) = \Phi(\eta_{ij}),$$

em que $\Phi(\cdot)$ é função de distribuição acumulada (FDA) da distribuição normal padrão.

3.3 Estimador de máxima verossimilhança

A função de verossimilhança é escrita em termos da probabilidade de falha do i -ésimo indivíduo no j -ésimo intervalo (I_j), dado que o indivíduo sobreviveu em a_{j-1} conforme definida em (3.2).

Então, a função de verossimilhança pode ser obtida considerando as covariáveis $\mathbf{x}_i$ tal que a contribuição do i -ésimo indivíduo no j -ésimo intervalo para a função de verossimilhança é dado por

- Se o i -ésimo indivíduo falhou no j -ésimo intervalo, sua contribuição na função de verossimilhança é dada por $p_j(\mathbf{x}_i) = 1 - S(a_j)/S(a_{j-1})$.
- Se o i -ésimo indivíduo sobreviveu (isto é, esta em risco) no j -ésimo intervalo, sua contribuição na função de verossimilhança é dada por $1 - p_j(\mathbf{x}_i) = S(a_j)/S(a_{j-1})$.
- Se o i -ésimo indivíduo censurou no j -ésimo intervalo, sua contribuição na função de verossimilhança é dada por $1 - p_j(\mathbf{x}_i) = S(c_i)/S(a_{j-1})$ em que $c_i \in I_j$.

Assim, a função de verossimilhança considerando o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ é dado por

$$L(\boldsymbol{\theta}) = \prod_{j=1}^k \left\{ \prod_{i \in F_j} [p_j(\mathbf{x}_i)] \prod_{i \in R_j} [1 - p_j(\mathbf{x}_i)] \prod_{i \in C_j} [S_i(c_i)/S_i(a_{j-1})] \right\}. \quad (3.5)$$

em que F_j denota o conjunto de indivíduos que falharam no j -ésimo intervalo, R_j denota o conjunto de indivíduos que sobreviveram no j -ésimo intervalo e C_j denota o conjunto de indivíduos que foram censurados no j -ésimo intervalo.

Segundo Thompson (1977), na prática a expressão (3.5) é complicado de ser utilizada, pois os tempos de c_i são desconhecidos, se os dados foram agrupados em intervalos.

Nesta situação, uma alternativa é associar ao tempo de censura c_i uma distribuição uniforme no intervalo $[a_{j-1}, a_j]$ e aproximar a taxa de falha como sendo constante sobre o j -ésimo intervalo. Após manipulações algébricas (ver apêndice B1), a contribuição dos indivíduos censurados no j -ésimo intervalo é dado por

$$\frac{S(c_i)}{S(a_{j-1})} = \left[\frac{S(a_j)}{S(a_{j-1})} \right]^{1/2}. \quad (3.6)$$

Agora substituindo a equação (3.2) em (3.6) temos que

$$\frac{S(c_i)}{S(a_{j-1})} = [1 - p_j(\mathbf{x}_i)]^{1/2}. \quad (3.7)$$

Segundo Colosimo e Giolo (2006), a interpretação da quantidade $1/2$ é que observações para as quais a censura ocorreu no intervalo I_j são tratadas como se estivessem sob risco durante a metade do intervalo considerado.

Substituindo a equação (3.7) em (3.5), a função de verossimilhança é expressa por

$$L(\boldsymbol{\theta}) = \prod_{j=1}^k \left\{ \prod_{i \in F_j} [p_j(\mathbf{x}_i)] \prod_{i \in R_j} [1 - p_j(\mathbf{x}_i)] \prod_{i \in C_j} [1 - p_j(\mathbf{x}_i)]^{(1/2)} \right\} \quad (3.8)$$

Finalmente o logaritmo da função de verossimilhança considerando o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ é dado por

$$l(\boldsymbol{\theta}) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \log[p_j(\mathbf{x}_i)] + \sum_{i \in R_j} \log[1 - p_j(\mathbf{x}_i)] + \frac{1}{2} \sum_{i \in C_j} \log[1 - p_j(\mathbf{x}_i)] \right\}. \quad (3.9)$$

Observe que as probabilidades $p_j(\mathbf{x}_i)$ podem ser modeladas utilizando diferentes funções de ligação, ver seção (1.2.1).

Sabe-se que os estimadores de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$, respectivamente, são obtidos a partir da solução do sistema de equações $U(\beta_m) = 0$, em que $m = 1, \dots, p$ e $U(\gamma_j) = 0$, em que $j = 1, \dots, k$. As equações $U(\beta_m) = 0$ e $U(\gamma_j) = 0$ são não lineares e, em geral, não possuem soluções exatas. Soluções aproximadas são encontradas por métodos iterativos, por exemplo, Newton-Raphson ou um algoritmo de Quase-Newton. Para os dados de sobrevivência agrupados, Lawless (2003) sugere como estimativa inicial para os parâmetros $\boldsymbol{\gamma}'s$, a quantidade $(n_j - d_j)/n_j$,

em que n_j é o número de indivíduos em risco no j -ésimo intervalo e d_j é o números de falhas no j -ésimo intervalo.

Considerando as diferentes funções de ligação a função escore assume as seguintes formas

- **Ligação logito**

$$U(\beta_m) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \left[\frac{x_{im}}{1 + \exp(\eta_{ij})} \right] - \sum_{i \in R_j} \left[\frac{x_{im} \exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{x_{im} \exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \right] \right\}.$$

$$U(\gamma_j) = \sum_{i \in F_j} \left[\frac{1}{1 + \exp(\eta_{ij})} \right] - \sum_{i \in R_j} \left[\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \right].$$

- **Ligação complemento log-log**

$$U(\beta_m) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \left[\frac{ax_{im} \exp(\eta_{ij})}{1 - a} \right] - \sum_{i \in R_j} [x_{im} \exp(\eta_{ij})] - \frac{1}{2} \sum_{i \in C_j} [x_{im} \exp(\eta_{ij})] \right\}.$$

$$U(\gamma_j) = \sum_{i \in F_j} \left[\frac{a \exp(\eta_{ij})}{1 - a} \right] - \sum_{i \in R_j} [\exp(\eta_{ij})] - \frac{1}{2} \sum_{i \in C_j} [\exp(\eta_{ij})].$$

- **Ligação log-log**

$$U(\beta_m) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} [x_{im} \exp(-\eta_{ij})] - \sum_{i \in R_j} \left[\frac{bx_{im} \exp(-\eta_{ij})}{1 - b} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{bx_{im} \exp(-\eta_{ij})}{1 - b} \right] \right\}.$$

$$U_{\gamma_j} = \sum_{i \in F_j} [\exp(-\eta_{ij})] - \sum_{i \in R_j} \left[\frac{b \exp(-\eta_{ij})}{1 - b} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{b \exp(-\eta_{ij})}{1 - b} \right].$$

• **Ligação proibito**

$$U(\beta_m) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \left[\frac{x_{im} f(\eta_{ij})}{\Phi(\eta_{ij})} \right] - \sum_{i \in R_j} \left[\frac{x_{im} f(\eta_{ij})}{1 - \Phi(\eta_{ij})} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{x_{im} f(\eta_{ij})}{1 - \Phi(\eta_{ij})} \right] \right\}.$$

$$U_{\gamma_j} = \sum_{i \in F_j} \left[\frac{f(\eta_{ij})}{\Phi(\eta_{ij})} \right] - \sum_{i \in R_j} \left[\frac{f(\eta_{ij})}{1 - \Phi(\eta_{ij})} \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{f(\eta_{ij})}{1 - \Phi(\eta_{ij})} \right],$$

em que $\eta_{ij} = \gamma_j + \mathbf{x}_i^T \boldsymbol{\beta}$, $f(\cdot)$ e $\Phi(\cdot)$ são a função densidade de probabilidade (fdp) e FDA de uma distribuição normal padrão, respectivamente.

Neste trabalho, a linguagem matricial de programação Ox (DOORNIK, 2007) foi utilizada para calcular as estimativas de máxima verossimilhança por meio da sub-rotina MAXBFGS (ver apêndice B2). No caso em que o tamanho da amostra é grande e sob certas condições de regularidade para a função de verossimilhança, intervalos de confiança e testes de hipóteses podem ser obtidos utilizando o fato de que a distribuição assintótica para os estimadores de máxima verossimilhança é a distribuição normal com média $\boldsymbol{\theta}$ e matriz de covariância dada pela inversa da matriz de informação de Fisher. Portanto, a matriz de covariância assintótica é dada por $\mathbf{I}^{-1}(\boldsymbol{\theta})$ em que $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ e $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$.

Como não é possível determinar a matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\theta})$ devido a presença de observações censuradas, uma forma alternativa é utilizar a matriz Hessiana $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, avaliada no estimador de máxima verossimilhança $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e também conhecida como matriz de informação observada. A distribuição normal assintótica de $\hat{\boldsymbol{\theta}}$ é expressa como

$$\hat{\boldsymbol{\theta}}^T \sim N_{(p+k)}\{\boldsymbol{\theta}^T; \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\},$$

em que $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ é a matriz de informação observada $(p+k)(p+k)$ expressa por

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{\beta\beta} & \mathbf{L}_{\beta\gamma} \\ \cdot & \mathbf{L}_{\gamma\gamma} \end{pmatrix}$$

com as correspondentes submatrizes descrita no apêndice B3.

Assim, o intervalo de confiança assintótico para β_m com $m = 1, \dots, p$ considerando $(1 - \alpha)100\%$ de confiança é expresso por

$$\hat{\beta}_m \pm z_{\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\beta}_m)}$$

em que $z_{\alpha/2}$ é o valor que é excedido com probabilidade $\alpha/2$ para a distribuição normal padrão.

Segundo Chalita (1997), a estimativa do efeito dos intervalos γ_j é interpretada como o efeito sobre a probabilidade de sobrevivência $(1 - p_j(\mathbf{x}_i))$. As estimativas dos parâmetros γ_j , no caso da

ligação complemento log-log, é utilizada para obter as estimativas de $p_j(\mathbf{x}_i)$, que pode ser interpretada como a probabilidade de ocorrência do evento de interesse no j -ésimo intervalo, para um paciente sobre uma dada covariável (COLLETT, 2003).

3.4 Estimador Jackknife

O método Jackknife foi introduzido por Quenouille (1949) para reduzir o vício de um estimador de correlação, como base na divisão da amostra original em duas semi-amostras e melhorado posteriormente com o objetivo de reduzir o vício de um estimador e gerar procedimentos de testes razoavelmente poderosos e com grande aplicabilidade para problemas em que os procedimentos clássicos são sensíveis a não-normalidade da população (EFRON; TIBISHIRANI, 1993). O método Jackknife produz melhores estimativas amostrais de alguns parâmetros de posição e fornece variância e intervalos de confiança aproximados para as estimativas.

Um importante trabalho de implementação do método Jackknife é dado por Lipsitz, Laird e Harrington (1990) que sugerem um estimador robusto alternativo da matriz de covariância baseada no Jackknife para análise de dados de um estudo de medidas repetidas.

Suponha $T_1, \dots, T_n$ uma amostra aleatória de tamanho n e a média amostral dada por

$$\bar{T} = \sum_{i=1}^n \frac{T_i}{n},$$

que é utilizada para estimar a média da população.

Assim, a média amostral sem a l -ésima observação é obtida pela seguinte expressão

$$\bar{T}_{-l} = \frac{\sum_{i=1}^n T_i - T_l}{n-1}.$$

Dessa forma, obtém-se

$$T_l = n\hat{T} - (n-1)\bar{T}_{-l}. \quad (3.10)$$

Em geral, considera-se que θ é um parâmetro estimado por $\hat{E}(T_1, \dots, T_n)$. Finalmente, é calculado $\hat{E}_{-l}$ que é obtido sem a l -ésima observação. Segue da equação 3.10 que os pseudo-valores podem ser obtidos por meio da seguinte expressão

$$\hat{E}_l^* = n\hat{E} - (n-1)\hat{E}_{-l}, \quad l = 1, \dots, n.$$

A média dos pseudo-valores é dado por

$$\hat{E}^* = \frac{\sum_{l=1}^n \hat{E}_l^*}{n}$$

que é o estimador Jackknife de θ .

Manly (1997) sugere que um intervalo de confiança ao nível de significância de $100(1 - \alpha)\%$ aproximado para θ é dado por $\hat{E}^* \pm t_{\alpha/2, n-1} s / \sqrt{n}$, em que $t_{\alpha/2, n-1}$ é o valor que é excedido com probabilidade $\alpha/2$ para a distribuição t com $(n - 1)$ graus de liberdade e o estimador Jackknife tem o efeito de retirar o vício de ordem $1/n$.

3.5 Critérios seleção de modelos

Critérios de informação são fundamentadas na teoria da decisão que penalizam os modelos com um grande número de parâmetros. Alguns desses critérios são baseados nas estatísticas AIC (informação de Akaike), BIC (critério de informação bayesiana) e CAIC (critério de informação de Akaike consistente) definido pelas seguintes expressões,

- $AIC = -2 \log(l) + 2p$,
- $BIC = -2 \log(l) + p \log(n)$,
- $CAIC = -2 \log(l) + p[\log(n) + 1]$,

em que l é o máximo do logaritmo da função de verossimilhança denotado por $l(\hat{\theta})$, p é o número de parâmetros no modelo e n é o tamanho da amostra.

Quanto menor for o valor dessas estatísticas, maior evidência favorável ao modelo em questão. O critério BIC penaliza mais fortemente modelos com um maior número de parâmetros do que o critério AIC tendendo, desta forma, a selecionar modelos com um menor número de parâmetros. Maiores detalhes sobre critérios de seleção podem ser encontrados em Nobre (2004), Paula (2004), Verbeke e Molenberghs (1997), entre outros.

3.6 Análise de sensibilidade

A análise de sensibilidade é um conjunto de métodos de diagnósticos que tem por objetivo, em uma análise de regressão, detectar anomalias no modelo ajustado, por exemplo, presença de *outliers* e observações influentes que podem causar importantes distorções nos resultados da análise. A análise de sensibilidade também é importante para avaliar a robustez do modelo sujeito à algum tipo de perturbação.

Dentre os métodos de diagnóstico se destacam a influência global proposta por Cook (1977), a medida é baseada em deleção de casos e tem por objetivo estudar o efeito da exclusão do i -ésimo indivíduo de um conjunto de dados. Entretanto, as medidas de influência global não permitem dizer em que aspecto específico do modelo o indivíduo tem influência.

Como um método alternativo, Cook (1986) propõe na área de diagnóstico de influência, avaliar a influência conjunta das observações sob pequenas perturbações no modelo ou aos dados, ao invés da avaliação pela retirada individual dos pontos. Esta metodologia é conhecida como influência local se baseia na análise do afastamento da verossimilhança e tem por finalidade avaliar como os resultados da análise são modificados sob pequenas perturbações inseridas no modelo e se essas perturbações apresentam interpretações específicas.

Inúmeras aplicações de diagnósticos de influência podem ser encontrados na literatura estatística para vários modelos de regressão, por exemplo, Cook e Weisberg (1982) e Carroll e Ruppert (1993) descreveram métodos de influência baseados na deleção de casos em modelos de regressão linear, Ortega (2001) apresentou técnicas de influência global e local em um modelo de tempo de vida e Paula (2004) descreveu a metodologia de influência local para modelos de regressão linear. Souza(2006) apresentou as técnicas de influência local nos modelos de regressão logística. A seguir descreve-se os métodos de influência global e influência local.

3.6.1 Influência Global

A primeira ferramenta para realizar análise de sensibilidade como declarado antes é por meio do que denomina-se influência global a partir da deleção de casos (COOK, 1977). Deleção de casos é uma abordagem comum para estudar o efeito da retirada do i -ésimo caso do conjunto de dados. A deleção de casos para o modelo (3.1) é dado por

$$p_j(\mathbf{x}_{(i)}) = P(T_{(i)} < a_j | T_{(i)} \geq a_{j-1}, \mathbf{x}_{(i)}), \quad i = 1, \dots, n, \quad (3.11)$$

em que o subscrito "(i)" significa que o i -ésimo caso foi retirado da amostra. Para o modelo (3.11), o logaritmo da função de verossimilhança é denotado por $l_{(i)}(\boldsymbol{\theta})$.

Seja $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\boldsymbol{\gamma}}_{(i)}^T, \hat{\boldsymbol{\beta}}_{(i)}^T)^T$, as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ de $l_{(i)}(\boldsymbol{\theta})$. Para avaliar a influência do i -ésimo caso sobre o estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}^T, \hat{\boldsymbol{\beta}}^T)^T$, a idéia básica é comparar a diferença entre $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}$. Se a retirada de um indivíduo influencia seriamente as estimativas, maior atenção deverá ser dada àquele caso. Conseqüentemente, se $\hat{\boldsymbol{\theta}}_{(i)}$ é distante de $\hat{\boldsymbol{\theta}}$, então o indivíduo é considerado como uma observação influente. Uma primeira medida de influência global é definida como a norma padronizada de $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$, também conhecida como distância generalizada de Cook e é dada por

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T [-\ddot{\mathbf{L}}(\boldsymbol{\theta})] (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}),$$

em que $\hat{\boldsymbol{\theta}}$ é o vetor de estimativas da amostra completa, $\hat{\boldsymbol{\theta}}_{(i)}$ é o vetor de estimativas sem a i -ésima observação e $\ddot{\mathbf{L}}$ é a matriz de informação observada. Outra medida popular da diferença entre $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}$, é o que se denomina de afastamento da verossimilhança e é expresso por

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

em que $l(\hat{\theta})$ é o logaritmo da função de verossimilhança da amostra completa e $l(\hat{\theta}_{(i)})$ é o logaritmo da função de verossimilhança sem a i -ésima observação.

As medidas propostas por Wang et al. (1996) foram utilizadas no modelo de regressão para dados agrupados. Define-se as medidas $IE(\theta)$, $IE(\gamma)$ e $IE(\beta)$ para cada indivíduo da seguinte forma

$$IE(\theta)_i = \frac{1}{p+k} \sum_{v=1}^{p+k} \frac{|\hat{\theta}_v - \hat{\theta}_{(i)v}|}{EP(\hat{\theta}_v)}, \quad IE(\gamma)_i = \frac{1}{k} \sum_{j=1}^k \frac{|\hat{\gamma}_j - \hat{\gamma}_{(i)j}|}{EP(\hat{\gamma}_j)},$$

e

$$IE(\beta)_i = \frac{1}{p} \sum_{m=1}^p \frac{|\hat{\beta}_m - \hat{\beta}_{(i)m}|}{EP(\hat{\beta}_m)},$$

em que $EP(\cdot)$ é erro padrão associado as estimativas, θ é o vetor de parâmetros completo, γ é o vetor de parâmetros somente dos efeitos do intervalo e β é o vetor de parâmetros dos coeficientes da regressão.

Segundo Wang et al. (1996), a medida de influência global $IE(\cdot)_i$ é interpretada como uma mudança do coeficiente da média relativa para um conjunto de estimativas e também é útil para avaliar o efeito das estimativas dos parâmetros por exclusão da i -ésima observação. Se os valores das medidas de influência global para algum indivíduo forem relativamente grandes, há fortes indícios de que tal indivíduo seja uma observação influente e que pode causar instabilidade no modelo ajustado.

3.6.2 Influência Local

A idéia principal de influência local é verificar, através de alguma medida apropriada de influência, o efeito de pequenas perturbações no modelo ou nos dados. Se essas perturbações causarem efeitos desproporcionais em determinados resultados do modelo, podem ser indícios de que o modelo está mal ajustado ou que existem afastamentos importantes das suposições feitas para o mesmo. A identificação das observações responsáveis por essas discrepâncias pode ajudar na escolha de um modelo mais adequado aos dados. A medida de influência mais conhecida é o afastamento da verossimilhança $LD(\omega) = 2[(\hat{\theta}) - l(\hat{\theta}_\omega)]$, em que $\hat{\theta}_\omega$ denota a estimativa de máxima verossimilhança sob o modelo perturbado e $\omega = (\omega_1, \dots, \omega_q)^T$ é o vetor de perturbações aplicados ao modelo. A proposta de Cook (1986) é estudar o comportamento de $LD(\omega)$ em torno do vetor de não perturbação ω_0 . Tem-se que $LD(\omega = 0)$. Logo, desde que $LD(\omega) \geq 0$, ω_0 é um ponto de mínimo da função $LD(\omega)$. A sugestão de Cook (1986) é investigar a curvatura normal da linha projetada $LD(\omega_0 + a\mathbf{d})$, em que $a \in \mathbb{R}$, em torno de $a = 0$ para alguma direção arbitrária $\mathbf{d}$, $\|\mathbf{d}\| = 1$.

Assim, a curvatura normal é definida por $C_d(\theta) = 2|\mathbf{d}^T \Delta [\ddot{L}(\theta)]^{-1} \Delta \mathbf{d}|$, em que Δ é uma matriz $(p \times q)$ com elementos $\Delta_{ij} = \partial^2 l(\theta|\omega) / \partial \theta_j \partial \omega_i$, $i = 1, \dots, n$ e $j = 1, \dots, p+k$, com

todas as qualidades sendo avaliadas em $\omega = \omega_0$ e $\theta = \hat{\theta}_0$. Cook (1986) sugere tomar a direção correspondente à maior curvatura, denotada por $\mathbf{d}_{\max}$, que é o maior autovetor e o correspondente maior autovalor $C_{\mathbf{d}_{\max}}$, da matriz $\ddot{F} = \Delta^T [\ddot{\mathbf{L}}(\theta)]^{-1} \Delta$.

Por outro lado, Lesaffre e Verbeke (1998), sugerem avaliar a curvatura normal na direção da i -ésima observação que consiste na avaliação de $C_{\mathbf{d}}(\theta)$ em que $\mathbf{d}$ é um vetor $n \times 1$ formado por zeros com um na i -ésima posição. Essa curvatura é denominada por C_i . Sugere-se que as observações tais que $C_i > 2\bar{C}$, em $\bar{C} = \sum_{i=1}^n C_i/n$ tenham uma atenção especial.

Portanto, os gráficos dos elementos de $\mathbf{d}_{\max}$ *versus* o índice das observações bem como os gráficos dos elementos C_i *versus* o índice das observações são utilizadas para detectar possíveis pontos influentes.

3.6.3 Cálculo das curvaturas máximas

As medidas de influência são calculadas após um esquema de perturbação ser adotado modelo ou nos dados. Diferentes esquemas de perturbação podem ser considerados de acordo com o objetivo da análise estatística. Em seguida, calcula-se para os esquemas de perturbação adotados a matriz de curvaturas máximas, expressa por

$$\Delta = (\Delta_{vi})_{(p+k) \times n} = \left(\frac{\partial^2 l(\theta|\omega)}{\partial \theta_v \partial \omega_i} \right)_{(p+k) \times n}, \quad v = 1, \dots, p+k \quad \text{e} \quad i = 1, \dots, n,$$

considerando o modelo definido em (3.1) e o logaritmo da função de verossimilhança dado por (3.9). Considere também o vetor de ponderação $\omega = (\omega_1, \dots, \omega_n)^T$.

A seguir apresenta-se o logaritmo da função de verossimilhança perturbada para dois esquemas de perturbação bem como o cálculo das curvaturas.

Perturbação de casos

Nesse caso, a função de verossimilhança perturbada é dada por

$$l(\theta|\omega) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \omega_i \log[g(\eta_{ij})] + \sum_{i \in R_j} \omega_i \log[1 - g(\eta_{ij})] + \frac{1}{2} \sum_{i \in C_j} \omega_i \log[1 - g(\eta_{ij})] \right\},$$

em que $\eta_{ij} = \gamma_j + \mathbf{x}_i^T \beta$, $0 \leq \omega_i \leq 1$ e $\omega_0 = (1, \dots, 1)^T$ é o vetor de não perturbação.

Seja $\Delta = (\Delta_\beta, \Delta_\gamma)^T$, logo os elementos da matriz Δ_β , para $m = 1, \dots, p$ são dados por

$$\Delta_{mi} = \begin{cases} - \sum_{l=1}^{j-1} \left\{ \frac{[\dot{g}(\eta_{il})]_{\beta_m}}{1 - \hat{g}(\eta_{il})} \right\} + \frac{[\dot{g}(\eta_{ij})]_{\beta_m}}{\hat{g}(\eta_{ij})} & \text{se } i \in F_j, \\ - \sum_{l=1}^{j-1} \left\{ \frac{[\dot{g}(\eta_{il})]_{\beta_m}}{1 - \hat{g}(\eta_{il})} \right\} - \frac{1}{2} \left\{ \frac{[\dot{g}(\eta_{ij})]_{\beta_m}}{1 - \hat{g}(\eta_{ij})} \right\} & \text{se } i \in C_j. \end{cases}$$

Assim, especificando para cada função de ligação, tem-se que, os elementos da curvatura Δ_{mi} podem ser expressos da seguinte forma

- **Ligação logito**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \frac{\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \quad \text{e} \quad [\dot{g}(\eta_{il})]_{\beta_m} = \frac{x_{im} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \quad \text{e} \quad [\dot{g}(\eta_{ij})]_{\beta_m} = \frac{x_{im} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2}.$$

- **Ligação complemento log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = 1 - \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\dot{g}(\eta_{il})]_{\beta_m} = x_{im} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = 1 - \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

- **Ligação log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\dot{g}(\eta_{il})]_{\beta_m} = x_{im} \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\}.$$

- **Ligação proibito**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \Phi(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{e} \quad [\dot{g}(\eta_{il})]_{\beta_m} = x_{im} f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}).$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \Phi(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{e} \quad [\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}).$$

em que $f(\cdot)$ é a fdp da normal padrão e $\Phi(\cdot)$ é a FDA da normal padrão.

Os elementos da matriz $\Delta_{\boldsymbol{\gamma}}$, para $j = 1, \dots, k$, pode ser expressado como

$$\Delta_{ji} = \begin{cases} \left\{ \frac{1}{\hat{g}(\eta_{ij})} [\dot{g}(\eta_{ij})]_{\gamma_j} \right\} & \text{se } i \in F_j, \\ -\frac{1}{2} \left\{ \frac{1}{1-\hat{g}(\eta_{ij})} [\dot{g}(\eta_{ij})]_{\gamma_j} \right\} & \text{se } i \in R_j, \\ -\left\{ \frac{1}{1-\hat{g}(\eta_{ij})} [\dot{g}(\eta_{ij})]_{\gamma_j} \right\} & \text{se } i \in C_j. \end{cases}$$

Considerando as funções de ligação tem-se que

- **Ligação logito**

$$\hat{g}(\eta_{ij}) = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \quad \text{e} \quad [\dot{g}(\eta_{ij})]_{\gamma_j} = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2}.$$

- **Ligação complemento log-log**

$$\hat{g}(\eta_{ij}) = 1 - \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\dot{g}(\eta_{ij})]_{\gamma_j} = \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

- **Ligação log-log**

$$\hat{g}(\eta_{ij}) = \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\dot{g}(\eta_{ij})]_{\gamma_j} = \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\}.$$

- **Ligação proibito**

$$\hat{g}(\eta_{ih}) = \Phi(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{e} \quad [\dot{g}(\eta_{ij})]_{\gamma_j} = f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}).$$

em que $f(\cdot)$ é a fdp da normal padrão e $\Phi(\cdot)$ é a FDA da normal padrão.

Perturbação de uma covariável

Considere agora, uma perturbação aditiva em uma particular covariável contínua, denotada por X_t , fazendo $x_{it\omega} = x_{it} + \omega_i S_x$, em que S_x é um fator escala, no geral, considera-se S_x como o desvio padrão da covariável e $\omega_i \in \mathbf{R}$. Este esquema de perturbação leva a seguinte expressão para a função de verossimilhança e para os elementos da matriz Δ .

Nesse caso, o logaritmo da função de verossimilhança perturbada é dada por

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{j=1}^k \left\{ \sum_{i \in F_j} \log[g(\eta_{ij}^*)] + \sum_{i \in S_j} \log[1 - g(\eta_{ij}^*)] + \frac{1}{2} \sum_{i \in C_j} \log[1 - g(\eta_{ij}^*)] \right\},$$

em que $\eta_{ij}^* = \gamma_j + \mathbf{x}_i^{*T} \boldsymbol{\beta}$, $\mathbf{x}_i^{*T} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_t (x_{it} + \omega_i S_t) + \dots + \beta_p x_{ip}$ e $\boldsymbol{\omega}_0 = (0, 0, \dots, 0)^T$ é o vetor de não perturbação.

Assim, os elementos da matriz $\Delta_{\boldsymbol{\beta}}$ com $m = 1, \dots, p$ e $m \neq t$ são expressos como

$$\Delta_{mi} = \begin{cases} -\sum_{l=1}^{j-1} \left\{ \frac{[\ddot{g}(\eta_{il})]_{\beta_m \omega_0}}{1 - \hat{g}(\eta_{il})} + \frac{[\dot{g}(\eta_{il})]_{\beta_m} [\dot{g}(\eta_{il})]_{\omega_0}}{[1 - \hat{g}(\eta_{il})]^2} \right\} + \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \omega_0}}{\hat{g}(\eta_{ij})} - \frac{[\dot{g}(\eta_{ij})]_{\beta_m} [\dot{g}(\eta_{ij})]_{\omega_0}}{[\hat{g}(\eta_{ij})]^2} \right] & \text{se } i \in F_j, \\ -\sum_{l=1}^{j-1} \left\{ \frac{[\ddot{g}(\eta_{il})]_{\beta_m \omega_0}}{1 - \hat{g}(\eta_{il})} + \frac{[\dot{g}(\eta_{il})]_{\beta_m} [\dot{g}(\eta_{il})]_{\omega_0}}{[1 - \hat{g}(\eta_{il})]^2} \right\} - \left\{ \frac{[\ddot{g}(\eta_{ij})]_{\beta_m \omega_0}}{1 - \hat{g}(\eta_{ij})} + \frac{[\dot{g}(\eta_{ij})]_{\beta_m} [\dot{g}(\eta_{ij})]_{\omega_0}}{[1 - \hat{g}(\eta_{ij})]^2} \right\} & \text{se } i \in C_j. \end{cases}$$

No caso das funções de ligação tem-se que

• Ligação logito

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \frac{\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}, \quad [\dot{g}(\eta_{il})]_{\beta_m} = \frac{x_{im} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2},$$

$$[\dot{g}(\eta_{il})]_{\omega_0} = \frac{\hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2} \text{ e}$$

$$[\ddot{g}(\eta_{il})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \frac{1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^3}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}, \quad [\dot{g}(\eta_{ij})]_{\beta_m} = \frac{x_{im} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2},$$

$$\dot{g}(\eta_{ij})]_{\omega_0} = \frac{\hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2} \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \frac{1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^3}.$$

• **Ligação complemento log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = 1 - \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{\hat{g}}(\eta_{il})]_{\beta_m} = x_{im} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$\dot{\hat{g}}(\eta_{il})]_{\omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\ddot{\hat{g}}(\eta_{il})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \\ [1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = 1 - \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{\hat{g}}(\eta_{ij})]_{\beta_m} = x_{im} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$\dot{\hat{g}}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\ddot{\hat{g}}(\eta_{ij})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \\ [1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

• **Ligação log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{\hat{g}}(\eta_{il})]_{\beta_m} = x_{im} \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$\dot{g}(\eta_{il})_{\omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\ddot{g}(\eta_{il})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \\ \{\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] - 1\}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$\dot{g}(\eta_{ij})_{\omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \omega_0} = x_{im} \hat{\beta}_t S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \\ \{\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] - 1\}$$

• **Ligação proibido**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \Phi(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad [\dot{g}(\eta_{il})]_{\beta_m} = x_{im} f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad \dot{g}(\eta_{il})_{\omega_0} = \hat{\beta}_t S_x f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{e}$$

$$[\ddot{g}(\eta_{il})]_{\beta_m \omega_0} = -x_{im} \hat{\beta}_t S_x (\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}).$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \Phi(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad [\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad \dot{g}(\eta_{ij})_{\omega_0} = \hat{\beta}_t S_x f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \omega_0} = -x_{im} \hat{\beta}_t S_x (\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}),$$

em que $f(\cdot)$ é a fdp da normal padrão e $\Phi(\cdot)$ é a FDA da normal padrão.

Os elementos do vetor Δ_t para $m = t$ são dados por

$$\Delta_{ti} = \begin{cases} -\sum_{l=1}^{j-1} \left\{ \frac{[\ddot{g}(\eta_{il})]_{\beta_t \omega_0}}{1 - \hat{g}(\eta_{il})} + \frac{[\dot{g}(\eta_{il})]_{\beta_t} [\dot{g}(\eta_{il})]_{\omega_0}}{[1 - \hat{g}(\eta_{il})]^2} \right\} + \left\{ \frac{[\ddot{g}(\eta_{ij})]_{\beta_t \omega_0}}{\hat{g}(\eta_{ij})} - \frac{[\dot{g}(\eta_{ij})]_{\beta_t} [\dot{g}(\eta_{ij})]_{\omega_0}}{[\hat{g}(\eta_{ij})]^2} \right\} & \text{se } i \in F_j, \\ -\sum_{l=1}^{j-1} \left\{ \frac{[\ddot{g}(\eta_{il})]_{\beta_t \omega_0}}{1 - \hat{g}(\eta_{il})} + \frac{[\dot{g}(\eta_{il})]_{\beta_t} [\dot{g}(\eta_{il})]_{\omega_0}}{[1 - \hat{g}(\eta_{il})]^2} \right\} - \left\{ \frac{[\ddot{g}(\eta_{ij})]_{\beta_t \omega_0}}{1 - \hat{g}(\eta_{ij})} + \frac{[\dot{g}(\eta_{ij})]_{\beta_t} [\dot{g}(\eta_{ij})]_{\omega_0}}{[1 - \hat{g}(\eta_{ij})]^2} \right\} & \text{se } i \in C_j. \end{cases}$$

No caso das funções de ligação tem-se que

• **Ligação logito**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \frac{\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}, \quad [\dot{\hat{g}}(\eta_{il})]_{\beta_t} = \frac{x_{it} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2},$$

$$[\dot{\hat{g}}(\eta_{il})]_{\omega_0} = \frac{\hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2} \mathbf{e}$$

$$[\ddot{\hat{g}}(\eta_{il})]_{\beta_t \omega_0} = \frac{S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^3} \left\{ 1 + \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) + x_{it} \hat{\beta}_t [1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \right\}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}, \quad [\dot{\hat{g}}(\eta_{ij})]_{\beta_t} = \frac{x_{it} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2}$$

$$[\dot{\hat{g}}(\eta_{ij})]_{\omega_0} = \frac{\hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^2} \mathbf{e}$$

$$[\ddot{\hat{g}}(\eta_{ij})]_{\beta_t \omega_0} = \frac{S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]^3} \left\{ 1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) + x_{it} \hat{\beta}_t [1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \right\}.$$

• **Ligação complemento log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = 1 - \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{\hat{g}}(\eta_{il})]_{\beta_t} = x_{it} \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{\hat{g}}(\eta_{il})]_{\omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \mathbf{e}$$

$$[\ddot{\hat{g}}(\eta_{il})]_{\beta_t \omega_0} = S_x \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \left\{ 1 + x_{it} \hat{\beta}_t [1 - \exp(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \right\}.$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = 1 - \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{g}(\eta_{ij})]_{\beta_t} = x_{it} \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})],$$

$$[\dot{g}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\beta_t \omega_0} = S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})].$$

- **Ligação log-log**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{il})]_{\beta_t} = x_{it} \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{il})]_{\omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\ddot{g}(\eta_{il})]_{\beta_t \omega_0} = S_x \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \\ \{1 - x_{it} \hat{\beta}_t [1 - \exp[-(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]]\}.$$

Para $j = 1, 2, \dots, k$.

$$\hat{g}(\eta_{ij}) = \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{ij})]_{\beta_t} = x_{it} \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\beta_t \omega_0} = S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \\ \{1 - x_{it} \hat{\beta}_t [1 - \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]]\}.$$

- **Ligação proibito**

Para $l = 1, \dots, j - 1$.

$$\hat{g}(\eta_{il}) = \Phi(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad [\dot{g}(\eta_{il})]_{\beta_t} = x_{it} f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}),$$

$$[\dot{g}(\eta_{il})]_{\omega_0} = \hat{\beta}_t S_x f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\beta}) \quad \text{e} \quad [\ddot{g}(\eta_{il})]_{\beta_t \omega_0} = S_x f(\hat{\gamma}_l + \mathbf{x}_i^T \hat{\beta}) [1 - x_{it} \hat{\beta}_t (\hat{\gamma}_l + \mathbf{x}_i^T \hat{\beta})].$$

Para $j = 1, \dots, k$.

$$\hat{g}(\eta_{ij}) = \Phi(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}), \quad [\dot{g}(\eta_{ij})]_{\beta_t} = x_{it} f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}),$$

$$[\dot{g}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) \quad \text{e} \quad [\ddot{g}(\eta_{ij})]_{\beta_t \omega_0} = S_x f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) [1 - x_{it} \hat{\beta}_t (\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})],$$

em que $f(\cdot)$ é a fdp da normal padrão e $\Phi(\cdot)$ é a FDA da normal padrão.

E os elementos da matriz Δ_{γ} com $j = 1, \dots, k$ são expressos como

$$\Delta_{ji} = \begin{cases} \frac{[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0}}{\hat{g}(\eta_{ij})} - \frac{[\dot{g}(\eta_{ij})]_{\gamma_j} [\dot{g}(\eta_{ij})]_{\omega_0}}{[\hat{g}(\eta_{ij})]^2} & \text{se } i \in F_j, \\ - \left\{ \frac{[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0}}{1 - \hat{g}(\eta_{ij})} + \frac{[\dot{g}(\eta_{ij})]_{\gamma_j} [\dot{g}(\eta_{ij})]_{\omega_0}}{[1 - \hat{g}(\eta_{ij})]^2} \right\} & \text{se } i \in R_j, \\ - \frac{1}{2} \left\{ \frac{[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0}}{1 - \hat{g}(\eta_{ij})} + \frac{[\dot{g}(\eta_{ij})]_{\gamma_j} [\dot{g}(\eta_{ij})]_{\omega_0}}{[1 - \hat{g}(\eta_{ij})]^2} \right\} & \text{se } i \in C_j. \end{cases}$$

No caso das funções de ligação tem-se que

- **Ligação logito**

$$\hat{g}(\eta_{ij}) = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})}{1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})}, \quad [\dot{g}(\eta_{ij})]_{\gamma_j} = \frac{\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})]^2}$$

$$[\dot{g}(\eta_{ij})]_{\omega_0} = \frac{\hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})}{[1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})]^2} \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) \frac{1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})}{[1 + \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})]^3}.$$

- **Ligação complemento log-log**

$$\hat{g}(\eta_{ij}) = 1 - \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})],$$

$$[\dot{g}(\eta_{ij})]_{\gamma_j} = \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})],$$

$$[\dot{g}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})] \quad \text{e}$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0} = \hat{\beta}_t S_x \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta}) \exp[-\exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})] [1 - \exp(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\beta})].$$

- **Ligação log-log**

$$\hat{g}(\eta_{ij}) = \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$[\dot{g}(\eta_{ij})]_{\gamma_j} = \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\},$$

$$\dot{g}(\eta_{ij})_{\omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \quad e$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0} = \hat{\beta}_t S_x \exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] \exp\{-\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})]\} \{\exp[-(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})] - 1\}.$$

- **Ligação proibito**

$$\hat{g}(\eta_{ij}) = \Phi(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad [\dot{g}(\eta_{ij})]_{\gamma_j} = f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \quad [\dot{g}(\eta_{ij})]_{\omega_0} = \hat{\beta}_t S_x f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})$$

$$e \quad [\ddot{g}(\eta_{ij})]_{\gamma_j \omega_0} = -\hat{\beta}_t S_x (\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) f(\hat{\gamma}_j + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}),$$

em que $f(\cdot)$ é a fdp da normal padrão e $\Phi(\cdot)$ é a FDA da normal padrão.

3.7 Análise de resíduos

De maneira geral, define-se o resíduo referente a i -ésima observação por meio de uma função que depende da variável resapost e do vetor de parâmetros estimado, que objetiva identificar a discrepância entre o modelo ajustado e o conjunto de dados.

Dados de sobrevivência grupados são complicados de tratar pelo fato das observações incompletas e os tempos grupados em intervalos. Na literatura encontra-se resíduos modificados para incorporar os tempos de censura, por exemplo, Collet (2003), Colosimo e Giolo (2006), Klein e Moeschberger (1997) descreveram os resíduos de Cox-Snell, Martingale e Deviance, usuais na área de análise de sobrevivência.

Sabe-se que a metodologia de modelos de regressão grupados baseia-se nos métodos de análise binária, por esta razão este trabalho propõe como uma medida de resíduos, o resíduo de Pearson.

Logo, adota-se o resíduo de Pearson transformado como uma alternativa para estudar os resíduos, uma vez que se assume a modelagem binária. O resíduo de Pearson é dado pela expressão

$$r_{Pi} = \frac{y_i - \hat{g}(\eta_{ij})}{\sqrt{\hat{g}(\eta_{ij})[1 - \hat{g}(\eta_{ij})]}},$$

em que $\hat{g}(\eta_{ij})$ é uma função de ligação dependendo das estimativas dos parâmetros. O resíduo de Pearson é baseado em modelos de regressão linear logística, porém, aproximações de resíduo podem ser construídas para modelos baseados nas outras funções de ligação, complemento log-log, log-log e proibito (COLLETT, 1991).

3.8 Aplicação

Nesta seção ilustrou-se o modelo de regressão para dados grupados proposto utilizando um conjunto de dados de vitamina A. Os dados são resultados de um estudo realizado por Barreto et al.(1994), cujo objetivo foi avaliar o efeito da suplementação da vitamina A em episódios de diarreia. Para avaliar o efeito da suplementação foi realizado um experimento em uma pequena cidade no nordeste do Brasil entre dezembro de 1990 e dezembro de 1991. A amostra é composta de 1207 crianças em idade pré-escolar, com idade entre 6 a 48 meses no início do experimento, alocadas para receber placebo ou vitamina A. O tempo de sobrevivência foi definido como o tempo da primeira dose de vitamina A até a ocorrência de um episódio de diarreia, ou seja, uma sequência de dias com diarreia. As covariáveis associadas a cada criança foram: idade no início do experimento (em meses), tratamento (0=placebo, 1=vitamina A) e sexo (0= feminino, 1=masculino). Informações adicionais sobre este estudo podem ser encontradas em Silva et. al (2008). Os tempos de vida foram categorizados em oito intervalos $\{(4, 21], (21, 38], \dots, (126, 185]\}$. Não há um critério que estabeleça o número de intervalos, a única suposição que se tem é a presença de pelo menos uma falha em cada intervalo, dessa forma por conveniência adotou-se $k = 8$ intervalos.

Para cada criança, $i = 1, \dots, 1207$, foi associado as seguintes variáveis

- t_i : tempo observado;
- I_j : intervalo de tempo;
- x_{i1} : idade (em meses);
- x_{i2} : tratamento (0 = placebo, 1 = suplementação de vitamina A);
- x_{i3} : sexo (0 = menina, 1 = menino).

Na Tabela 3.1 é apresentado um resumo dos dados de vitamina A. Observa-se que das 1207 crianças em estudo, 925 apresentaram um episódio de diarreia, o que significa que 282 crianças foram censuras aleatoriamente durante o tempo de observação.

Tabela 3.1 – Resumo do número de valores censurados e não censurados

total de observações	número de falhas	número de censuras	porcentagens de censura
1207	925	282	23,36

Com o objetivo de se ter uma idéia do comportamento dos dados de vitamina A foi construído um histograma de frequência do conjunto de dados (Figura 3.1)

Figura 3.1 – Gráfico de frequência para os dados de vitamina A

A Figura 3.1 apresenta uma frequência alta de observações nos tempos iniciais e próximos ao término do estudo.

A seguir foi construído uma tabela de vida considerando oito intervalos.

Tabela 3.2 – Tabela de vida para os dados de vitamina A

Intervalo I_j	número de falhas	número de censuras	número sob risco	$\hat{S}(t)$
[4, 21)	292	0	1207	1,000
[21, 38)	243	4	915	0,758
[38, 55)	138	6	668	0,556
[55, 73)	101	2	524	0,441
[73, 90)	46	3	421	0,356
[90, 108)	49	6	372	0,317
[108, 126)	46	11	317	0,275
[126, 185)	10	250	260	0,234

Os resultados da Tabela 3.2 indicam que há um número considerável de empates e por essa razão, a metodologia de dados de sobrevivência agrupados é uma alternativa para modelar o conjunto de dados de vitamina A.

Dessa forma, o conjunto de dados de vitamina A é analisado sob a perspectiva do modelo de regressão de dados de sobrevivência agrupados utilizando as quatro funções de ligação, expressado por

- **Ligação logito**

$$p_j \mathbf{x}_i = g(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$

$$= \frac{\exp(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})}{1 + \exp(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})},$$

- **Ligação complemento log-log**

$$p_j \mathbf{x}_i = g(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$

$$= 1 - \exp[-\exp(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})],$$

- **Ligação log-log**

$$p_j \mathbf{x}_i = g(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$

$$= \exp\{-\exp[-(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})]\},$$

- **Ligação probito**

$$p_j \mathbf{x}_i = g(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$

$$= \Phi(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}),$$

em que $i = 1, \dots, 1207$, $j = 1, \dots, 8$ e $\Phi(\cdot)$ é FDA da normal padrão

3.8.1 Seleção de modelos

Na Tabela 3.7 são apresentadas as estimativas de máxima verossimilhança com cada uma das funções de ligação. As estimativas que apresentaram menor erro padrão foi dado pelo modelo com função de ligação.

Tabela 3.3 – Estimativas de máxima verossimilhança (erro padrão) para os parâmetros do modelo de regressão para dados grupados de vitamina A e os valores dos critérios de seleção de modelos, usando as quatro funções de ligação

Parâmetro	Estimativa(erro padrão)			
	Ligação logito	Ligação complemento log-log	Ligação log-log	Ligação probito
γ_1	-0,079(0,119)	-0,371(0,099)	0,206(0,065)	-0,081(0,070)
γ_2	0,104(0,127)	-0,213(0,106)	0,299(0,070)	0,025(0,075)
γ_3	-0,174(0,143)	-0,448(0,122)	0,149(0,076)	-0,140(0,084)
γ_4	-0,212(0,156)	-0,489(0,135)	0,138(0,082)	-0,156(0,091)
γ_5	-0,859(0,192)	-1,085(0,174)	-0,132(0,088)	-0,501(0,105)
γ_6	-0,627(0,191)	-0,861(0,174)	-0,058(0,091)	-0,390(0,107)
γ_7	-0,465(0,198)	-0,712(0,176)	0,008 (0,095)	-0,305(0,111)
γ_8	-1,212(0,350)	-1,424(0,331)	-0,255(0,134)	-0,672(0,176)
β_1	-0,037(0,003)	-0,003(0,003)	-0,019(0,002)	-0,021(0,002)
β_2	-0,161(0,076)	-0,141(0,066)	-0,081(0,038)	-0,093(0,043)
β_3	-0,029(0,076)	-0,018(0,066)	-0,024(0,038)	-0,021(0,043)
AIC	4378,86	4378,87	4379,93	4382,26
BIC	4434,91	4434,93	4435,99	4438,31
CAIC	4445,91	4379,88	4446,99	4449,31

Com a finalidade de escolher uma função de ligação, entre a logito, complemento log-log, log-log e probito para seguir a análise, utilizou-se os critérios de AIC, BIC e CAIC.

Os resultados da Tabela 3.7 mostram que as medidas de AIC, BIC e CAIC são similares, porém o menor valor dos critérios de seleção foi observado para a função de ligação logito, exceto quando o critério de CAIC é considerado. Assim, sendo optou-se por continuar analisando os dados da vitamina A utilizando a função de ligação logito.

3.8.2 Estimação de máxima verossimilhança e Jackknife

As Tabelas 3.4 e 3.5 apresentam os resultados das estimativas dos parâmetros utilizando o método de máxima verossimilhança e Jackknife, respectivamente.

Tabela 3.4 – Estimativas de máxima verossimilhança para os parâmetros do modelo de regressão para dados grupados assumindo a função de ligação logito ajustado ao conjunto de dados de vitamina A

Parâmetros	Estimativa	Erro padrão	<i>p</i> -valor
γ_1	-0,079	0,119	0,506
γ_2	0,104	0,127	0,415
γ_3	-0,174	0,143	0,222
γ_4	-0,212	0,156	0,175
γ_5	-0,859	0,192	<0,001
γ_6	-0,627	0,191	0,001
γ_7	-0,465	0,198	0,019
γ_8	-1,212	0,350	0,001
β_1	-0,037	0,003	0,000
β_2	-0,161	0,076	0,033
β_3	-0,029	0,076	0,699

Tabela 3.5 – Estimativas Jackknife para os parâmetros do modelo de regressão para dados grupados assumindo a função de ligação logito ajustado ao conjunto de dados de vitamina A

Parâmetros	Estimativa	Erro padrão	Intervalo de confiança(95%)
γ_1	-0,081	0,123	(-0,323; 0,161)
γ_2	0,104	0,133	(-0,156; 0,364)
γ_3	-0,177	0,148	(-0,467; 0,114)
γ_4	-0,218	0,162	(-0,536; 0,100)
γ_5	-0,858	0,199	(-1,248;-0,467)
γ_6	-0,627	0,197	(-1,013;-0,241)
γ_7	-0,471	0,202	(-0,868;-0,074)
γ_8	-1,164	0,368	(-1,885;-0,442)
β_1	-0,038	0,003	(-0,044;-0,031)
β_2	-0,162	0,076	(-0,311;-0,012)
β_3	-0,029	0,076	(-0,179; 0,120)

Observa-se que as covariáveis x_1 que representa a idade e x_2 que representa o tratamento são significativas no modelo ao nível de significância de 5%. Porém observa-se que a covariável sexo não é significativa. Por meio da Tabela 3.4 e da Tabela 3.5, notou-se que os dois métodos de estimação apresentam resultados similares.

3.8.3 Análise de influência global

Nesta subseção, a linguagem de programação Ox foi utilizada para calcular as medidas de deleção de casos $GD_i(\theta)$ e $LD_i(\theta)$ apresentados na subseção 3.6.1. Os resultados dessas medidas de influência são mostradas na Figura 3.2, na qual observou-se que os indivíduos #128 e #438 são possíveis observações influentes.

As observações #128 e #438 representam indivíduos grupados no oitavo intervalo com a menor idade no conjunto de observações que foram falha no tratamento com suplementação de vitamina A e do sexo feminino.

Na Figura 3.3 é apresentada as medidas de estimativas influentes de θ , β e γ , na qual a observação #461 destacou-se como ponto influente. A observação #461 representa o indivíduo grupado no oitavo intervalo com uma das menores idade no conjunto de observações que foram censuradas recebendo o tratamento com o placebo e do sexo feminino.

Figura 3.2 – Gráficos de medidas de influência global para o modelo de regressão com a função de ligação logito ajustado aos dados de vitamina A. (a) Gráfico da distância de Cook generalizada e (b) gráfico do afastamento da verossimilhança

Figura 3.3 – Gráficos da estimativa influente para o modelo de regressão com a função de ligação logito ajustado aos dados de vitamina A. (a) Gráfico de $IE(\theta)_i$, (b) Gráfico de $IE(\beta)_i$, e (c) Gráfico de $IE(\gamma)_i$

3.8.4 Análise de influência local e total

Nesta seção, fez-se uma análise de influência local para o conjunto de dados de vitamina A utilizando o modelo de regressão para dados grupados com função de ligação logito.

Perturbação de casos

Para aplicar a teoria de influência local desenvolvida na subseção 3.6.2, na qual o esquema de perturbação de casos é utilizado, o valor $C_{d_{max}} = 1,14$ foi obtida como a curvatura máxima. Na Figura 3.4a é apresentada o gráfico de autovetores correspondentes a direção máxima d_{max} , onde a observação #461 destaca-se como um possível ponto influente. A Figura 3.4b mostra os valores das medidas de influência local total C_i e indica que a observação #438 é um possível ponto influente.

Figura 3.4 – Gráficos de medidas de influência para θ do modelo de regressão com a função de ligação logito sob o esquema de perturbação de casos nos dados de vitamina A. (a) Gráfico de influência local d_{max} e (b) gráfico de influência local total C_i

Perturbação de uma covariável

Continuando com a análise de influência local, nesta seção é apresentado os resultados das perturbações feitas na covariável idade do conjunto de dados de vitamina A. Ao perturbar a covariável o valor $C_{d_{max}} = 1,41$ foi encontrado para a curvatura máxima. A Figura 3.5a mostra o gráfico da direção máxima d_{max} versus os índices das observações e na Figura 3.5b é apresentada a medida de influência local total C_i . A Figura 3.5 mostrou que a observações #461 destaca-se entre as demais observações.

Figura 3.5 – Gráficos de medidas de influência para θ do modelo de regressão com a função de ligação logito sob o esquema de perturbação na covariável idade nos dados de vitamina A. (a) Gráfico de influência local $d_{max,i}$ e (b) gráfico de influência local total C_i

3.8.5 Análise de resíduo

Com o objetivo de verificar possíveis observações discrepantes no ajuste do modelo de regressão para dados de sobrevivência grupados de vitamina A, adotou-se o resíduo de Pearson. A Figura 3.6 mostra o comportamento do resíduo. Por meio da análise gráfica desse resíduo, observou-se um comportamento razoavelmente aleatório, apesar da concentração dos indivíduos em duas regiões. Tal fato é explicado devido a utilização dos métodos de análise binária, que apresentam esse tipo de comportamento, ver por exemplo, McGullagh e Nelder (1989), Collett (1991), Souza (2006), entre outros.

Na Figura 3.6, a observação #783 destaca-se das demais observações. A observações #783 representa indivíduo grupados no oitavo intervalo com a segundo maior idade no conjunto de observações que foram falha no tratamento com suplementação de vitamina A e do sexo feminino.

Figura 3.6 – Gráfico do resíduo de Pearson para os dados de vitamina A

3.8.6 Impacto da deleção de observações influentes

Por meio da análise de sensibilidade (influência global e influência local) do modelo de regressão para dados de sobrevivência agrupados utilizando a ligação logito e análise de resíduos foram detectados como possíveis observações influentes os seguintes casos: #128, #438 e #461. Para analisar o impacto dessas observações nas estimativas dos parâmetros, reajustou-se o modelo eliminando individualmente cada observação, em seguida, eliminando as duas observações e finalmente, eliminar todos esses pontos.

Na Tabela 3.6 são apresentados as mudanças relativas (em porcentagens) de cada parâmetro estimado, definido por $\mathbf{RC}_{\theta_j} = \left[(\hat{\theta}_j - \hat{\theta}_{j(i)}) / \hat{\theta}_j \right] \times 100$, em que $\theta_{j(i)}$ é o estimador de máxima verossimilhança sem a i -ésima observação.

Tabela 3.6 – Mudança relativa (em porcentagem), estimativas de máxima verossimilhança e os p -valores correspondentes em parênteses, excluindo as observações #128, #438 e #461

Exclusão	β_1	β_2	β_3
Nenhuma	-	-	-
#461	[-1] -0,038 (0,000)	[-5] -0,169 (0,026)	[-26] -0,037 (0,625)
#128	[0] -0,037 (0,000)	[1] -0,159 (0,036)	[-7] -0,031 (0,680)
#438	[0] -0,038 (0,000)	[2] -0,158 (0,037)	[-11] -0,033 (0,667)
#128 e #438	[0] -0,038 (0,000)	[3] -0,156 (0,040)	[-18] -0,035 (0,650)
#128 e #461	[-1] -0,038 (0,000)	[-4] -0,167 (0,028)	[-33] -0,039 (0,607)
#438 e #461	[-2] -0,038 (0,000)	[-3] -0,166 (0,029)	[-37] -0,040 (0,595)
#128 e #438 e #461	[-2] -0,038 (0,000)	[-2] -0,164 (0,031)	[-44] -0,042 (0,578)

Os resultados da Tabela 3.6 mostraram que a significância das estimativas dos parâmetros não mudou ao nível de significância de 1% depois de remover a i -ésima observação do conjunto de dados original. Logo, não se tem mudanças inferenciais depois de remover as observações influentes detectadas nos gráficos de diagnóstico.

Assim, o modelo final proposto retirando a covariável sexo que não foi significativa é dada por

$$p_j(\mathbf{x}_i) = \frac{\exp(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\gamma_j + \beta_1 x_{i1} + \beta_2 x_{i2})}, \quad i = 1, \dots, 1207 \quad \text{e} \quad j = 1, \dots, 8,$$

em que x_{i1} denota a covariável idade e x_{i2} denota a covariável tratamento. As estimativas de máxima verossimilhança dos parâmetros são apresentados na Tabela 3.7.

Tabela 3.7 – Estimativas de máxima verossimilhança para os parâmetros do modelo de regressão para dados grupados assumindo a função de ligação logito ajustado ao conjunto de dados de vitamina A

Parâmetros	Estimativa	Erro padrão	<i>p</i> -valor
γ_1	-0,094	0,112	0,400
γ_2	0,088	0,121	0,465
γ_3	-0,190	0,137	0,167
γ_4	-0,227	0,151	0,134
γ_5	-0,874	0,187	<0,001
γ_6	-0,643	0,186	0,001
γ_7	-0,481	0,194	0,013
γ_8	-1,227	0,348	0,000
β_1	-0,037	0,003	0,000
β_2	-0,162	0,076	0,032

Por meio da Tabela 3.7 observou-se que a covariável idade e tratamento são significativos.

3.9 Conclusão

Neste trabalho, o modelo de regressão usual em dados grupados foi modificado para incluir a contribuição dos indivíduos que foram censurados. O modelo proposto foi baseado em uma estrutura de regressão modelada por quatro funções de ligação. Observou-se que ligação logito mostrou-se uma boa função de ligação para incluir covariáveis no modelo, porém, a ligação log-log merece uma atenção, uma vez que as estimativas dos erros padrões foram os menores dentre as quatro ligações propostas. O modelo proposto apresentou indicações que seja robusto a perturbações feitas no modelo e nos dados. As técnicas de influência local, influência global e análise de resíduos no modelo de regressão para dados binários assumindo a ligação logito mostraram-se bastante úteis na detecção de possíveis pontos influentes. Na parte de análise de resíduos, como um alternativa para ilustrar as técnicas de resíduos foi proposto o resíduo de Pearson, uma vez que a análise de dados grupados é realizada por meio de métodos binários (CHALITA, 1997). Entretanto, a análise de resíduos para dados de sobrevivência precisa ser melhorada.

3.9.1 Trabalhos futuros

Como possíveis trabalhos futuros pode-se considerar os seguintes temas de pesquisa:

1. Desenvolver um estudo de simulação para estudar a distribuição empírica dos resíduos propostos.

2. Propor um modelo paramétrico de regressão locação-escala no caso de dados de sobrevivência grupados.
3. Considerar o modelo de regressão para dados discretos com fração de cura, proposto por Yu et al. (2004) na ausência de covariáveis.
4. Desenvolver a metodologia de dados de sobrevivência grupados sob o enfoque bayesiano.

Referências

- ALLISON, P.D. **Survival analysis using the SAS system: a practical guide**. Cary: SAS Institute, 1995. 292p.
- ANDRADE, A.C.G. **Efeitos da especificação incorreta da função de ligação no modelo de regressão beta**. 2007, 97p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2007.
- ARANDA-ORDAZ, F.J. An extension of proportional-hazard model for grouped data. **Biometrics**, Washington, v.39, p.109-117, 1983.
- BARRETO, M.L.; SANTOS, L.M.P; ASSIS, A.M.O.; ARAÚJO, M.P.N.; FARENZENA, G.G.; SANTOS, P.A.B.; FIACCONE, R.L. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. **Lancet**, New York, v.344, p.228-231, 1994.
- BERKSON, J. Application of the logistic function to bio-assay.. **Journal of the American Statistical Association**, Boston, v.39, p.357-365, 1944.
- BLISS, C.I. The calculation of the dosage-mortality curve. **Annals of Applied Biology**, Cambridge, v.22, p.134-167, 1935.
- BOLFARINE, H.; RODRIGUES, J.; ACHCAR, J.A. **Análise de sobrevivência**. Rio de Janeiro: Associação Brasileira de Estatística, 1991. 22p.
- CARRASCO, J.M.F. **Modelo de regressão log-Weibull modificado e a nova distribuição Weibull modificada generalizada**. 2007. 128p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.
- CARROLL, R.J.; RUPPERT, D. **Transformation and weighting in regression**. London: Chapman & Hall, 1993. 174p.

CHALITA, L.V.A.S. **Modelos para dados agrupados e censurados**. 1997.135p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1997.

CHRISTENSEN, R.; PEARSON, L.; JOHNSON, W. Case-deletion diagnostics for mixed models. **Technometrics**, Alexandria, v.34, p.38-45, 1992.

COLLETT, D. **Modelling survival data in medical research**. 2nd ed. London: Chapman & Hall, 2003. 391p.

_____. **Modelling binary data**. London: Chapman & Hall, 1991. 369p.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 374p.

COLOSIMO, E.A.; CHALITA, L.V.A.S.; DEMÉTRIO, G.B. Tests of proportional hazards and proportional odds models for grouped survival data. **Biometrics**, Washington, v.56, p.1233-1240, 2000.

COOK, R. D. Detection of influential observations in linear regression. **Technometrics**, Alexandria, v.19, p.15-18, 1977.

_____. Assessment of local influence. **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.

COOK, R. D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman & Hill, 1982. 230p.

COOK, R. D.; PEÑA, D.; WEISBERG, S. The likelihood displacement: a unifying principle for influence. **Communications in Statistics: Theory and Methods**, New York, v.17, p.623-640, 1988.

COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society B**, London, v.34, p.187-220, 1972.

DEMÉTRIO, C.G.B. **Modelos lineares generalizados em experimentação agronômica**, Piracicaba: ESALQ/USP, 2002, 113p.

DOORNIK, J.A. **Ox**: an object-oriented matrix programming language. Disponível em: <<http://www.nuff.ox.ac.uk/User/Doonik/>>. Acesso em: 15 ago. 2007.

ESCOBAR, L.A.; MEEKER, W.Q. Assessing influence in regression analysis with censored data. **Biometrics**, Washington, v.48, p.507-528, 1992.

EFRON, B.; TIBISHIRANI, R.J. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993. 436p.

GIOLO, S.R.; COLOSIMO, E.A.; DEMÉTRIO, C.G.B. Different approaches for modelling grouped survival data: a mango tree study. **Journal of Agricultural Biological and Environmental Statistics**, Alexandria, v.42, p.165-186, 2008.

GOMES, E.M.C. **Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas**. 2007. 104p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

HERTZ-PICCIOTO, I.; ROCKHILL, B. Validity and efficiency of approximation methods for tied survival times in cox regression. **Biometrics**, Washington, v.53, p.1151-1156, 1997.

HEITJAN, D.F. Inference from grouped continuous data: a review. **Statistical Science**, Hayward, v.4, p.164-183, 1989.

HINES, R.J.O.; LAWLESS, J.F.; CARTER, E.M. Diagnostics for a cumulative multinomial generalized linear model, with applications to grouped toxicological mortality data. **Journal of the American Statistical Association**, Boston, v.87, p.1059-1069, 1992.

KLEIN, J.P.; MOESCHBERGER, M.L. **Survival analysis techniques for censored and truncated data**. New York: Springer-Verlang, 1997. 357p.

LAM, K.F.; IP, D. REML and ML estimation for clustered grouped survival data. **Statistics in medicine**, Chichester, v.22, p.2025-2034, 2003.

LAWLESS, J.F. **Statistical models and methods for lifetime Data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 637p.

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

LI, J. **Modeling of silent event and evaluation of biomarkers in a prospective study with scheduled follow-up**. 2006. 110p. Thesis (Phylosophy Doctor) - Mathematics and Statistics Institute, Boston University, Boston, 2006.

LIPSITZ, S.R.; LAIRD, N.M.; HARRINGTON, D.P. Using the Jackknife to estimate the variance of regression estimators from repeated measures studies. **Communications in Statistics: Theory Methods**, New York, v.19, p.821-845, 1990.

MANLY, B.F.J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2nd ed. London: Chapman & Hall, 1997. 399p.

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

MILLER, R.G. The jackknife-a review. **Biometrika**, London, v.61, p.1-15, 1974.

MOREIRA, J.A. **Utilização do método bootstrap na escolha entre os modelos de cox e logístico para dados de sobrevivência com censura intervalar**. 2001. 65p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2001.

MORENO-REBOLLO, J.L.; MUÑOZ-REYES, A.; MUÑOZ-PICHARDO J. Influence diagnostic in survey sampling: conditinal bias. **Biometrika**, London, v.86, p.923-928, 1999.

MUTHÉN, B.; MASYN, K. Discrete-time survival mixture analysis. **Journal of Education and Behavioral Statistics**, Washington, v.30, p.27-58, 2005.

NOBRE, J.S. **Métodos de diagnóstico para modelos lineares mistos**. 2004. 99p. Dissertação(Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2004.

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

ORTEGA E.M.M.; BOLFARINE H; PAULA G.A. Influence diagnostic in generalized log-gamma regression model. **Computational Statistics and Data Analysis**, Amsterdam, v.42, p.165-186, 2003.

ORTEGA, E.M.M.; CANCHO, V.G.; PAULA, G.A. Generalized log-gamma regression models with cure. **Lifetime Data Analysis**, 2008, In Press.

PARKER, C.B.; DELONG, E.R. A diagnostic for cox regression with discrete failure-time models. **Biometrics**, Washington, v.56, p.996-1001, 2000.

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

PETTITT, A.N.; BIN DAUD, I. Case-weighted measures of influence for proportional hazards regression. **Applied Statistics**, London, v.38, p.51-67, 1989.

POON, W.Y.; POON, Y.S. Conformal normal curvature and assessment of local influence. **Journal of the Royal Statistical Society Series B**, London, v.61, p.51-61, 1999.

PRENTICE, R.L.; GLOECKLER, L.A. Regression analysis of grouped survival data with application to breast cancer data. **Biometrics**, Washington, v.34, p.57-67, 1978.

QUENOUILLE, M.H. Approximate tests of correlation in time-series. **Journal of the Royal Statistical Society**, London, v.11, p.68-84, 1949.

_____. Notes on bias in estimation. **Biometrika**, Cambridge, v.43, p.353-360, 1956.

RESCHE-RIGON, M.; CHEVRET, S. Local influence for the subdistribution of a competing risk. **Statistics in medicine**, Chichester, v.25, p.1937-1947, 2006.

SILVA, G.O.; ORTEGA, E.M M.; CANCHO, V.G.; BARRETO, M.L. Log-Burr XII regression models with censored Data. **Computational Statistics and Data Analysis**, Amsterdam, v.52, p.3820-3842, 2008.

SINGER, J.D.; WILLETT, J.B. It's about time: using discrete-time survival analysis to study duration and the timing of events. **Journal of Education Statistics Summer**, Washington, v.18, p.155-195, 1993.

SOUZA, E.C. **Análise de influência local no modelo de regressão logística**. 2006. 101p. Dissertação (Mestrado em Biometria) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2006.

THOMPSON JR., W.A. On the treatment of grouped observation in life studies. **Biometrics**, Washington, v.33, p.463-470, 1977.

YU, B; TIWARI, R.C.; CRONIN, K.A.; FEUER, E.J. Cure fraction estimation from the mixture cure models for grouped survival data. **Statistics in Medicine**, Chichester, v.23, p.1733-1747, 2004.

WANG, P.; PUTERMAN, M.L.; COCKBURN, I.; LE, N. Mixed poisson regression models with covariate dependent rates. **Biometrics**, Washington, v.52, p.381-400, 1996.

APÊNDICES

APÊNDICE A - Modelo de regressão Weibull-exponenciada com censura intervalar

A1. Demonstração do Teorema

Deriva-se uma representação de soma infinita para o k -ésimo momento ordinário para a distribuição da $f(z)$ (1.5), denotado por $\mu'_k = E(Z^k)$. Tem-se que

$$\mu'_k = \int_{-\infty}^{\infty} z^k \lambda \left\{ 1 - \exp[-\exp(z)] \right\}^{\lambda-1} \exp[z - \exp(z)] dz. \quad (1)$$

Utilizando a representação da série

$$(1+x)^a = \sum_{j=0}^{\infty} \frac{\Gamma(a+1)x^j}{\Gamma(a-j+1)j!}. \quad (2)$$

Considere $u = \exp(z)$ e utilizando a expressão (2), o k -ésimo momento (1) pode ser expandido como

$$\mu'_k = \lambda \Gamma(\lambda) \sum_{j=0}^{\infty} \frac{(-1)^j}{\Gamma(\lambda-j)j!} \int_0^{\infty} [\log(u)]^k \exp[-u(j+1)] du, \quad (3)$$

a integral em (3) é calculada em Prudnikov, Brychkov e Marichev (1986) e Nadarajah (2006), assim, tem-se

$$\int_0^{\infty} [\log(u)]^k \exp[-u(j+1)] du = \frac{\partial^k [(j+1)^{-a} \Gamma(a)]}{\partial a^k} \Big|_{a=1}. \quad (4)$$

Finalmente, substituindo (4) em (3), o k -ésimo momento de Z é dado por

$$\mu'_k = \lambda \Gamma(\lambda) \sum_{j=0}^{\infty} \frac{(-1)^j}{\Gamma(\lambda-j)j!} \frac{\partial^k [(j+1)^{-a} \Gamma(a)]}{\partial a^k} \Big|_{a=1}. \quad (5)$$

A2. Estimação de parâmetros do modelo de regressão Weibull-exponenciada

/*****

PROGRAMA: Estimação de parametros

OBJETIVO: Estimar os parametros do modelo de regressão weibull-exponenciada

DADOS: Cancer de mama

MODELO: Modelo apresentado por Colosimo e Giolo (2006)

*****/

#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 n;
log_vero(const vP, const adFunc, const avScore, const amHessiam)
{
    decl cont, t1, t2, xx0, xx1;
    decl uns=ones(n,1);
    decl vero=zeros(1,n);
    decl theta=fabs(vP[0][0]);
    decl delta=fabs(vP[1][0]);
    decl beta0=vP[2][0];
    decl beta1=vP[3][0];
    for(cont=0;cont<n;++cont)
    {
        t1=gm_Y[cont][0];
        t2=gm_Y[cont][1];
        xx0=gm_Y[cont][3];
        xx1=gm_Y[cont][4];
        decl xbeta=beta0*xx0+beta1*xx1;
        decl z1=(log(t1)-xbeta)/delta;
        decl z2=(log(t2)-xbeta)/delta;
        if(gm_Y[cont][2]==1)
            vero[0][cont]=log(((1-((1-exp(-exp(z1)))^theta))-(1-((1-exp(-exp(z2)))^theta))));
        if(gm_Y[cont][2]==0)
            vero[0][cont]=log(1-((1-exp(-exp(z1)))^theta));
    }
    adFunc[0]=double(vero*uns);
    return 1;
}
main()
{
    decl vP, dfunc;

```

```

gm_Y=loadmat("cancer_mama.txt");
n=rows(gm_Y);
print("Amostra", gm_Y);
vP=< 0.6; 0.6193; 3.8993; -0.5676 >; SAS
log_vero(vP, &dfunc, 0,0);
println("vero=", dfunc);
MaxControl(-1,10);
decl mhess=0.00001*unit(4);
decl ir;
ir=MaxBFGS(log_vero, &vP, &dfunc, &mhess, 1);
Num2Derivative(log_vero, vP, &mhess);
print("Imprimir mhess", mhess);
decl ep=sqrt(diagonal(-1/mhess));
print(" \n Estimativa Erro Padrao p-valor ", vP~ep'~(2*(1-probn(fabs(vP'./ep))))');

/*=====Intervalo de confiança =====*/
print(" \n ===== Intervalo de Confianca: WE-Censura intervalar ===== \n ");
/* Nivel de significancia */
decl alfa=0.05;
decl z=quann(1-(alfa/2));
print(" \n Quantil da normal= ", z);
decl m=rows(vP);
decl ic=zeros(m,2);
decl i;
for(i=0;i<m;++i)
{
    ic[i][0]=vP[i]-z*ep[i];
    ic[i][1]=vP[i]+z*ep[i];
}
print(" \n Intervalo de confianca EMV ",ic);
/*===== Fim: Intervalo de confiança =====*/

/*===== Resíduos =====*/
decl j,ui,vi,xbeta,su,sv;
decl lambda=vP[0];
decl sigma=vP[1];

```

```

decl b0=vP[2];
decl b1=vP[3];
decl rci=zeros(n,1);
decl rmi=zeros(n,1);
decl rdi=zeros(n,1);
for(j=0;j<n;++j)
{
    ui=log(gm_Y[j][0]);
    vi=log(gm_Y[j][1]);
    xbeta=b0*gm_Y[j][3]+b1*gm_Y[j][4];
    su=1-(1-exp(-exp((ui-xbeta)/sigma)))^lambda;
    sv=1-(1-exp(-exp((vi-xbeta)/sigma)))^lambda;
    if(gm_Y[j][2]==1)
    {
        rci[j]=(su*(1-log(su))-sv*(1-log(sv)))/(su-sv);
        rmi[j]=(su*log(su)-sv*log(sv))/(su-sv);
    }
    if(gm_Y[j][2]==0)
    {
        rci[j]=1-log(su);
        rmi[j]=log(su);
    }
    rdi[j]=(rmi[j]/fabs(rmi[j]))*((-2*(rmi[j]+gm_Y[j][2]*log(gm_Y[j][2]-rmi[j])))^(1/2));
}
print("\n Resíduos", rci~rmi~rdi);
/*===== Fim: Resíduos =====*/
}

```

A3. Matriz de informação observada $\ddot{L}(\theta)$

Elementos da matriz de informação observada para o modelo de regressão Weibull-exponenciada com censura intervalar. Após algumas manipulações algébricas, tem-se

$$\begin{aligned}
 \mathbf{L}_{\lambda\lambda} = & \sum_{i \in F} \left\{ \frac{(b_i)_\lambda \log(1 - b_i) - (a_i)_\lambda \log(1 - a_i)}{(1 - b_i)^\lambda - (1 - a_i)^\lambda} - \left[\frac{(b_i)_\lambda - (a_i)_\lambda}{(1 - b_i)^\lambda - (1 - a_i)^\lambda} \right]^2 \right\} - \\
 & - \sum_{i \in C} \left\{ \frac{(a_i)_\lambda \log(1 - a_i)}{1 - (1 - a_i)^\lambda} + \left[\frac{(a_i)_\lambda}{1 - (1 - a_i)^\lambda} \right]^2 \right\},
 \end{aligned}$$

$$\begin{aligned}
\mathbf{L}_{\lambda\sigma} &= \sum_{i \in F} \left\{ \frac{\sigma^{-1}(zu_i g_i p_i - zv_i h_i q_i)}{(1-b_i)^\lambda - (1-a_i)^\lambda} - \left[\frac{[\lambda\sigma^{-1}((a_i)_\lambda - (b_i)_\lambda)(zu_i g_i - zv_i h_i)]}{[(1-b_i)^\lambda - (1-a_i)^\lambda]^2} \right] \right\} + \\
&\quad + \sum_{i \in C} \left\{ \frac{\sigma^{-1}zu_i g_i}{1 - (1-a_i)^\lambda} \left[p_i + \frac{\lambda(1-a_i)^{\lambda-1} \log(1-a_i)}{1 - (1-a_i)^\lambda} \right] \right\}, \\
\mathbf{L}_{\lambda\beta_j} &= \sum_{i \in F} \left\{ \frac{\sigma^{-1}x_{ij}[g_i p_i - h_i q_i]}{(1-b_i)^\lambda - (1-a_i)^\lambda} - \left[\frac{\lambda\sigma^{-1}x_{ij}((a_i)_\lambda - (b_i)_\lambda)(g_i - h_i)}{[(1-b_i)^\lambda - (1-a_i)^\lambda]^2} \right] \right\} + \\
&\quad + \sum_{i \in C} \left\{ \frac{\sigma^{-1}x_{ij}g_i}{1 - (1-a_i)^\lambda} \left[p_i + \frac{\lambda(1-a_i)^{\lambda-1} \log(1-a_i)}{1 - (1-a_i)^\lambda} \right] \right\}, \\
\mathbf{L}_{\sigma\sigma} &= \sum_{i \in F} \left\{ \frac{\lambda\sigma^{-2}[zv_i h_i (b_i)_\sigma - zu_i g_i (a_i)_\sigma]}{(1-b_i)^\lambda - (1-a_i)^\lambda} - \left[\frac{\lambda\sigma^{-1}(zu_i g_i - zv_i h_i)}{[(1-b_i)^\lambda - (1-a_i)^\lambda]} \right]^2 \right\} - \\
&\quad - \sum_{i \in C} \left\{ \frac{\lambda\sigma^{-2}zu_i g_i}{1 - (1-a_i)^\lambda} \left[(a_i)_\sigma + \frac{\lambda zu_i g_i}{1 - (1-a_i)^\lambda} \right] \right\}, \\
\mathbf{L}_{\sigma\beta_j} &= \sum_{i \in F} \left\{ \frac{\lambda\sigma^{-2}x_{ij}[h_i (b_i)_{\sigma\beta} - g_i (a_i)_{\sigma\beta}]}{(1-b_i)^\lambda - (1-a_i)^\lambda} - \left[\frac{\lambda^2\sigma^{-2}x_{ij}(zu_i g_i - zv_i h_i)(g_i - h_i)}{[(1-b_i)^\lambda - (1-a_i)^\lambda]^2} \right] \right\} - \\
&\quad - \sum_{i \in C} \left\{ \frac{\lambda\sigma^{-2}x_{ij}g_i}{1 - (1-a_i)^\lambda} \left[(a_i)_{\sigma\beta} + \frac{\lambda zu_i g_i}{1 - (1-a_i)^\lambda} \right] \right\}, \\
\mathbf{L}_{\beta_j\beta_s} &= \sum_{i \in F} \left\{ \frac{\lambda\sigma^{-2}x_{ij}x_{is}[h_i (b_i)_\beta - g_i (a_i)_\beta]}{(1-b_i)^\lambda - (1-a_i)^\lambda} - \left[x_{ij}x_{is} \left(\frac{\lambda\sigma^{-1}(g_i - h_i)}{(1-b_i)^\lambda - (1-a_i)^\lambda} \right)^2 \right] \right\} - \\
&\quad - \sum_{i \in C} \left\{ \frac{\lambda\sigma^{-2}x_{ij}x_{is}g_i}{1 - (1-a_i)^\lambda} \left[(a_i)_\beta + \frac{\lambda g_i}{1 - (1-a_i)^\lambda} \right] \right\}
\end{aligned}$$

em que $zu_i = (\log(u_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $zv_i = (\log(v_i) - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $a_i = \exp[-\exp(zu_i)]$,
 $b_i = \exp[-\exp(zv_i)]$, $(a_i)_\lambda = (1-a_i)^\lambda \log(1-a_i)$, $(b_i)_\lambda = (1-b_i)^\lambda \log(1-b_i)$,
 $g_i = (1-a_i)^{\lambda-1} a_i \exp(zu_i)$, $h_i = (1-b_i)^{\lambda-1} b_i \exp(zv_i)$, $p_i = \lambda \log(1-a_i) + 1$,
 $q_i = \lambda \log(1-b_i) + 1$, $(a_i)_\sigma = 2 + (1-a_i)^{-1} zu_i a_i \exp(zu_i)[\lambda - 1] + zu_i[1 - \exp(zu_i)]$,
 $(b_i)_\sigma = 2 + (1-b_i)^{-1} zv_i b_i \exp(zv_i)[\lambda - 1] + zv_i[1 - \exp(zv_i)]$,
 $(a_i)_{\sigma\beta} = 1 + (1-a_i)^{-1} zu_i a_i \exp(zu_i)[\lambda - 1] + zu_i[1 - \exp(zu_i)]$,
 $(b_i)_{\sigma\beta} = 1 + (1-b_i)^{-1} zv_i b_i \exp(zv_i)[\lambda - 1] + zv_i[1 - \exp(zv_i)]$,
 $(a_i)_\beta = 1 + (1-a_i)^{-1} a_i \exp(zu_i)[\lambda - 1] - \exp(zu_i)$ e
 $(b_i)_\beta = 1 + (1-b_i)^{-1} b_i \exp(zv_i)[\lambda - 1] - \exp(zv_i)$.

A4. Simulação de resíduos

/*****

PROGRAMA: Simulação de Resíduos

OBJETIVO: Estudar residuos de Cox-Snell, Martingal e Deviance modificados para acomodar as

censura intervalares.

DISTRIBUIÇÃO: Weibull-exponenciada

```

*****/
#include<oxstd.h>
#include<oxprob.h>
#import<maximize>
static decl n,l;
static decl datm1;
log_vero(const vP, const adFunc, const avScore, const amHessian)
{
    decl cont,t1,t2,xx0,xx1;
    decl uns=ones(n,1);
    decl vero=zeros(1,n);
    decl theta=fabs(vP[0]);
    decl delta=fabs(vP[1]);
    decl beta0=vP[2];
    decl beta1=vP[3];
    for(cont=0;cont<n;++cont)
    {
        t1=datm1[cont][0];
        t2=datm1[cont][1];
        xx0=datm1[cont][3];
        xx1=datm1[cont][4];
        decl xbeta=beta0*xx0+beta1*xx1;
        decl z1=(log(t1)-xbeta)/delta;
        decl z2=(log(t2)-xbeta)/delta;
        if(datm1[cont][2]==1)
            vero[0][cont]=log((1-((1-exp(-exp(z1)))^theta))-(1-((1-exp(-exp(z2)))^theta)));
        if(datm1[cont][2]==0)
            vero[0][cont]=log(1-((1-exp(-exp(z1)))^theta));
    }
    adFunc[0]=double(vero*uns);
    return 1;
}
main()
{

```

```

/* Tamanho da amostra */
n=200;
/* Parametros do modelo */
decl theta=0.5;
decl sigma=0.5;
decl beta0=4;
decl beta1=2;
decl r,param;
/* Numero de replicas */
r=1000;
/* Matriz para guardar os residuos */
/* Residuo Cox-Snell */
decl rci=zeros(n,r);
/* Residuo Martingal */
decl rmi=rci; /* Residuo Deviance */
decl rdi=rci;
/*===== gerar r replicas de amostra de censura intervalar =====*/
param=zeros(r,4,0);
l=0;
while(l<r)
{
    print("===== ", l, " Replica =====\n");
    /* Intecepto */
    decl x0=ones(n,1);
    /* Covariavel */
    decl x1=ranu(n,1);
    decl obs,tem,m,len,a,b,c,delta,cont,np,p;
    tem=zeros(n,1,0);
    len=tem;
    c=tem;
    obs=tem;
    delta=tem;
    /* Limites da uniforme para gerar o comprimento */
    a=1;
    b=5.5;
    /* Porcentagem de censura a direita */

```

```

p=0.5;
/* Numero de observações censuradas a direita na amostra */
np=n*p;
/* Gerar Y=ln(tempo) e o comprimento do intervalo */
for(m=0;m<n;++m)
{
    decl u=ranu(1,1);
    /* Y=ln(tempo) */
    tem[m]=x0[m]*beta0+x1[m]*beta1+(sigma*log(-log(1-(u^(1/theta)))));
    /* Comprimento dos intervalos */
    len[m]=int((b-a)*ranu(1,1)+a);
}
//print("\n Tempo gerado ", tem len);
/* Determinando o valo minimo e maximo de Y */
decl maxi=max(exp(tem));
decl mini=min(exp(tem));
//print("\n maximo e minimo", maxi mini);
/* Gerar uma amostra com np observações censuradas */
decl v=0;
while(v<7000)
{
    cont=0;
    for(m=0;m<n;++m)
    {
        // c[m]=10*ranu(1,1)+(maxi+3); //para 0 de censura a direita
        c[m]=200*ranu(1,1); //para 10
        //c[m]=30*ranu(1,1);
        /* Gerando censura intervalar e a direita */
        if(exp(tem[m])<c[m])
        {
            delta[m]=1;
            obs[m]=exp(tem[m]);
        }
        if(c[m]<exp(tem[m]))
        {
            delta[m]=0;

```

```

        obs[m]=c[m];
    }          if(delta[m]==0) cont=cont+1;
}
//      print(v,"Contador= ", cont,"\n");
      if(np==cont) break;
      if(np<cont || np>cont) ++v;
}
/* Contador para o numero de observações censuradas a direita na amostra */
decl s;
decl kon=0;
for(s=0;s<n;++s)
{
    if(delta[s]==0) kon=kon+1;
}
print(1," Contador= ", kon, "\n");
/* Construindo os intervalos */
decl i, j;
decl li=zeros(n,1,0);
decl ls=li;
for(i=0;i<n;++i)
{
    decl sel=int((len[i])*ranu(1,1)); //selecionar um dos candidatos a intervalo
    //println("ind-",i, " Num selecionado ", sel,"\n");
    /* gerar os candidatos a intervalo */
    decl yi=zeros(len[i],2,0);
    for(j=0;j<len[i];++j)
    {
        /* Para censura intervalar */
        if(delta[i]==1)
        {
            if(round(obs[i])-j>0)
            {
                yi[j][0]=round(obs[i])-j;
                yi[j][1]=round(obs[i])+(len[i]-j);
            }
            if(round(obs[i])-j<=0)

```



```

        {
            yi[j][0]=0.0001;
            yi[j][1]=round(obs[i])+j+1;
        }
    }
    /* Para censura a direita */
    if(delta[i]==0)
    {
        yi[j][0]=obs[i];
        yi[j][1]=.NaN;
    }
    /* Escolhendo os intervalos selecionados para compor o conjunto de dados */
    if(len[i]>1)
    {
        li[i]=yi[sel][0];
        ls[i]=yi[sel][1];
    }
    if(len[i]<=1)
    {
        li[i]=yi[j][0];
        ls[i]=yi[j][1];
    }
}
//print(i," Candidatos a intervalo", yi,"\n");
}
/* Conjunto de dados gerados*/
datm1=li ls delta x0 x1;
// print("\n Amostra M1 ", datm1);
MaxControl(-1, 5);
decl vp,dfunc,ir,mhess;
vp = < 0.5; 0.5; 4; 2 >; // valor nicial
ir = MaxBFGS(log_vero, &vp, &dfunc, 0, 1);
/* Se convergir: */
if (ir == MAX_CONV)
{
    if (Num2Derivative(log_vero, vp, &mhess))

```

```

{
    param[l][0]=fabs(vp[0]);
    param[l][1]=fabs(vp[1]);
    param[l][2]=vp[2];
    param[l][3]=vp[3];
    mhess = -invert(mhess) / n;
    // print("standard errors:", sqrt(diagonal(mhess)'));
}
decl j;
for(j=0;j<n;++j)
{
    decl li=log(datm1[j][0]);
    decl ui=log(datm1[j][1]);
    decl xbeta1=param[l][2]*datm1[j][3]+param[l][3]*datm1[j][4];
    decl sl=1-((1-exp(-exp((li-xbeta1)/param[l][1])))^param[l][0]);
    decl su=1-((1-exp(-exp((ui-xbeta1)/param[l][1])))^param[l][0]);
    if(datm1[j][2]==1)
    {
        rci[j][l]=(sl*(1-log(sl))-su*(1-log(su)))/(sl-su);
        rmi[j][l]=(sl*log(sl)-su*log(su))/(sl-su);
        rdi[j][l]=(rmi[j][l]/fabs(rmi[j][l]))*((-2*(rmi[j][l]+log(1-rmi[j][l])))^(1/2));
    }
    if(datm1[j][2]==0)
    {
        rci[j][l]=1-log(sl);
        rmi[j][l]=log(sl);
        rdi[j][l]=(rmi[j][l]/fabs(rmi[j][l]))*((-2*rmi[j][l])^(1/2));
    }
}
++l;
}
}

/*===== Fim: replicas =====*/

/*===== Imprimindo as estimativas dos parametros =====*/
/* Estimativas e erro padrão: M1 */

```

```

decl final_m1=sumc(param)/r;
print("\n Estimativa M1", final_m1);
decl varian_m1=varc(param)/r;
print("\n Erro padrao M1", sqrt(varian_m1));
/*===== Fim: parametros =====*/

/*===== Imprimindo os resíduos =====*/
// print("\n Matriz de Residuo Cox-Snell \n", rc_m1);
// print("\n Matriz de Residuo Martingale \n", rdi);
decl media_rci=meanr(sortc(rci));
decl media_rmi=meanr(sortc(rmi));
decl media_rdi=meanr(sortc(rdi));
// print("Media dos residuos Classica", media_rci~media_rmi~media_rdi);
/*===== Fim: Resíduos =====*/

/*===== Gravando os resultados da simulação =====*/
decl resm1_file=fopen("c:/w1_reg/res_50200.txt","w");
fprintf(resm1_file, media_rci~media_rmi~media_rdi);
/*===== Fim: Gravando =====*/
}

```

APÊNDICE B - Modelo de regressão para dados de sobrevivência grupados

B1. Contribuição das observações censuradas

Neste apêndice, estão os cálculos desenvolvidos para obter a contribuição dos indivíduos censurados em termos do intervalo.

Considere a contribuição de um indivíduo censurado, dada pela seguinte expressão

$$\frac{S(c_i)}{S(a_{j-1})}, \quad i = 1, \dots, n \quad \text{e} \quad j = 1, \dots, k, \quad (1)$$

em que $S(\cdot)$ é função de sobrevivência e c_i é o tempo de censura do i -ésimo indivíduo.

Uma alternativa, proposta por Thompson (1977), para escrever a expressão (1) em termos dos intervalo $[a_{j-1}, a_j]$ é associar a c_i uma probabilidade uniforme no intervalo e aproximar a taxa de falha como constante sobre o intervalo. Então,

$$\frac{S(a_j)}{S(a_{j-1})} \cong \frac{\exp(-r_j a_j)}{\exp(-r_j a_{j-1})} = \exp[-r_j(a_j - a_{j-1})]$$

e

$$\frac{S(c_i)}{S(a_{j-1})} \cong \frac{\exp(-r_j c_i)}{\exp(-r_j a_{j-1})} = \exp[-r_j(c_i - a_{j-1})], \quad (2)$$

em que r_j é o parâmetro da distribuição exponencial para o j -ésimo intervalo.

Mas observe que

$$\frac{S(a_j)}{S(a_{j-1})} \cong \exp[-r_j(a_j - a_{j-1})] \Rightarrow \left[\frac{S(a_j)}{S(a_{j-1})} \right]^{(a_j - a_{j-1})^{-1}} = \exp(-r_j).$$

Então, substituindo $\left[\frac{S(a_j)}{S(a_{j-1})} \right]^{(a_j - a_{j-1})^{-1}} = \exp(-r_j)$ na expressão (2), obtém-se

$$\frac{S(c_i)}{S(a_{j-1})} \cong \exp[-r_j(c_i - a_{j-1})] = \left[\frac{S(a_j)}{S(a_{j-1})} \right]^{(a_j - a_{j-1})^{-1}(c_i - a_{j-1})}.$$

Sabendo que c_i tem distribuição uniforme no intervalo $[a_{j-1}, a_j]$, tem-se

$$\frac{(c_i - a_{j-1})}{a_j - a_{j-1}} = \begin{cases} 1 & \text{se } c_i = a_{j-1} \\ 0 & \text{se } c_i = a_j \end{cases}.$$

Assim, o valor esperado da quantidade $(c_i - a_{j-1})/(a_j - a_{j-1})$ é dado por

$$E \left[\frac{c_i - a_{j-1}}{a_j - a_{j-1}} \right] = \int_{a_{j-1}}^{a_j} \left[\frac{1}{a_j - a_{j-1}} \left(\frac{c_i - a_{j-1}}{a_j - a_{j-1}} \right) \right] dc_i = \frac{1}{2}.$$

Portanto, tem-se

$$\frac{S(c_i)}{S(a_{j-1})} = \left[\frac{S(a_j)}{S(a_{j-1})} \right]^{1/2}.$$

B2. Estimação dos parâmetros do modelo de regressão

/******

PROGRAMA: Estimar parâmetros dos dados de Vitamina A

OBJETIVO: Estimar parâmetros no caso de dados agrupados

DISTRIBUIÇÃO: Logística

*****/

```
#include<oxstd.h>
```

```
#include<oxprob.h>
```

```
#include<maximize.h>
```

```
#include<simula.h>
```

```
#import<maximize>
```

```
#pragma link("maximize.oxo")
```

```
//Variável global
```

```
static decl g_mY;
```

```
static decl inter;
```

```
static decl d1;
```

```
static decl d2;
```

```
static decl n;
```

```
static decl k;
```

```
//Função de verossimilhança
```

```
lv(const vP, const adFunc, const avScore, const amHessian)
```

```
{
```

```
    decl xbeta, cont1,cont2;
```

```
    decl xx1,xx2,xx3;
```

```
    decl unsk=ones(1,k);
```

```
    decl unsn=ones(n,1);
```

```
    decl vero=zeros(k,n);
```

```
    decl beta1=vP[8];
```

```
    decl beta2=vP[9];
```

```
    for(cont1=0;cont1<k;++cont1) //fazer o somatorio dos intervalos
```

```
    {
```

```
        for(cont2=0;cont2<n;++cont2) //fazer o somatorio dos individuos
```

```
        {
```

```
            xx1=g_mY[cont2][3];
```

```
            xx2=g_mY[cont2][4];
```

```

xx3=g_mY[cont2][5];
xbeta=xx1*beta1+xx2*beta2;

if(d1[cont1][cont2]==0 && d2[cont1][cont2]==1)/*Falha*/
    vero[cont1][cont2]=log(exp(vP[cont1]+xbeta)/(1+exp(xbeta+vP[cont1])));
if(d1[cont1][cont2]==0 && d2[cont1][cont2]==0)/*Censura*/
    vero[cont1][cont2]=(1/2)*log(1/(1+exp(xbeta+vP[cont1])));
if(d1[cont1][cont2]==1 && d2[cont1][cont2]==0)/*Risco*/
    vero[cont1][cont2]=-log(1+exp(xbeta+vP[cont1]));
if(d1[cont1][cont2]==.NaN && d2[cont1][cont2]==.NaN)/*Individuos que passaram para
o proximo intervalo*/
    vero[cont1][cont2]=0;
    }
}
adjunct[0]=double((unsk*vero)*unsn);
return 1;
}
main()
{
    decl dados=loadmat("vitamina.txt");
    n=rows(dados);
    inter=< 4,21;21,38;38,55;55,73; 73,90;90,108;108,126;126,185>;
    k=rows(inter);
    decl i,j,t,cens;
    d1=ones(k,n);
    d2=zeros(k,n);
    for(i=0;i<n;++i)
    {
        for(j=0;j<k;++j)
        {
            t=dados[i][0];
            cens=dados[i][1];
            if(t>=inter[j][0] _ t<inter[j][1])
            {
                d1[j][i]=0;
                if(cens==1)

```

```

        {
            d2[j][i]=1;
        }
        if(cens==0)
        {
            d2[j][i]=0;
        }
    }
    if(t<inter[j][0])
    {
        d1[j][i]=.NaN;
        d2[j][i]=.NaN;
    }
}
}
g_mY=dados;
decl dfunc;
decl vP=<0.75808;0.73443;0.79341;0.80725; 0.89074;0.86828;0.85489;0.96154;
    0.02210;0.08980>;
lv(vP,&dfunc,0,0);
println("vero=" ,dfunc);
MaxControl(-1,20);
decl mhess=0.00001*unit(10);
decl ir;
ir=MaxBFGS(lv,&vP,&dfunc,&mhess,1);
Num2Derivative(lv,vP,&mhess);
print("Imprimir mhess \n", mhess);
decl ep= sqrt(diagonal(-1/mhess));
print(" \n Estimativa Erro Padrao p-valor ", vP~ep'~(2*(1-probn(fabs(vP'./ep)))));
}

```

B3. Matriz de informação observada $\ddot{L}(\theta)$

As expressões necessárias para obter as derivadas parciais de segunda ordem do logaritmo da função de verossimilhança estão neste apêndice.

$$\begin{aligned}
\mathbf{L}_{\beta_m \beta_m} &= \sum_{j=1}^k \left\{ \sum_{i \in F_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_m}}{g(\eta_{ij})} - \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m}}{g(\eta_{ij})} \right)^2 \right] - \sum_{i \in R_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_m}}{1 - g(\eta_{ij})} + \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m}}{1 - g(\eta_{ij})} \right)^2 \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_m}}{1 - g(\eta_{ij})} + \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m}}{1 - g(\eta_{ij})} \right)^2 \right] \right\}, \\
\mathbf{L}_{\beta_m \beta_y} &= \sum_{j=1}^k \left\{ \sum_{i \in F_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y}}{g(\eta_{ij})} - \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m} [\dot{g}(\eta_{ij})]_{\beta_y}}{[g(\eta_{ij})]^2} \right) \right] - \sum_{i \in R_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y}}{1 - g(\eta_{ij})} + \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m} [\dot{g}(\eta_{ij})]_{\beta_y}}{[1 - g(\eta_{ij})]^2} \right) \right] - \frac{1}{2} \sum_{i \in C_j} \left[\frac{[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y}}{1 - g(\eta_{ij})} + \left(\frac{[\dot{g}(\eta_{ij})]_{\beta_m} [\dot{g}(\eta_{ij})]_{\beta_y}}{[1 - g(\eta_{ij})]^2} \right) \right] \right\}, \\
\mathbf{L}_{\beta_m \gamma_h} &= \sum_{i \in F_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\beta_m \gamma_h}}{g(\eta_{ih})} - \left(\frac{[\dot{g}(\eta_{ih})]_{\beta_m} [\dot{g}(\eta_{ih})]_{\gamma_h}}{[g(\eta_{ih})]^2} \right) \right] - \sum_{i \in S_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\beta_m \gamma_h}}{1 - g(\eta_{ih})} + \left(\frac{[\dot{g}(\eta_{ih})]_{\beta_m} [\dot{g}(\eta_{ih})]_{\gamma_h}}{[1 - g(\eta_{ih})]^2} \right) \right] - \frac{1}{2} \sum_{i \in C_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\beta_m \gamma_h}}{1 - g(\eta_{ih})} + \left(\frac{[\dot{g}(\eta_{ih})]_{\beta_m} [\dot{g}(\eta_{ih})]_{\gamma_h}}{[1 - g(\eta_{ih})]^2} \right) \right], \\
\mathbf{L}_{\gamma_h \gamma_h} &= \sum_{i \in F_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\gamma_h \gamma_h}}{g(\eta_{ih})} - \left(\frac{[\dot{g}(\eta_{ih})]_{\gamma_h}}{g(\eta_{ih})} \right)^2 \right] - \sum_{i \in S_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\gamma_h \gamma_h}}{1 - g(\eta_{ih})} + \left(\frac{[\dot{g}(\eta_{ih})]_{\gamma_h}}{1 - g(\eta_{ih})} \right)^2 \right] - \frac{1}{2} \sum_{i \in C_h} \left[\frac{[\ddot{g}(\eta_{ih})]_{\gamma_h \gamma_h}}{1 - g(\eta_{ih})} + \left(\frac{[\dot{g}(\eta_{ih})]_{\gamma_h}}{1 - g(\eta_{ih})} \right)^2 \right],
\end{aligned}$$

em que $\eta_{ij} = \gamma_j + \mathbf{x}_i^T \boldsymbol{\beta}$, e as manipulações matemáticas associadas a cada função de ligação são dadas por

- **Ligação logito**

$$g(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}, \quad [\dot{g}(\eta_{ij})]_{\beta_m} = \frac{x_{im} \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^2}, \quad [\dot{g}(\eta_{ij})]_{\beta_y} = \frac{x_{iy} \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^2},$$

$$[\dot{g}(\eta_{ij})]_{\gamma_h} = \frac{\exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^2}, \quad [\ddot{g}(\eta_{ij})]_{\beta_m \beta_m} = (x_{im})^2 \exp(\eta_{ij}) \frac{1 - \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^3},$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y} = x_{im} x_{iy} \exp(\eta_{ij}) \frac{1 - \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^3}, \quad [\ddot{g}(\eta_{ij})]_{\gamma_h \gamma_h} = \exp(\eta_{ij}) \frac{1 - \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^3}$$

e

$$[\ddot{g}(\eta_{ij})]_{\beta_m \gamma_h} = x_{im} \exp(\eta_{ij}) \frac{1 - \exp(\eta_{ij})}{[1 + \exp(\eta_{ij})]^3}.$$

• **Ligação complemento log-log**

$$g(\eta_{ij}) = \exp[-\exp(\eta_{ij})], \quad [\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})],$$

$$[\dot{g}(\eta_{ij})]_{\beta_y} = x_{iy} \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})], \quad [\dot{g}(\eta_{ij})]_{\gamma_h} = \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})],$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \beta_m} = (x_{im})^2 \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})][1 - \exp(\eta_{ij})],$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y} = x_{im} x_{iy} \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})][1 - \exp(\eta_{ij})],$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_h \gamma_h} = \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})][1 - \exp(\eta_{ij})]$$

e

$$[\ddot{g}(\eta_{ij})]_{\beta_m \gamma_h} = x_{im} \exp(\eta_{ij}) \exp[-\exp(\eta_{ij})][1 - \exp(\eta_{ij})].$$

• **Ligação log-log**

$$g(\eta_{ij}) = \exp[-\exp(-\eta_{ij})], \quad [\dot{g}(\eta_{ij})]_{\beta_m} = x_{im} \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})],$$

$$[\dot{g}(\eta_{ij})]_{\beta_y} = x_{iy} \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})], \quad [\dot{g}(\eta_{ij})]_{\gamma_h} = \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})],$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \beta_m} = (x_{im})^2 \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})][\exp(-\eta_{ij}) - 1],$$

$$[\ddot{g}(\eta_{ij})]_{\beta_m \beta_y} = x_{im} x_{iy} \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})][\exp(-\eta_{ij}) - 1],$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_h \gamma_h} = \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})][\exp(-\eta_{ij}) - 1]$$

e

$$[\ddot{g}(\eta_{ij})]_{\beta_m \gamma_h} = x_{im} \exp(-\eta_{ij}) \exp[-\exp(-\eta_{ij})][\exp(-\eta_{ij}) - 1].$$

• **Ligação proibito**

$$g(\eta_{ij}) = \Phi(\eta_{ij}), \quad [\dot{g}(\eta_{ij})]_{\beta_m} = x_{im}f(\eta_{ij}), \quad [\dot{g}(\eta_{ij})]_{\beta_y} = x_{iy}f(\eta_{ij}),$$

$$[\dot{g}(\eta_{ij})]_{\gamma_h} = f(\eta_{ij}), \quad [\ddot{g}(\eta_{ij})]_{\beta_m\beta_m} = -(x_{im})^2\eta_{ij}f(\eta_{ij}), \quad [\ddot{g}(\eta_{ij})]_{\beta_m\beta_y} = -x_{im}x_{iy}\eta_{ij}f(\eta_{ij}),$$

$$[\ddot{g}(\eta_{ij})]_{\gamma_h\gamma_h} = -\eta_{ij}f(\eta_{ij}) \quad \text{e} \quad [\ddot{g}(\eta_{ij})]_{\beta_m\gamma_h} = -x_{im}\eta_{ij}f(\eta_{ij}),$$

$\Phi(\cdot)$ é a função distribuição acumulada da distribuição da normal padrão e $f(\cdot)$ é a função densidade de probabilidade da distribuição da normal padrão.