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

A nova família de distribuições odd log-logística: teoria e aplicações

José Nilton da Cruz

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba
2016

José Nilton da Cruz
Licenciado em Matemática

A nova família de distribuições odd log-logística: teoria e aplicações
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba
2016

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Cruz, José Nilton da

A nova família de distribuições odd log-logística: teoria e aplicações / José Nilton da Cruz . - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

155 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Distribuição odd log-logística Weibull 2. Distribuição odd log-logística half-normal generalizada 3. Distribuição odd log-logística Fréchet 4. Momentos 5. Estatística de ordem 6. Modelo de regressão log-odd log-logístico Weibull 7. Envelope simulado 8. Censura informativa I. Título

CDD 519.532
C957n

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

“A alegria está na luta, na tentativa, no sofrimento envolvido e não na vitória propriamente dita.”

Mahatma Gandhi

DEDICATÓRIA

Aos meus queridos pais,

José da Cruz Sobrinho e Zildete Lopes da Cruz,
meu “porto seguro”, alicerce das minhas conquistas.

Aos meus irmãos, Zélia Lopes da Cruz, Manoel Messias da Cruz,
Maria Aparecida Lopes da Cruz, Antônio da Cruz Neto e
Zelma Lopes da Cruz, *pela amizade, carinho e apoio incondicional*

A eles, dedico este trabalho

AGRADECIMENTOS

A Deus, por sempre nortear minhas escolhas.

Meus amados pais, José da Cruz Sobrinho e Zildete Lopes da Cruz, pelo amor incondicional, ensinamentos e por me mostrarem que a simplicidade é uma virtude que pode serenar a inveja, decepções e preocupações.

Aos meus irmãos, cunhados e sobrinhos, pelos momentos de descontração, indispensáveis nos momentos de tensão.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela concessão da bolsa de doutorado, auxílio que contribuiu para que eu pudesse me dedicar exclusivamente as atividades de pós-graduação, a produção de artigos científicos e de eventos científicos.

Ao meu orientador, Prof. Dr. Edwin Moisés Marcos Ortega, a orientação, confiança, paciência e amizade que contribuíram incomensuravelmente para o meu desenvolvimento profissional e pessoal.

Ao Prof. Dr. Gauss Moutinho Cordeiro, por toda a colaboração e nos momentos de dúvidas ser sempre tão solícito.

Aos professores da banca de qualificação, Prof. Dr. Adriano Kamimura, Profa. Dra. Liciane Vaz de Arruda Silveira e a Profa. Dra. Sônia Maria de Stefano Piedade, pelas sugestões e contribuições valiosíssimas para desenvolvimento deste trabalho.

A minha amiga e “irmã científica”, Elizabeth Mie Hashimoto, pela amizade, ajuda e ensinamentos durante a elaboração deste trabalho.

Aos meus incansáveis amigos, Sérgio e Marisol, por me aturarem, auxiliarem e proporcionarem momentos ímpares na minha vida.

Aos meus amigos, Kuang e Fabi pela amizade, ajuda, conselhos e momentos de alegria.

Aos meus amigos de república, Pedro Amoedo, Djair, Elias, Ricardo Klein, Guto e Júlio, pelo prazer incomensurável de poder viver momentos alegres ao lado dos mesmos.

Aos meus amigos Rodrigo Pescim e Altemir, pela amizade e discussões sobre análise de sobrevivência que muito enriqueceram meu conhecimento à respeito.

A todos os amigos dos cursos de mestrado e doutorado do Departamento de Ciências Exatas Da Esalq/USP (alguns egressantes), em especial a Iábita, Adriele, Josiane, Valiana, Marcelo, Pedro Amoedo, Iuri, Edilan, Davi, Rick, Erasnilson, Cássio, Ezequiel, Rafael Moral, Alessandra, Maria Cristina, Marina, Lucas, pela amizade, conselhos e momentos de descontração.

A todos os professores do programa de Pós-graduação em Estatística e Experimentação Agronômica da ESALQ/USP.

Aos professores, funcionários e estagiários do Departamento de Estatística da UFMT, pela amizade e incentivo.

A todos aqueles que contribuíram, direta ou indiretamente, neste trabalho meus sinceros agradecimentos.

SUMÁRIO

RESUMO	11
ABSTRACT	13
1 INTRODUÇÃO	15
Referências	18
2 A FAMÍLIA ODD LOG-LOGÍSTICA: TEORIA E APLICAÇÕES	23
Resumo	23
Abstract	23
2.1 Introdução	24
2.2 Modelos especiais	26
2.2.1 O modelo odd log-logístico Weibull (OLLW)	26
2.2.2 O modelo odd log-logístico Fréchet (OLLFr)	27
2.2.3 O modelo odd log-logístico Half-Normal generalizada (OLLHNG)	29
2.2.4 O modelo odd log-logístico log-Logístico (OLLLL)	30
2.2.5 O modelo odd log-logístico Lognormal (OLLLN)	32
2.3 Assíntotas e formas	34
2.4 Expansões úteis	34
2.5 Série de potência quantílica	36
2.6 Momentos	38
2.7 Entropias	40
2.8 Valores extremos	41
2.9 Estatísticas de ordem	43
2.10 Extensão bivariada	44
2.11 Estimação por máxima verossimilhança	45
2.11.1 Dados com censura à direita	45
2.11.2 Dados de sobrevivência com censura intervalar	47
2.12 Aplicações	49
2.12.1 Mortalidade de pessoas aposentadas	49
2.12.2 Longevidade de moscas adultas	51
2.12.3 Estudo Signal-Tandmobiel	52
2.13 Conclusões finais	54
Referências	54
3 O MODELO DE REGRESSÃO LOG-ODD LOG-LOGÍSTICO WEIBULL: MO- DELAGEM, ESTIMAÇÃO, ANÁLISE DE DIAGNÓSTICOS DE INFLUÊNCIA E RESIDUAL	57
Resumo	57
Abstract	57
3.1 Introdução	58

3.2	A distribuição LOLLW	59
3.3	Propriedades matemáticas	61
3.4	Modelo de regressão log-odd log-Logístico Weibull	67
3.4.1	Estimação por máxima verossimilhança	68
3.4.2	Estudo de simulação	69
3.4.3	Probabilidade de cobertura	70
3.5	Estudo de má-especificação do modelo	71
3.6	Influência local	72
3.7	Análise de resíduos	74
3.7.1	Resíduos martingale	74
3.7.2	Resíduo deviance modificado	75
3.8	Estudos de simulação	75
3.9	Aplicação	82
3.9.1	Transplante de coração em Stanford	83
3.9.2	Análises de sensibilidade e residual	86
3.9.3	Modelo final	89
3.10	Conclusões	90
	Referências	91
4	O MODELO DE REGRESSÃO LOG-ODD LOGÍSTICA WEIBULL COM CEN- SURA INFORMATIVA	95
	Resumo	95
	Abstract	95
4.1	Introdução	96
4.2	Modelo de regressão log-odd log-Logístico Weibull	98
4.3	Censura informativa no modelo de regressão	100
4.4	Análise de sensibilidade	105
4.5	Análise residual	108
4.6	Aplicação	110
4.7	Análise de influência local	114
4.7.1	Perturbação de casos	114
4.7.2	Perturbação da variável resposta	115
4.7.3	Impacto das observações influentes detectadas	116
4.7.4	Análise residual	116
4.8	Conclusões	117
	Referências	118
5	MODELO DE REGRESSÃO LOG-ODD LOG-LOGÍSTICO WEIBULL PARA QUALIDADE DE VIDA DE SAÚDE BUCAL	121
	Resumo	121
	Abstract	121

5.1 Introdução	122
5.2 Modelos de resposta bivariada usando as cópulas de Frank-LOLLW e Clayton- LOLLW	123
5.3 Modelo de regressão e métodos de inferência	128
5.4 Diagnósticos de influência	130
5.4.1 Perturbação de casos	131
5.4.2 Perturbação da variável resposta	132
5.4.3 Perturbação de uma variável explanatória	133
5.5 Aplicação: dados de qualidade de vida relacionada à saúde bucal	133
5.6 Análise de sensibilidade	140
5.7 Considerações Finais	144
Referências	144
APÊNDICES	147

RESUMO

A nova família de distribuições odd log-logística: teoria e aplicações

Neste trabalho, foi proposta uma nova família de distribuições, a qual permite modelar dados de sobrevivência quando a função de risco tem formas unimodal e U (banheira). Ainda, foram consideradas as modificações das distribuições Weibull, Fréchet, half-normal generalizada, log-logística e lognormal. Tomando dados não-censurados e censurados, considerou-se os estimadores de máxima verossimilhança para o modelo proposto, a fim de verificar a flexibilidade da nova família. Além disso, um modelo de regressão localização-escala foi utilizado para verificar a influência de covariáveis nos tempos de sobrevida. Adicionalmente, conduziu-se uma análise de resíduos baseada nos resíduos deviance modificada. Estudos de simulação, utilizando-se de diferentes atribuições dos parâmetros, porcentagens de censura e tamanhos amostrais, foram conduzidos com o objetivo de verificar a distribuição empírica dos resíduos tipo martingale e deviance modificada. Para detectar observações influentes, foram utilizadas medidas de influência local, que são medidas de diagnóstico baseadas em pequenas perturbações nos dados ou no modelo proposto. Podem ocorrer situações em que a suposição de independência entre os tempos de falha e censura não seja válida. Assim, outro objetivo desse trabalho é considerar o mecanismo de censura informativa, baseado na verossimilhança marginal, considerando a distribuição log-odd log-logística Weibull na modelagem. Por fim, as metodologias descritas são aplicadas a conjuntos de dados reais.

Palavras-chave: Distribuição odd log-logística Weibull; Distribuição odd log-logística half-normal generalizada; Distribuição odd log-logística Fréchet; Momentos; Estatística de ordem; Modelo de regressão log-odd log-logístico Weibull; Envelope simulado; Censura informativa

ABSTRACT

The new family of odd log-logistic distributions: theory and applications

In this study, a new family of distributions was proposed, which allows to model survival data when the function of risk has unimodal shapes and U (bathtub). Modifications of the Weibull, Fréchet, generalized half-normal, log-logistic and lognormal distributions were considered. Taking censored and non-censored data, we consider the maximum likelihood estimators for the proposed model, in order to check the flexibility of the new family. Also, it was considered a location-scale regression model, to verify the influence of covariates on survival times. Additionally, a residual analysis was conducted based on modified deviance residuals. For different parameters fixed, percentages of censoring and sample sizes, several simulation studies were performed with the objective of verify the empirical distribution of the martingale type and modified deviance residuals. To detect influential observations, measures of local influence were used, which are diagnostic measures based on small perturbations in the data or in the proposed model. It can occur situations in which the assumption of independence between the failure and censoring times is not valid. Thus, another objective of this work is to consider the informative censoring mechanism based on the marginal likelihood, considering the log-odd log-logistic Weibull distribution in modelling. Finally, the methodologies described are applied to sets of real data.

Keywords: Odd log-logistic Weibull distribution; Odd log-logistic generalized half-normal distribution; Odd log-logistic Fréchet; Moments; Order statistics; Log-odd log-logistic Weibull regression model; Simulated envelope; Informative censoring

1 INTRODUÇÃO

A análise de sobrevivência é um conjunto de procedimentos estatísticos usados para analisar dados em que a variável resposta é o tempo até a ocorrência de um evento de interesse. Este tempo é denominado de tempo de falha, e este pode caracterizar o tempo até a morte de um paciente, aparecimento de um tumor, recidiva de uma doença, falha de um dispositivo eletrônico (análise de confiabilidade) e etc.

O tempo de sobrevivência de um indivíduo (observação), t , é considerado como o valor de uma variável aleatória T não-negativa. Ainda, supondo que essa variável T tenha função densidade de probabilidade $f(t)$, a função de sobrevivência, definida como a probabilidade do tempo de sobrevida ser maior ou igual a t , é dada por

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - \int_0^t f(u)du = 1 - F(t).$$

Desta maneira, a função de sobrevivência pode ser utilizada para representar a probabilidade de um indivíduo sobreviver a partir da origem (tempo de entrada no estudo) para além de um tempo t .

Outra função de suma importância em análise de sobrevivência, é a função de risco. Amplamente usada para expressar o risco de morte (risco de falha), em algum momento t , e é obtida como a probabilidade de que um indivíduo morra (falhe) no tempo t , dado que ele sobreviveu a esse tempo. Uma função de risco crescente (decrecente) indica que a taxa de falha do paciente aumenta (diminui) com o passar do tempo. Função de risco constante indica que a taxa de falha não se altera com o transcorrer do tempo. Uma função de risco em forma de banheira (ou de U), indica que o risco de falha decresce no período inicial, fica constante na faixa intermediária e cresce na porção final. Por fim, uma função de risco unimodal, apresenta um risco crescente no período inicial e decrescente na fase seguinte.

A expressão matemática para a função de risco é dada por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log(S(t)).$$

A função de risco também é chamada de taxa de falha, taxa de morte instantânea, taxa de intensidade ou a força de mortalidade (COLLETT, 2003).

A principal característica em análise de sobrevivência é a presença de censura, que é a observação parcial da resposta. Essa ocorre quando o evento de interesse não é observado durante o período de estudo, o qual pode ocorrer posteriormente. Dentre as

causas para o aparecimento da censura em um estudo, estão: o fim do acompanhamento de um indivíduo ou dispositivo eletrônico, morte de um paciente por uma causa diferente da estudada, cura do indivíduo, abandono do tratamento (não adaptação, mudança de cidade e etc) entre outros fatores.

De acordo com Colosimo e Giolo (2006), as observações censuradas devem ser incorporadas na análise estatística por duas razões: (i) mesmo sendo incompletas, elas fornecem informações sobre o tempo de vida de pacientes e (ii) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar conclusões viciadas.

A condução de experimentos, muitas vezes, se dá de forma a não permitir a observação do tempo exato de ocorrência do evento de interesse, mas é sabido que o mesmo ocorreu dentro de um intervalo de tempo. Uma vez que o tempo de falha T não é conhecido exatamente, mas pertence a um intervalo, ou seja, $T \in [U, V]$, os dados são chamados de *dados com censura intervalar*. De acordo com Lindsey e Ryan (1998), os tempos exatos de falha, assim como censuras à esquerda e à direita, são casos especiais de dados com censura intervalar, sendo que $U = V$ caracteriza o tempo exato de falha, $U = 0$ para censura à esquerda e $V = \infty$ para censura à direita.

Muitos estudos envolvendo censuras à direita e intervalar são propostos na literatura. No contexto de modelos não-paramétricos e semi-paramétricos, Ning, Qin e Shen (2010) desenvolveram um teste escore não-paramétrico para dados censurados à direita com amostra viesada. Dehghan e Duschene (2011) propuseram uma simples modificação do estimador de Turnbull para estimação não-paramétrica de uma função de sobrevivência condicional dada uma covariável contínua, e Wang, Chen e Yan (2013) apresentaram um modelo de regressão dinâmico bayesiano para dados de sobrevida com censura intervalar, o qual é uma generalização do modelo de Cox. No que diz respeito a modelos paramétricos, Balakrishnan e Mitra (2013), desenvolveram o algoritmo EM para ajustar uma distribuição Gama para dados truncados à esquerda e censurados à direita. Hashimoto et al. (2013) introduziram e estudaram o modelo de regressão log-Gama generalizada para dados com censura intervalar, obtendo e comparando para os parâmetros do modelo, os estimadores de máxima verossimilhança, de *jackknife* e de *bootstrap* não-paramétrico.

Todos os trabalhos citados anteriormente pressupuseram o mecanismo de censura não-informativa. Contudo, em algumas situações, essa suposição de que os tempos de falha e de censura são independentes é violada. Como exemplo de uma situação em que a censura é informativa, suponha que indivíduos são retirados de um estudo por apresentarem alto risco de morte devido a efeitos colaterais de um tratamento particular. O efeito disso será que as taxas observadas de sobrevivência, sobre esse particular tratamento, aparecerão maiores do que deveriam ser, levando a conclusões incorretas. Para considerar censura informativa, modelos marginais e de fragilidade têm sido desenvolvidos na litera-

tura. Emoto e Matthews (1990) assumiram distribuição Weibull para os tempos de falha e de censura. Huang e Wolfe (2002) consideraram um modelo de fragilidade para censura informativa, assumindo que os tempos de falha e de censura tornam-se independentes quando condicionados à uma fragilidade. Ebrahimi e Molefe (2003) consideraram um modelo para censura dependente (informativa) e obtiveram um estimador consistente e assintoticamente Normal para a função de sobrevivência de uma amostra censurada. Trabalhos mais recentes podem ser citados, como Zhao, Liu e Xu (2012), que apresentaram um modelo de fragilidade semi-paramétrico aplicados a eventos de dados recorrentes com covariáveis dependentes do tempo e censura informativa. E, Freitas e Rodrigues (2013) que apresentaram o modelo de taxa de cura Exponencial padrão com censura informativa e realizaram um estudo de simulação a fim de verificar o impacto da censura informativa sobre os parâmetros do modelo.

Nesse trabalho serão considerados dados com censuras à direita e intervalar sob o mecanismo de censura não-informativo e dados com censura à direita sob o mecanismo informativo.

Devido a simplicidade e flexibilidade em acomodar diferentes formas de funções de risco, distribuições como Lognormal, log-logística e Weibull são muito utilizadas em análises estatísticas de tempos de sobrevida. Contudo, em algumas situações práticas, surge a necessidade de utilizar modelos que acomodem funções de taxas de falha unimodal e em forma de banheira. Nessa perspectiva, generalizações destas distribuições tem surgido ao longo dos anos. Mudholkar, Srivastava e Freimer (1995) apresentaram uma extensão da família Weibull: a família Weibull exponenciada, extendendo a Weibull para uma classe mais ampla com taxas de falha nas formas unimodal e banheira (adequadas quando não se tem riscos crescentes ou decrescentes) e computacionalmente convenientes para dados censurados. Outras generalizações podem ser vistas em Mudholkar, Srivastava e Kollia (1996) e Marshall e Olkin (1997).

Carrasco, Ortega e Paula (2008) propuseram a distribuição Weibull modificada generalizada, uma distribuição com quatro parâmetros capaz de modelar as formas unimodais e de banheira da função de risco e que tem, dentre outras, como casos particulares as distribuições Exponencial exponencializada, Weibull exponencializada e Rayleigh generalizada. Nessa mesma linha, Silva, Ortega e Cordeiro (2010) propuseram uma nova distribuição, a Beta Weibull modificada, com cinco parâmetros, tendo como casos particulares: Weibull modificada generalizada, Beta Weibull, Weibull exponenciada, Beta exponencial, Weibull modificada, dentre outras. Outros trabalhos recentes podem ser citados, tais como Cordeiro, Ortega e Nadarajah (2010), Gusmão, Ortega e Cordeiro (2011) e Cordeiro, Silva e Ortega (2013).

Na maioria dos dados relacionados a sobrevivência, o tempo é influenciado por

variáveis que elucidam determinadas características dos indivíduos. Assim, a inclusão de covariáveis em um modelo de regressão é uma maneira de representar a heterogeneidade em uma população. Essas, por sua vez, devem ser consideradas, de alguma forma, no modelo de modo a aumentar o seu poder de predição. No contexto de análise de sobrevivência são muito utilizados os modelos de regressão locação-escala ou tempos de vida acelerado (LAWLESS, 2003).

O objetivo deste trabalho será o estudo de uma nova família de distribuições, que modifica distribuições contínuas por meio da adição de um parâmetro de forma. Esta família de distribuições também tem como vantagem permitir a obtenção de funções de risco unimodais e banheira (dependendo da distribuição de base). No presente trabalho serão consideradas as modificações das distribuições Weibull, Fréchet, Gumbel, Half-Normal generalizada, log-Logística e Lognormal.

Além do estudo das distribuições, pretende-se considerar a aplicação dos modelos Odd log-logístico Weibull (OLLW), Odd log-Logístico Half-Normal generalizada (OLLHNG) e Odd log-logístico Fréchet (OLLFr) para dados com censura à direita e intervalar.

Posteriormente ao ajuste de um modelo, torna-se necessário verificar a adequacidade do mesmo ao conjunto de dados considerados, visto que a não-adequacidade pode levar a conclusões viesadas. Para tal, será considerada uma análise de resíduos baseada nos trabalhos de Cox e Snell (1968), Fleming e Harrington (1991), Klein e Moeschberger (1997), Collet (2003) e Ortega, Paula e Bolfarine (2008), que descreveram os resíduos de Cox-Snell, Martingale e Deviance modificado. Ainda como um dos requisitos nesta etapa, será construído o gráfico Normal de probabilidade com envelope simulado, para se averiguar a distribuição dos resíduos e serão discutidas medidas de diagnóstico de influência com intuito de detectar possíveis observações que podem causar alterações substanciais nas estimativas do modelo. Aqui, serão consideradas as medidas de influência local, sendo essas encontradas com base em pequenas perturbações na verossimilhança (ponderação de casos), na variável resposta ou em covariáveis (COOK, 1977, 1986).

Por fim, essas metodologias serão aplicadas a conjuntos de dados reais.

Referências

BALAKRISHNAN, N.; MITRA, D. Likelihood inference based on left truncated and right censored data from a Gamma distribution. **IEEE transactions on reliability**, New York, v.62, p.679-688, 2013.

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**, New York, v.52, p.4021-4039, 2008.

CHEN, G. Generalized log-Normal distributions with reliability application. **Computational Statistics and Data Analysis**, New York, v.19, p.309-319, 1995.

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. 367p.

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.

CORDEIRO, G.M.; ORTEGA, E.M.M.; NADARAJAH, S. The Kumaraswamy Weibull distribution with application to failure data. **Journal of the Franklin Institute**, London, v.347, p.1399-1429, 2010.

CORDEIRO, G.M.; SILVA, G.O.; ORTEGA, E.M.M. The Beta-Weibull Geometric distribution. **Statistics**, Berlin, v.47, p.34-78, 2013.

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.

DEHGHAN, M.H.; DUSCHESNE, T. A generalization of Turnbull's estimator for nonparametric estimation of the conditional survival function with interval-censored data. **Lifetime Data Anal**, New York, v.17, p.234-255, 2011.

EBRAHIMI, N.; MOLEFE, D. Survival function estimation when lifetime and censoring time are dependent. **Journal of Multivariate Analysis**, Amsterdam, v.87, p.101-132, 2003.

EMOTO, S.E.; MATTHEWS, P.C. A Weibull model for dependent censoring. **Annals of Statistics**, Philadelphia, v.18, p.1556-1577, 1990.

FLEMING, T.T.; HARRINGTON, D.P. **Counting process and survival analysis**. New York: Wiley, 1991. 448p.

FREITAS, L.A.; RODRIGUES, J. Standard Exponential cure rate model with informative censoring. **Communications in Statistics-Simulation and Computation**, London, v.42, p.8-23, 2013.

GUSMÃO, F.R.S.; ORTEGA, E.M.M.; CORDEIRO, G.M. The generalized inverse Weibull distribution. **Statistical Papers**, London, v.52, p.591-619, 2011.

HASHIMOTO, E.M.; ORTEGA, E.M.M.; CANCHO, V.G.; CORDEIRO, G.M. On estimation and diagnostics analysis in log-generalized Gamma regression model for interval-censored data. **Statistics**, Berlin, v. 47, p.379-398, 2013.

HUANG, X.; WOLFE, R.A. A frailty model for informative censoring. **Biometrics**, Washington, v.58, p.510-520, 2002.

KLEIN, J.P.; MOESCHBERGER, M.L. **Survival Analysis Techniques for Censored and Truncated Data**. New York: Springer-Verlang, 2003. 536p.

LAWLESS, J.F. **Statistical Models and Methods for Lifetime Data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 664p.

LINDSEY, J.C.; RYAN, L.M. Tutorial in Biostatistics: methods for interval-censored data. **Statistics in Medicine**, London, v.17, 219-238, 1998.

MARSHALL, A.W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the Exponential and Weibull families. **Biometrika**, London, v.84, p.641-652, 1997.

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

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 American Statistical Association**, New York, v.91, p.1575-1583, 1996.

NING, J.; QIN, J.; SHEN, Y. Non-parametric tests for right-censored data with biased sampling. **Journal of the Royal Statistical Society**, Chichester, v.72, p.609-630, 2010.

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

WANG, X.; CHEN, M.H.; YAN, J. Bayesian dynamic regression models for interval censored survival data with application to children dental health. **Lifetime Data Analysis**, New York, v.19, p.297-316.

SILVA, G.O.; ORTEGA, E.M.M.; CORDEIRO, G.M. The Beta modified Weibull distribution. **Lifetime Data Analysis**, New York, v.16, p.409-430, 2010.

ZHAO, X.; LIU, L.; XU, W. Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. **Biometrical Journal**, Berlin, v.54, p.585-599.

2 A FAMÍLIA ODD LOG-LOGÍSTICA: TEORIA E APLICAÇÕES

Resumo

Fornecer uma família ampla de distribuições é sempre precioso para a estatística. Neste trabalho, é proposto um novo método adicionando um parâmetro à uma distribuição contínua. Esse método leva a uma nova classe de distribuições, chamada família odd log-Logística, que pode ser facilmente interpretada. Alguns modelos especiais são discutidos. Propriedades matemáticas gerais dessa família são obtidas, incluindo os momentos ordinários e incompletos, funções quantílica e geradora, valores extremos assintóticos, estatísticas de ordem e dois tipos de entropias. Estimação por máxima verossimilhança e inferência para grandes amostras são usados para dados de sobrevivência com censura à direita. Um modelo baseado na família odd log-logística é proposto para modelar dados com censura intervalar. A potencialidade da nova família é demonstrada por meio de três conjuntos de dados reais. De fato, a nova família produz melhores ajustes que algumas distribuições conhecidas.

Palavras-chave: Distribuição exponenciada; Distribuição odd log-Logística Fréchet; Distribuição odd log-Logística Weibull; Estatísticas de ordem; Extensão bivariada; Momento probabilisticamente ponderado; Valores extremos

Abstract

Providing a wider family of distributions is always precious for statisticians. It is proposed a new method adding a parameter to a continuous distribution. This method leads to a new class of distributions called the odd log-Logistic family that can be easily interpreted. Some special models are discussed. Were derived general mathematical properties of this family including the ordinary and incomplete moments, generating and quantile functions, asymptotic extreme values, order statistics and two types of entropies. Maximum likelihood estimation and inference for large samples are addressed for survival data with right censoring. A model based on the odd log-logistic family is proposed for modeling interval-censored data. The potentiality of the new family is demonstrated by means of three real data sets. In fact, the new family can produce better fits than some well-known distributions.

Keywords: Exponentiated distribution; Odd log-logistic Fréchet distribution; Odd log-logistic Weibull distribution; Order statistics; Bivariate extension; Probability weighted moment; Extreme values

2.1 Introdução

A literatura estatística está preenchida com centenas de distribuições contínuas univariadas. Desenvolvimentos recentes focam em novas técnicas para construir modelos significativos. Mais recentemente, muitos métodos de introduzir um ou mais parâmetros para gerar novas distribuições têm sido propostos. Dentre estes métodos, a composição de algumas distribuições discretas e de tempo de vida têm sido propostas na vanguarda da modelagem de tempo de vida. Assim, muitas famílias de distribuições foram investigadas pela composição de algumas distribuições de tempo de vida e discretas truncadas. A distribuição log-Logística (LL), com um parâmetro de forma α , é um modelo muito útil para análise de sobrevivência e é uma alternativa para a distribuição Lognormal. Diferentemente da distribuição Weibull, a distribuição LL tem uma função de risco não-monótona, o que a torna apropriada para modelar dados de sobrevivência de câncer. Se $\alpha > 1$, a função de risco é unimodal e quando $\alpha = 1$, a função de risco decresce monotonicamente. O fato de sua função de distribuição acumulada ter uma forma fechada é, particularmente, muito útil para análise de dados de sobrevivência com censura. A LL compartilha de algumas propriedades das distribuições Lognormal e Normal (AHMAD; SINCLAIR; WERRITTY, 1998), isto é, se T tem uma distribuição LL, então $Y = \log(T)$ tem uma distribuição Logística. Algumas aplicações da distribuição LL são discutidas em economia para modelar a riqueza e renda (KLEIBER; KOTZ, 2003), e dados de vazão em hidrologia (ASHKAR; MAHDI, 2006). Collet (2003) também sugeriu essa distribuição para modelar os tempos de vida após transplante de coração.

Neste trabalho, propõe-se uma nova família de distribuições para estender algumas distribuições contínuas e obter algumas de suas propriedades estruturais. Dada uma função de densidade acumulada contínua de base $G(x; \xi)$, com um vetor de parâmetros ξ , a função de densidade acumulada da distribuição *odd log-Logística-G* (“OLL-G”), com um parâmetro de forma adicional $\alpha > 0$, é definida por

$$F(x) = \int_0^{\frac{G(x; \xi)}{1-G(x; \xi)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{G(x; \xi)^\alpha}{G(x; \xi)^\alpha + \bar{G}(x; \xi)^\alpha}. \quad (2.1)$$

Ainda, o parâmetro α pode ser escrito como

$$\alpha = \frac{\log \left[\frac{F(x)}{F(x)} \right]}{\log \left[\frac{G(x)}{\bar{G}(x)} \right]} \quad \text{e} \quad \bar{G}(x; \xi) = 1 - G(x; \xi).$$

Assim, o parâmetro α representa o quociente do log da razão entre distribuições gerada e de base. Note que não há uma função complicada na equação (2.1) em contraste com a família Beta generalizada (EUGENE; LEE; FAMOYE, 2002), a qual inclui dois

parâmetros extras e também envolve a função Beta incompleta. A distribuição de base $G(x; \boldsymbol{\xi})$ é claramente um caso especial de (2.1) quando $\alpha = 1$. Se $G(x; \boldsymbol{\xi}) = x/(1+x)$, ela reduz-se a distribuição LL. Muitas distribuições podem ser geradas a partir da equação (2.1). As distribuições odd log-Logística Fréchet e odd log-Logística Gama são obtidas tomando $G(x; \boldsymbol{\xi})$ como sendo distribuições acumuladas Fréchet e Gama, respectivamente. A função densidade de probabilidade (fdp) da nova família é dada por

$$f(x) = \frac{\alpha g(x; \boldsymbol{\xi}) \{G(x; \boldsymbol{\xi})[1 - G(x; \boldsymbol{\xi})]\}^{\alpha-1}}{\{G(x; \boldsymbol{\xi})^\alpha + [1 - G(x; \boldsymbol{\xi})]^\alpha\}^2}. \quad (2.2)$$

Também, a função de risco da nova família é dada por

$$h(x) = \frac{\alpha g(x; \boldsymbol{\xi}) G(x; \boldsymbol{\xi})^{\alpha-1}}{\bar{G}(x; \boldsymbol{\xi}) [G(x; \boldsymbol{\xi})^\alpha + \bar{G}(x; \boldsymbol{\xi})^\alpha]}. \quad (2.3)$$

Doravante, serão consideradas G e g como distribuição e função densidade de probabilidade de base, respectivamente, de uma variável aleatória T e F como uma variável aleatória gerada X tendo a função de densidade (2.2). A família de densidades OLL (2.2) permite uma maior flexibilidade em suas caldas e pode ser aplicada em muitas áreas da engenharia e biologia. Uma vez que essa família estende muitas distribuições conhecidas, algumas de suas propriedades matemáticas foram estudadas. Note que mesmo se $g(x)$ for uma distribuição simétrica, a distribuição $f(x)$ não será uma distribuição simétrica. A equação (2.1) tem propriedades tratáveis especialmente por simulações, uma vez que a função quantílica de X toma forma simples

$$y = Q_G \left(\frac{u^{1/\alpha}}{[1-u]^{1/\alpha} + u^{1/\alpha}} \right), \quad (2.4)$$

em que $Q_G(u) = G^{-1}(u)$ é a função quantílica da distribuição de base e $u \sim U(0, 1)$.

Para simplificar a notação, às vezes, será usada $G(x) = G(x; \boldsymbol{\xi})$. Seja Y uma variável aleatória descrevendo um sistema estocástico tendo uma distribuição de base G . A odd y que o sistema não trabalhará no tempo x é $G(x)/\bar{G}(x)$. Se o interesse estiver na modelagem de y por uma variável aleatória Y tendo um parâmetro da distribuição LL, pode-se escrever $\Pr(Y \leq y) = \Pi_Y(y)$. Substituindo y da função densidade acumulada da LL pela razão de odds $G(x; \boldsymbol{\xi})/\bar{G}(x; \boldsymbol{\xi})$, obtém-se (2.2).

O resto do trabalho está organizado como segue. Na Seção 2.2, são apresentados alguns modelos especiais da família OLL, correspondendo as distribuições de base Weibull, Fréchet, half-Normal generalizada, log-Logística e Lognormal. Na Seção 2.4, é fornecida uma representação na forma de mistura muito útil para a função de densidade OLL. Na Seção 2.5, é obtida uma série de potências para a função quantílica da nova família. Seus

momentos são derivados na Seção 2.6. Duas entropias são determinadas na Seção 2.7. Na Seção 2.8, são fornecidos os valores extremos e a Seção 2.9 trata das estatísticas de ordem. Na Seção 2.10, é fornecida uma extensão bivariada da nova família. Estimação por máxima verossimilhança é investigada na Seção 2.11 para dados de sobrevivência com censura à direita e intervalar. Três aplicações para dados reais são desenvolvidas na Seção 2.12. Alguns comentários e conclusões são dados na Seção 2.13.

2.2 Modelos especiais

Aqui, serão discutidas algumas distribuições OLL especiais, em que $\alpha > 0$ em todos os casos. A função densidade (2.1) será mais tratável quando a função densidade acumulada $G(x)$ e a função densidade de probabilidade $g(x)$ tiverem expressões analíticas simples.

2.2.1 O modelo odd log-logístico Weibull (OLLW)

A função densidade acumulada Weibull é dada por $G(x; \xi) = 1 - \exp \left[- \left(\frac{x}{\lambda} \right)^\gamma \right]$, em que $\xi = (\lambda, \gamma)^T$, $\gamma > 0$ é o parâmetro de forma e $\lambda > 0$ é o parâmetro de escala. Assim, a função de densidade OLLW (para $x \geq 0$) reduz-se a

$$f(x) = \frac{\alpha \gamma x^{\gamma-1} \left\{ \exp \left[- \left(\frac{x}{\lambda} \right)^\gamma \right] \right\}^\alpha \left\{ 1 - \exp \left[- \left(\frac{x}{\lambda} \right)^\gamma \right] \right\}^{\alpha-1}}{\lambda^\gamma \left\{ \left\{ 1 - \exp \left[- \left(\frac{x}{\lambda} \right)^\gamma \right] \right\}^\alpha + \left\{ \exp \left[- \left(\frac{x}{\lambda} \right)^\gamma \right] \right\}^\alpha \right\}^2}.$$

Gráficos da função de densidade OLLW para alguns valores dos parâmetros são apresentados nas Figuras 2.1 e 2.2 e das funções de risco da OLLW são mostrados na Figura 2.3. Para alguns valores de α no intervalo $(0, 1)$ e certos valores de λ e γ , a função de risco tem a forma de banheira. Para alguns valores de $\alpha > 1$ e combinações especiais de λ e γ , a função de risco é unimodal.

Figura 2.1 - Gráficos da função de densidade da OLLW para alguns valores dos parâmetros

Figura 2.2 - Gráfico da função de densidade OLLW para alguns valores dos parâmetros

Figura 2.3 - Gráficos da função de risco OLLW para alguns valores dos parâmetros

2.2.2 O modelo odd log-logístico Fréchet (OLLFr)

A função de densidade acumulada da distribuição Fréchet (para $x > 0$) é $G(x; \xi) = \exp \left\{ - \left(\frac{\sigma}{x} \right)^\lambda \right\}$, em que $\lambda > 0$, $\sigma > 0$ e $\xi = (\sigma, \lambda)^T$. Então, a função de densidade OLLFr (para $x > 0$) pode ser obtida de (2.2) como

$$f(x) = \frac{\alpha \lambda \sigma^\lambda \left\{ \exp \left[- \left(\frac{\sigma}{x} \right)^\lambda \right] \right\}^\alpha \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x} \right)^\lambda \right] \right\}^{\alpha-1}}{x^{\lambda+1} \left\{ \left\{ \exp \left[- \left(\frac{\sigma}{x} \right)^\lambda \right] \right\}^\alpha + \left\{ 1 - \exp \left[- \left(\frac{\sigma}{x} \right)^\lambda \right] \right\}^\alpha \right\}^2}.$$

Gráficos da função de densidade para alguns valores dos parâmetros são apresentados nas Figuras 2.4 e 2.5 e da função de risco da OLLFr são exibidos na Figura 2.6. Estas podem ser monotonicamente decrescente e unimodal, dependendo dos valores dos parâmetros.

Figura 2.4 - Gráficos da função de densidade OLLFr para alguns valores dos parâmetros

Figura 2.5 - Gráficos da função de densidade para alguns valores dos parâmetros

Figura 2.6 - Gráficos da função de risco da OLLFr para alguns valores dos parâmetros

2.2.3 O modelo odd log-logístico Half-Normal generalizada (OLLHNG)

A distribuição half-Normal generalizada (HNG) (COORAY; ANANDA, 2008) tem função densidade de probabilidade e função densidade acumulada (para $\lambda, \theta > 0$ e $x \geq 0$) dadas, respectivamente, por

$$g(x; \xi) = \sqrt{\frac{2}{\pi}} \left(\frac{\lambda}{x} \right) \left(\frac{x}{\theta} \right)^\lambda e^{-\frac{1}{2} \left(\frac{x}{\theta} \right)^{2\lambda}} \quad \text{e} \quad G(x; \xi) = 2\Phi \left[\left(\frac{x}{\theta} \right)^\lambda \right] - 1,$$

em que $\Phi(\cdot)$ é a função de densidade acumulada da Normal padrão e $\xi = (\lambda, \theta)^T$.

A função de densidade da OLLHNG reduz-se a

$$f(x) = \frac{\alpha \sqrt{\frac{2}{\pi}} \left(\frac{\lambda}{x} \right) \left(\frac{x}{\theta} \right)^\lambda e^{-\frac{1}{2} \left(\frac{x}{\theta} \right)^{2\lambda}} \left\{ \left[2\Phi \left[\left(\frac{x}{\theta} \right)^\lambda \right] - 1 \right] \left[2 - 2\Phi \left[\left(\frac{x}{\theta} \right)^\lambda \right] \right] \right\}^{\alpha-1}}{\left\{ \left[2\Phi \left[\left(\frac{x}{\theta} \right)^\lambda \right] - 1 \right]^\alpha + \left[2 - 2\Phi \left[\left(\frac{x}{\theta} \right)^\lambda \right] \right]^\alpha \right\}^2}.$$

Gráficos desta função de densidade para alguns valores dos parâmetros são apresentados nas Figuras 2.7 e 2.8. Gráficos da função de risco são mostradas na Figura 2.9. Elas são: (a) forma de banheira, (b) unimodal e (c) monotonicamente crescente e decrescente.

Figura 2.7 - Gráficos da função de densidade da OLLHNG para alguns valores dos parâmetros

Figura 2.8 - Gráficos da função de densidade da OLLHG para alguns valores dos parâmetros

Figura 2.9 - Gráficos da função de risco da OLLHG para alguns valores dos parâmetros

2.2.4 O modelo odd log-logístico log-Logístico (OLLLL)

A distribuição acumulada log-logística com 2 parâmetros (para $x > 0$) com parâmetro de escala $\lambda > 0$ e parâmetro de forma $\gamma > 0$ é dada por $G(x; \boldsymbol{\xi}) = \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-1}$, em que $\boldsymbol{\xi} = (\lambda, \gamma)^T$. A função de densidade OLLLL reduz-se a

$$f(x) = \frac{\alpha \frac{\gamma}{\lambda} \left(\frac{x}{\lambda}\right)^{\gamma-1} \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-2} \left\{ \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-1} \left[1 - \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-1}\right] \right\}^{\alpha-1}}{\left\{ \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-\alpha} + \left[1 - \left[1 + \left(\frac{x}{\lambda}\right)^{-\gamma}\right]^{-1}\right]^{\alpha} \right\}^2}.$$

Gráficos da função de densidade OLLLL para alguns valores dos parâmetros são apresentados nas Figuras 2.10 e 2.11 e da função de risco são dados na Figura 2.12. A função de risco pode ser unimodal e monotonicamente decrescente.

Figura 2.10 - Gráficos da função de densidade da OLLLL para alguns valores dos parâmetros

Figura 2.11 - Gráficos da função de densidade da OLLLL para alguns valores dos parâmetros

Figura 2.12 - Gráficos da função de risco da OLLL para alguns valores dos parâmetros

2.2.5 O modelo odd log-logístico Lognormal (OLLLN)

A distribuição Lognormal (LN) tem função densidade de probabilidade e função de densidade acumulada (para μ e $\sigma > 0$) dadas, respectivamente, por

$$g(x; \xi) = \frac{1}{\sqrt{2\pi}x\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{\log(x) - \mu}{\sigma} \right)^2 \right\} \quad \text{e} \quad G(x; \xi) = 1 - \Phi \left(\frac{-\log(x) + \mu}{\sigma} \right),$$

em que $x > 0$ e $\xi = (\mu, \sigma)^T$ denota os parâmetros do modelo. A função de densidade da OLLLN é dada por

$$f(x) = \frac{\alpha \exp \left\{ -\frac{1}{2} \left(\frac{-\log(x) + \mu}{\sigma} \right)^2 \right\} \left\{ \left[1 - \Phi \left(\frac{-\log(x) + \mu}{\sigma} \right) \right] \left[\Phi \left(\frac{-\log(x) + \mu}{\sigma} \right) \right] \right\}^{\alpha-1}}{\sqrt{2\pi}x\sigma \left\{ \left[1 - \Phi \left(\frac{-\log(x) + \mu}{\sigma} \right) \right]^\alpha + \left[\Phi \left(\frac{-\log(x) + \mu}{\sigma} \right) \right]^\alpha \right\}^2}.$$

Gráficos das função de densidade da OLLLN para alguns valores dos parâmetros são mostrados nas Figuras 2.13 e 2.14, e da função de risco da OLLLN são apresentados na Figura 2.15. Estas tem formas unimodal e decrescente.

Figura 2.13 - Gráficos da função de densidade da OLLN para alguns valores dos parâmetros

Figura 2.14 - Gráficos da função de densidade da OLLN para alguns valores dos parâmetros

Figura 2.15 - Gráficos da função de risco da OLLN para alguns valores dos parâmetros

2.3 Assíntotas e formas

proposição 1: As assíntotas das equações (2.1), (2.2) e (2.3) quando $G(x) \rightarrow 0$ são dadas por

$$\begin{aligned} F(x) &\sim G(x)^\alpha \quad \text{quando } G(x) \rightarrow 0, \\ f(x) &\sim \alpha g(x) G(x)^{\alpha-1} \quad \text{quando } G(x) \rightarrow 0, \\ h(x) &\sim \alpha g(x) G(x)^{\alpha-1} \quad \text{quando } G(x) \rightarrow 0. \end{aligned}$$

Proposição 2: As assíntotas das equações (2.1), (2.2) e (2.3) quando $x \rightarrow \infty$ são dadas por

$$\begin{aligned} 1 - F(x) &\sim \bar{G}(x)^\alpha \quad \text{quando } x \rightarrow \infty, \\ f(x) &\sim \alpha g(x) \bar{G}(x)^{\alpha-1} \quad \text{quando } x \rightarrow \infty, \\ h(x) &\sim \frac{\alpha g(x)}{\bar{G}(x)} \quad \text{quando } x \rightarrow \infty. \end{aligned}$$

As formas das funções densidade e de risco são descritas analiticamente. Os pontos críticos da função densidade da OLL-G são as raízes da equação:

$$\frac{g'(x)}{g(x)} + (\alpha - 1) \frac{g(x)}{G(x)} + (1 - \alpha) \frac{g(x)}{\bar{G}(x)} + 2\alpha g(x) \frac{G(x)^{\alpha-1} - \bar{G}(x)^{\alpha-1}}{G(x)^\alpha + \bar{G}(x)^\alpha} = 0. \quad (2.5)$$

Há mais que uma raiz para (2.5).

Os pontos críticos de $h(x)$ são obtidos da equação

$$\frac{g'(x)}{g(x)} + (\alpha - 1) \frac{g(x)}{G(x)} + \frac{g(x)}{\bar{G}(x)} + \frac{\alpha g(x) [G(x)^{\alpha-1} - \bar{G}(x)^{\alpha-1}]}{\bar{G}(x)^\alpha + G(x)^\alpha} = 0. \quad (2.6)$$

Há mais que uma raiz para (2.6).

2.4 Expansões úteis

Primeiro, define-se a distribuição G-exponenciada (“G-exp”) para uma distribuição de base arbitrária $G(x)$, $W \sim \text{Exp}^c G$, se W tem funções densidade acumulada e densidade de probabilidade dadas por

$$H_c(x) = G(x)^c \quad \text{e} \quad h_c(x) = c g(x) G(x)^{c-1},$$

respectivamente. Este modelo transformado é também chamado de distribuição de Lehmann tipo I, digamos $\text{Exp}^c(G)$. Para $c > 1$ ($c < 1$) e valores grandes de x , o fator multiplicativo $c G(x)^{c-1}$ é maior (menor) que um. A afirmação inversa também é válida para valores menores de x . A última, implica imediatamente que os momentos ordinários associados à função de densidade $h_c(x)$, são estritamente maiores (menores) que àqueles associados à densidade $g(x)$ quando $c > 1$ ($c < 1$).

A seguir, uma expansão para $F(x)$ será obtida. Primeiro, usa-se uma série de potências para $G(x)^\alpha$ ($\alpha > 0$, real) dada por

$$G(x)^\alpha = \sum_{k=0}^{\infty} a_k G(x)^k, \quad (2.7)$$

em que

$$a_k = a_k(\alpha) = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{\alpha}{j} \binom{j}{k}.$$

Para qualquer real $\alpha > 0$, considera-se a expansão binomial generalizada

$$[1 - G(x)]^\alpha = \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} G(x)^k. \quad (2.8)$$

Inserindo (2.7) e (2.8) na equação (2.1), obtém-se

$$F(x) = \frac{\sum_{k=0}^{\infty} a_k G(x)^k}{\sum_{k=0}^{\infty} b_k G(x)^k},$$

em que $b_k = a_k + (-1)^k \binom{\alpha}{k}$.

A razão das duas séries de potência pode ser expressa como

$$F(x) = \sum_{k=0}^{\infty} c_k G(x)^k, \quad (2.9)$$

em que os coeficientes c_k 's (para $k \geq 0$) são determinados a partir da equação de recorrência

$$c_k = b_0^{-1} \left(a_k - b_0^{-1} \sum_{r=1}^k b_r c_{k-r} \right).$$

A função densidade de probabilidade de X segue, derivando (2.9), como

$$f(x) = \sum_{k=0}^{\infty} c_{k+1} h_{k+1}(x), \quad (2.10)$$

em que $h_{k+1}(x) = (k+1) G(x)^k g(x)$ é a função de densidade G-exp, com parâmetro de potência $(k+1)$.

A equação (2.10) revela que a função de densidade OLL-G é uma mistura linear das densidades G-exp. Portanto, algumas propriedades estruturais da nova família, tais como momentos ordinários e incompletos e função geradora, podem ser imediatamente obtidas das propriedades bem estabelecidas das distribuições G-exp.

As propriedades das distribuições G-exp foram estudadas por muitos autores nos últimos anos, ver Mudholkar e Srivastava (1993) e Mudholkar, Srivastava e Freimer (1995) para Weibull exponenciada, Gupta et al. (1998) para Pareto exponenciada, Gupta e Kundu (1999) para Exponencial exponenciada, Nadarajah (2005) para Gumbel exponenciada, Shirke e Kakade (2006) para log-Normal exponenciada e Nadarajah e Gupta (2007) para distribuições Gama exponenciadas. Ver também, Nadarajah e Kotz (2006), entre outros. As equações (2.9) e (2.10) são os principais resultados desta Subseção.

2.5 Série de potência quantílica

Nesta seção, uma série de potências para a função quantílica $y = Q(u) = F^{-1}(u)$ de X , será obtida por meio da expansão de (2.4). Se a função quantílica G, $Q_G(u) = G^{-1}(u)$, não tiver uma expressão de forma fechada, usualmente pode ser expressa em termos de uma série de potência

$$Q_G(u) = \sum_{i=0}^{\infty} a_i u^i, \quad (2.11)$$

em que os coeficientes a_i são números reais adequadamente escolhidos, os quais dependem dos parâmetros da distribuição G. Para muitas distribuições importantes, tais como as distribuições Normal, t -Student, Gama e Beta, $Q_G(u)$ não tem expressões explícitas, mas ela pode ser expandida como na equação (2.11). Como um exemplo simples, para a distribuição $N(0, 1)$, $a_i = 0$ para $i = 0, 2, 4, \dots$, $a_1 = 1$, $a_3 = 1/6$, $a_5 = 7/120$ e $a_7 = 127/7560, \dots$

Todos os resultados do artigo de Gradshteyn e Ryzhik (2000, Seção 0.314) serão utilizados. Desta forma, para uma série de potência elevado a um número inteiro positivo

n (para $n \geq 1$)

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i u^i \right)^n = \sum_{i=0}^{\infty} c_{n,i} u^i, \quad (2.12)$$

em que os coeficientes $c_{n,i}$ (para $i = 1, 2, \dots$) são facilmente obtidos da equação de recorrência (com $c_{n,0} = a_0^n$)

$$c_{n,i} = (i a_0)^{-1} \sum_{m=1}^i [m(n+1) - i] a_m c_{n,i-m}. \quad (2.13)$$

Claramente, $c_{n,i}$ pode ser determinado de $c_{n,0}, \dots, c_{n,i-1}$ e em seguida a partir das quantidades $a_0, \dots, a_i$.

Em seguida, uma expansão para o argumento de $Q_G(\cdot)$ em (2.4) é obtida

$$A = \frac{u^{\frac{1}{\alpha}}}{u^{\frac{1}{\alpha}} + (1-u)^{\frac{1}{\alpha}}}.$$

Posteriormente, define-se $(\alpha)_i = \alpha(\alpha-1)\dots(\alpha-i+1)$ como fatorial decrescente (GRAHAM; KNUTH; PATASHNIK, 1994, pg. 48). Usando a expansão binomial generalizada para $u \in (0, 1)$, podemos escrever

$$A = \frac{\sum_{k=0}^{\infty} \alpha_k^* u^k}{\sum_{k=0}^{\infty} b_k^* u^k} = \sum_{k=0}^{\infty} \delta_k u^k,$$

em que

$$\alpha_k^* = \alpha_k \left(\frac{1}{\alpha} \right) \quad \text{e} \quad b_k^* = \alpha_k \left(\frac{1}{\alpha} \right) + (-1)^k \left(\frac{1}{\alpha} \right),$$

e o coeficiente δ_k (para $k \geq 1$) segue da equação de recorrência (2.8)

$$\delta_k = \frac{1}{b_0^*} \left(\alpha_k^* - \frac{1}{b_0^*} \sum_{r=1}^k b_r^* \delta_{k-r} \right).$$

Portanto, a função quantílica de X pode ser expressa de (2.4) como

$$Q(u) = Q_G \left(\sum_{k=0}^{\infty} \delta_k u^k \right). \quad (2.14)$$

Para uma distribuição de base G, (2.11) e (2.14) são combinadas, obtendo-se

$$Q(u) = Q_G \left(\sum_{m=0}^{\infty} \delta_m u^m \right) = \sum_{i=0}^{\infty} a_i \left(\sum_{m=0}^{\infty} \delta_m u^m \right)^i,$$

e então, usando (2.12) e (2.13), tem-se

$$Q(u) = \sum_{m=0}^{\infty} e_m u^m, \quad (2.15)$$

em que $e_m = \sum_{i=0}^{\infty} a_i d_{i,m}$, e para $i = 0, 1, \dots$, $d_{i,0} = \delta_0^i$ e (para $m > 1$)

$$d_{i,m} = (m \delta_0)^{-1} \sum_{n=1}^m [n(i+1) - m] \delta_n d_{i,m-n}.$$

A equação (2.15) é o principal resultado desta Subseção, uma vez que permite obter várias quantidades matemáticas da OLL-G.

As fórmulas obtidas ao longo do trabalho podem ser facilmente manipuladas na maioria das plataformas de *software* de computação simbólica, tais como Maple, Mathematica e Matlab. Estas plataformas têm atualmente a capacidade de lidar com expressões analíticas de tamanho e complexidade formidáveis. Expressões explícitas, estabelecidas para calcular medidas estatísticas, podem ser mais eficientes do que o cálculo diretamente por integração numérica. O limite infinito dessas somas podem ser substituídos por um número inteiro positivo grande, tal como 30 ou 50, para a maioria dos fins práticos.

2.6 Momentos

Os momentos ordinários de Y podem ser obtidos do (r, j) -ésimo momentos probabilisticamente ponderados (MPPs) da variável aleatória Z , resultando em G (digamos, $\tau_{r,j}$). Definindo $G(x) = u$ e usando a função quantílica paterna $Q_G(x) = G^{-1}(x)$, tem-se que

$$\tau_{r,j} = E[Z^r G(Z)^j] = \int_{-\infty}^{\infty} z^r G(z)^j g(z) dz = \int_0^1 Q_G(u)^r u^j du. \quad (2.16)$$

Então, o r -ésimo momento de X pode ser expresso a partir de (2.10) e (2.16) como

$$E(X^r) = \sum_{j=0}^{\infty} t_j \tau_{r,j}, \quad (2.17)$$

em que $t_j = (j+1)c_{j+1}$. Sendo assim, os momentos de qualquer distribuição OLL-G podem ser expressos como uma soma infinita de MPPs. Em seguida, são apresentadas três fórmulas explícitas de $E(X^r)$ de (2.17) para os modelos especiais.

A primeira distribuição considerada é a OLLFr, introduzida na Seção 2.2. O (r, j) -ésimo MPP da distribuição de Fréchet configura $u = (j+1) \left(\frac{\sigma}{x}\right)^\lambda$ como

$$\tau_{r,j} = \lambda \sigma^\lambda \int_0^\infty x^{r-(\lambda+1)} \exp \left\{ -(j+1) \left(\frac{\sigma}{x}\right)^\lambda \right\} dx = -\frac{\sigma^r}{(j+1)^{1-r/\lambda}} \int_0^\infty u^{-r/\lambda} e^{-u} du.$$

A integral converge absolutamente para $r < \lambda$, e assim

$$E(X^r) = -\sigma^r \Gamma \left(1 - \frac{r}{\lambda}\right) \sum_{j=0}^\infty \frac{t_j}{(j+1)^{1-r/\lambda}}.$$

A segunda, é a OLLLL introduzida na Seção 2.4. Usando um resultado de Prudnikov, Brychkov e Marichev (1986, Seção 2.6.13, equação 4) pode-se escrever de (2.16) (para $t < 1$)

$$\tau_{r,j} = \left(\frac{\partial}{\partial t} \right)^r B(t+j+1, 1-t) \Big|_{t=0},$$

em que $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \Gamma(a) \Gamma(b) / \Gamma(a+b)$ é a função beta. O r -ésimo momento da distribuição OLLL (para $t < 1$) segue de (2.17) como

$$E(X^r) = \alpha \beta \sum_{j=0}^\infty t_j \left(\frac{\partial}{\partial t} \right)^r B(t+j+1, 1-t) \Big|_{t=0}.$$

Por último, a terceira, é a distribuição odd log-Logística Exponencial (OLLE). Para a função acumulada da exponencial $G(x) = 1 - \exp(-\lambda x)$, $\lambda > 0$, a função quantílica é $Q_G(u) = -\lambda^{-1} \log(1-u)$. Da equação (2.16), $\tau_{r,j}$ pode ser escrito como

$$\tau_{r,j} = (-1)^r \lambda^{-r} \int_0^1 u^j [\log(1-u)]^r du = r! \lambda^r \sum_{m=0}^\infty \frac{(-1)^{m+r}}{(m+1)^{r+1}} \binom{j}{m}.$$

Portanto, o r -ésimo momento da distribuição OLLE pode ser facilmente determinado de (2.17) por

$$E(X^r) = r! \lambda^r \sum_{j,m=0}^\infty \frac{(-1)^{m+r} t_j}{(m+1)^{r+1}} \binom{j}{m}.$$

2.7 Entropias

A entropia é uma medida de variação ou incerteza de uma variável aleatória X . Duas medidas populares de entropia são as entropias de Rényi e de Shannon (SHANNON, 1948; RÉNYI, 1961). A entropia de Rényi, de uma variável aleatória com fdp $f(x)$, é definida como

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left(\int_0^\infty f^\gamma(x) dx \right),$$

para $\gamma > 0$ e $\gamma \neq 1$. A entropia de Shannon, de uma variável aleatória X , é definida por $E \{-\log [f(X)]\}$. Ela é um caso especial da entropia de Rényi quando $\gamma \uparrow 1$. Cálculos diretos produzem

$$\begin{aligned} E \{-\log [f(X)]\} &= -\log(\alpha) - E \{\log [g(X; \xi)]\} + (1-\alpha) E \{\log [G(x; \xi)]\} \\ &\quad + (1-\alpha) E \{\log [\bar{G}(x; \xi)]\} + 2 E \{G(x; \xi)^\alpha + \bar{G}(x; \xi)^\alpha\}. \end{aligned}$$

Primeiro, define-se e calcula-se

$$A(a_1, a_2, a_3; \alpha) = \int_0^1 \frac{u^{a_1}(1-u)^{a_2}}{[u^\alpha + (1-u)^\alpha]^{a_3}} du, \quad a_1, a_2, a_3, \alpha > 0, \quad (2.18)$$

usando a expansão binomial generalizada

$$A(a_1, a_2, a_3; \alpha) = \sum_{i=0}^{\infty} (-1)^i \binom{a_2}{i} \int_0^1 \frac{u^{a_1+i}}{[u^\alpha + (1-u)^\alpha]^{a_3}} du. \quad (2.19)$$

Mas,

$$\frac{u^{a_1+i}}{[u^\alpha + (1-u)^\alpha]^{a_3}} = \frac{\sum_{k=0}^{\infty} \beta_k^* u^k}{\sum_{k=0}^{\infty} \gamma_k^* u^k} = \sum_{k=0}^{\infty} \delta_k^* u^k, \quad (2.20)$$

em que $\beta_k^* = a_k(a_1 + i)$, $\gamma_k^* = h_k(\alpha, a_3)$ e $\delta_k^* = \delta_k^*(a_1, i, \alpha, a_3) = \frac{1}{\gamma_0^*} \left[\beta_k^* - \frac{1}{\gamma_0^*} \sum_{r=1}^k \gamma_r^* \delta_{k-r}^* \right]$.

Então,

$$A(a_1, a_2, a_3; \alpha) = \sum_{i,k=0}^{\infty} \frac{(-1)^i \binom{a_2}{i} \delta_k^*(a_1, i, \alpha, a_3)}{k+1}. \quad (2.21)$$

Proposição: Seja X uma variável aleatória com fdp dada na equação 2.2. Então,

$$E \{ \log [G(X)] \} = \alpha \frac{\partial}{\partial t} A(\alpha + t - 1, \alpha - 1, 2; \alpha) \Big|_{t=0} \quad (2.22)$$

$$E \{ \log [\bar{G}(X)] \} = \alpha \frac{\partial}{\partial t} A(\alpha - 1, \alpha + t - 1, 2; \alpha) \Big|_{t=0} \quad (2.23)$$

$$E \{ G(x; \boldsymbol{\xi})^\alpha + \bar{G}(X; \boldsymbol{\xi})^\alpha \} = \alpha \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2 - t; \alpha) \Big|_{t=0}. \quad (2.24)$$

A fórmula mais simples para a entropia de X é dada por

$$\begin{aligned} E \{ -\log[f(X)] \} &= -\log(\alpha) - E \{ \log [g(X; \boldsymbol{\xi})] \} \\ &+ (1 - \alpha) \alpha \frac{\partial}{\partial t} A(\alpha + t - 1, \alpha - 1, 2; \alpha) \Big|_{t=0} \\ &+ (1 - \alpha) \alpha \frac{\partial}{\partial t} A(\alpha - 1, \alpha + t - 1, 2; \alpha) \Big|_{t=0} \\ &+ 2 \alpha \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2 - t; \alpha) \Big|_{t=0}. \end{aligned}$$

Após alguns desenvolvimentos algébricos, uma expressão alternativa para $I_R(\gamma)$ é obtida

$$I_R(\gamma) = \frac{\gamma}{1-\gamma} \log(\alpha) + \frac{1}{1-\gamma} \log \left[\sum_{l,k=0}^{\infty} w_{l,k}^* E_{Y_k} (g^{\gamma-1} [G^{-1}(Y)]) \right],$$

em que $Y_k \sim B(k+1, 1)$, $w_{l,k}^* = \frac{(-1)^l \binom{\gamma(\alpha-1)}{l} \lambda_k^*}{k+1}$, $\rho_k^* = a_k(\gamma(\alpha-1) + l)$, $\epsilon_k = h_k(\alpha, 2\gamma)$ e $\lambda_k = \frac{1}{\epsilon_0} (\rho_k - \frac{1}{\epsilon_0} \sum_{r=1}^k \epsilon_r \lambda_{k-r})$.

2.8 Valores extremos

Se $\bar{X} = (X_1 + \dots + X_n)/n$ denota a média de uma amostra aleatória de (2.2), então, pelo Teorema Central do Limite, $\sqrt{n}(\bar{X} - E(X))/\sqrt{Var(\bar{X})}$ aproxima-se para a distribuição Normal padrão, quando $n \rightarrow \infty$, sob condições de regularidade. Às vezes, o interesse pode estar na assíntota dos valores extremos $M_n = \max(X_1, \dots, X_n)$ e $m_n = \min(X_1, \dots, X_n)$.

Primeiro, supondo que G pertence ao domínio de atração do máximo da distribuição valor extremo Gumbel. Então, por Leadbetter, Lindgren e Rootzén (1983, Capítulo 1),

existirá uma função estritamente positiva, $h(t)$, tal que

$$\lim_{t \rightarrow \infty} \frac{1 - G(t + xh(t))}{1 - G(t)} = e^{-x},$$

para todo $x \in (-\infty, \infty)$. Mas

$$\lim_{t \rightarrow \infty} \frac{1 - F(t + xh(t))}{1 - F(t)} = \lim_{x \rightarrow \infty} \frac{xf(tx)}{f(t)} = e^{-\alpha x},$$

para todo $x \in (-\infty, \infty)$. Assim, segue de Leadbetter, Lindgren e Rootzén (1987, Capítulo 1), que F pertence ao domínio de atração do máximo da distribuição valor extremo Gumbel com

$$\lim_{n \rightarrow \infty} P[a_n(M_n - b_n \leq x)] = \exp[-\exp(-ax)]$$

para algumas constantes de normalização adequadas $a_n > 0$ e b_n . Segundo, supondo que G pertence ao domínio de atração do max da distribuição valor extremo Fréchet. Então, de Leadbetter, Lindgren e Rootzén (1983, Capítulo 1), existirá um $\beta > 0$ tal que

$$\lim_{t \rightarrow \infty} \frac{1 - G(t + xh(t))}{1 - G(t)} = x^c$$

para todo $x \in (-\infty, \infty)$. Todavia,

$$\lim_{t \rightarrow \infty} \frac{1 - F(t + xh(t))}{1 - F(t)} = \lim_{t \rightarrow \infty} \frac{xf(tx)}{f(t)} = x^{\alpha c},$$

para todo $x > 0$. Desta forma, segue de Leadbetter, Lindgren e Rootzén (1983, Capítulo 1), que F pertence ao domínio de atração do max da distribuição valor extremo Gumbel com

$$\lim_{n \rightarrow \infty} P[a_n(M_n - b_n \leq x)] = \exp(-x^{c\alpha})$$

para constantes de normalização adequadas $a_n > 0$ and b_n . Terceiro, supondo que G pertence ao domínio de atração do max da distribuição valor extremo da Weibull. Então, de Leadbetter, Lindgren e Rootzén (1983, Capítulo 1), existirá um $\beta > 0$ tal que

$$\lim_{t \rightarrow 0} \frac{G(tx)}{G(t)} = x^\beta$$

para todo $x < 0$. Entretanto,

$$\lim_{t \rightarrow 0} \frac{F(tx)}{F(t)} = \lim_{t \rightarrow 0} \frac{xf(tx)}{f(t)} = x^{\alpha\beta}$$

para todo $x < 0$. Dessa maneira, segue de Leadbetter, Lindgren e Rootzén (1983, Capítulo 1), que F pertence ao domínio de atração do max da distribuição valor extremo Weibull

com

$$\lim_{n \rightarrow \infty} P[a_n(M_n - b_n \leq x)] = \exp[-(-x)^{\alpha\beta}]$$

para algumas constantes de normalização adequadas $a_n > 0$ e b_n . Portanto, conclui-se que F pertence ao mesmo domínio de atração do $\max$ que o de G .

O mesmo argumento aplica para o domínio de atração do mínimo, ou seja, F pertence ao mesmo domínio de atração do $\min$ que o de G .

2.9 Estatísticas de ordem

Estatísticas de ordem aparecem em muitas áreas da teoria e da prática estatística.

Sejam $X_1, \dots, X_n$ uma amostra aleatória da distribuição OLL-G. Considere $X_{i:n}$ denotando a i -ésima estatística de ordem de X em uma amostra de tamanho n . A fdp de $X_{i:n}$ é dada por

$$f_{i:n}(x) = K f(x) F^{i-1}(x) \{1 - F(x)\}^{n-i} = K \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F(x)^{j+i-1},$$

em que $K = n! / [(i-1)!(n-i)!]$.

Assim, a função de densidade da i -ésima estatística de ordem da OLL-G pode ser expressa por

$$f_{i:n}(x) = \sum_{r,k=0}^{\infty} \sum_{j=0}^{n-i} m_{j,r,k} h_{r+k+1}(x). \quad (2.25)$$

Aqui, $h_{r+k+1}(x)$ denota a função densidade G-exp com parâmetro de potência $r+k+1$,

$$m_{j,r,k} = \frac{n! (r+1) (i-1)! c_{r+1}}{(r+k+1)} \sum_{j=0}^{n-i} \frac{(-1)^j f_{j+i-1,k}}{(n-i-j)! j!},$$

e c_k é definido na Subseção 2.4. Além disso, as quantidade $f_{j+i-1,k}$ são determinadas recursivamente por $f_{j+i-1,0} = c_0^{j+i-1}$ e (para $k \geq 1$)

$$f_{j+i-1,k} = (k c_0)^{-1} \sum_{m=1}^k [m(j+i) - k] c_m f_{j+i-1,k-m}.$$

Baseado na equação (2.25), pode-se facilmente obter momentos ordinários e incompletos e função geradora das estatísticas de ordem da OLL-G para qualquer distribuição de base G .

2.10 Extensão bivariada

Define-se a distribuição *odd log-Logística bivariada* (OLL-GB) pela função densidade acumulada (fda)

$$F_{X,Y}(x, y) = \frac{G(x, y; \boldsymbol{\xi})^\alpha}{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha},$$

em que $G(x, y; \boldsymbol{\xi})$ é uma distribuição contínua bivariada com fda's marginais $G_1(x; \boldsymbol{\xi})$ e $G_2(y; \boldsymbol{\xi})$. As funções densidades acumuladas marginais correspondentes são dadas por

$$F_X(x) = \frac{G_1(x; \boldsymbol{\xi})^\alpha}{G_1(x; \boldsymbol{\xi})^\alpha + [1 - G_1(x; \boldsymbol{\xi})]^\alpha} \quad \text{e} \quad F_Y(y) = \frac{G_2(y; \boldsymbol{\xi})^\alpha}{G_2(y; \boldsymbol{\xi})^\alpha + [1 - G_2(y; \boldsymbol{\xi})]^\alpha}.$$

A fdp conjunta de (X, Y) é facilmente obtida por $f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$, e então

$$f_{X,Y}(x, y) = \frac{\alpha A(x, y; \alpha, \boldsymbol{\xi}) G(x, y; \boldsymbol{\xi})^{\alpha-1} [1 - G(x, y; \boldsymbol{\xi})]^{\alpha-1}}{\{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha\}^2},$$

em que

$$\begin{aligned} A(x, y; \alpha, \boldsymbol{\xi}) = & g(x, y; \boldsymbol{\xi}) + \frac{\alpha - 1}{G(x, y; \boldsymbol{\xi})} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial x} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial y} \\ & + \frac{1 - \alpha}{1 - G(x, y; \boldsymbol{\xi})} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial x} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial y} \\ & - 2\alpha \frac{G(x, y; \boldsymbol{\xi})^{\alpha-1} - [1 - G(x, y; \boldsymbol{\xi})]^{\alpha-1}}{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial x} \frac{\partial G(x, y; \boldsymbol{\xi})}{\partial y}. \end{aligned}$$

As fdp's marginais são dadas por

$$f_X(x) = \frac{\alpha g_1(x; \boldsymbol{\xi}) G_1(x; \boldsymbol{\xi})^{\alpha-1} [\bar{G}_1(x; \boldsymbol{\xi})]^{\alpha-1}}{\{G_1(x; \boldsymbol{\xi})^\alpha + \bar{G}_1(x; \boldsymbol{\xi})^\alpha\}^2}$$

e

$$f_Y(y) = \frac{\alpha g_2(y; \boldsymbol{\xi}) G_2(y; \boldsymbol{\xi})^{\alpha-1} [\bar{G}_2(y; \boldsymbol{\xi})]^{\alpha-1}}{\{G_2(y; \boldsymbol{\xi})^\alpha + \bar{G}_2(y; \boldsymbol{\xi})^\alpha\}^2}.$$

As fda's condicionais seguem como

$$F_{X|Y}(x, y) = \frac{G(x, y; \boldsymbol{\xi})^\alpha \{G_2(y; \boldsymbol{\xi})^\alpha + \bar{G}_2(y; \boldsymbol{\xi})^\alpha\}}{\{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha\} G_2(y; \boldsymbol{\xi})^\alpha}$$

e

$$F_{Y|X}(y, x) = \frac{G(x, y; \boldsymbol{\xi})^\alpha \{G_1(x; \boldsymbol{\xi})^\alpha + \bar{G}_1(x; \boldsymbol{\xi})^\alpha\}}{\{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha\} G_1(x; \boldsymbol{\xi})^\alpha}.$$

As funções de densidade condicionais são escritas como

$$f_{X|Y}(x, y) = \frac{A(x, y; \alpha, \boldsymbol{\xi}) G(x, y; \boldsymbol{\xi})^{\alpha-1} [\bar{G}(x, y; \boldsymbol{\xi})]^{\alpha-1} \{[G_2(y; \boldsymbol{\xi})^\alpha + \bar{G}_2(y; \boldsymbol{\xi})^\alpha]^2\}}{\{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha\}^2 \{g_2(y; \boldsymbol{\xi}) G_2(y; \boldsymbol{\xi})^{\alpha-1} [\bar{G}_2(y; \boldsymbol{\xi})]^{\alpha-1}\}}$$

e

$$f_{Y|X}(x, y) = \frac{A(x, y; \alpha, \boldsymbol{\xi}) G(x, y; \boldsymbol{\xi})^{\alpha-1} [\bar{G}(x, y; \boldsymbol{\xi})]^{\alpha-1} \{[G_1(x; \boldsymbol{\xi})^\alpha + \bar{G}_1(x; \boldsymbol{\xi})^\alpha]^2\}}{\{G(x, y; \boldsymbol{\xi})^\alpha + [1 - G(x, y; \boldsymbol{\xi})]^\alpha\}^2 g_1(x; \boldsymbol{\xi}) G_1(x; \boldsymbol{\xi})^{\alpha-1} [\bar{G}_1(x; \boldsymbol{\xi})]^{\alpha-1}}.$$

2.11 Estimação por máxima verossimilhança

2.11.1 Dados com censura à direita

As estimativas dos parâmetros da distribuição OLL-G foram determinadas pelo método de máxima verossimilhança. Seja X_i uma variável aleatória seguindo (2.2) com vetor de parâmetros $\boldsymbol{\tau} = (\alpha, \boldsymbol{\xi}^T)^T$. Os dados encontrados em estudos na análise de sobrevivência e confiabilidade são frequentemente censurados. Um mecanismo de censura aleatória muito simples e que é muitas vezes realista, é aquele em que cada indivíduo i é assumido como tendo um tempo de vida X_i e um tempo de censura C_i , em que X_i e C_i são variáveis aleatórias independentes. Suponhamos que os dados consistem de n observações independentes $x_i = \min(X_i, C_i)$ para $i = 1, \dots, n$. A distribuição de C_i não depende de nenhum dos parâmetros desconhecidos de X_i . A inferência paramétrica para estes dados é usualmente baseada no método de verossimilhança e sua teoria assintótica. O log-verossimilhança $l(\boldsymbol{\tau})$, considerando censura, para os parâmetros do modelo é

$$\begin{aligned} l(\boldsymbol{\tau}) = & r \log(\alpha) + \sum_{i \in F} \log[g(x_i; \boldsymbol{\xi})] + (\alpha - 1) \sum_{i \in F} \{\log[G(x_i; \boldsymbol{\xi})] + \log[1 - G(x_i; \boldsymbol{\xi})]\} \\ & - 2 \sum_{i \in F} \{\log\{G(x_i; \boldsymbol{\xi})^\alpha + [1 - G(x_i; \boldsymbol{\xi})]^\alpha\}\} + \sum_{i \in C} \log \left\{ \frac{\bar{G}(x_i; \boldsymbol{\xi})^\alpha}{G(x_i; \boldsymbol{\xi})^\alpha + \bar{G}(x_i; \boldsymbol{\xi})^\alpha} \right\}, \end{aligned} \quad (2.26)$$

em que r é o número de falhas e F e C denotam o conjunto de observações não-censuradas e censuradas, respectivamente. O logaritmo da função de verossimilhança pode ser maximizado diretamente usando o SAS (Proc NLMixed), o R (optim), a rotina MaxBFGS no programa de linguagem matricial Ox (ver, DOORNIK, 2007) ou resolvendo as equações não-lineares de verossimilhança obtidas derivando (2.26). Os componentes do vetor escore $U(\boldsymbol{\tau})$ são dados por

$$U_{\alpha}(\boldsymbol{\tau}) = \frac{r}{\alpha} + \sum_{i \in F} \left\{ \log[G(x_i; \boldsymbol{\xi})] + \log[\bar{G}(x_i; \boldsymbol{\xi})] - 2 \left\{ \frac{G(x_i; \boldsymbol{\xi})^{\alpha} \log[G(x_i; \boldsymbol{\xi})] + [\bar{G}(x_i; \boldsymbol{\xi})]^{\alpha} \log[\bar{G}(x_i; \boldsymbol{\xi})]}{G(x_i; \boldsymbol{\xi})^{\alpha} + [\bar{G}(x_i; \boldsymbol{\xi})]^{\alpha}} \right\} \right\} + \sum_{i \in C} \left\{ \log[\bar{G}(x_i; \boldsymbol{\xi})] - \frac{G(x_i; \boldsymbol{\xi})^{\alpha} \log[G(x_i; \boldsymbol{\xi})] + \bar{G}(x_i; \boldsymbol{\xi})^{\alpha} \log[\bar{G}(x_i; \boldsymbol{\xi})]}{G(x_i; \boldsymbol{\xi})^{\alpha} + \bar{G}(x_i; \boldsymbol{\xi})^{\alpha}} \right\},$$

$$U_{\xi_j}(\boldsymbol{\tau}) = \sum_{i \in F} \left\{ \frac{[\dot{g}(x_i; \boldsymbol{\xi})]_{\omega_j}}{g(x_i; \boldsymbol{\xi})} + (\alpha - 1) \left\{ \frac{[\dot{G}(x_i; \boldsymbol{\xi})]_{\omega_j}}{G(x_i; \boldsymbol{\xi})} - \frac{[\dot{\bar{G}}(x_i; \boldsymbol{\xi})]_{\omega_j}}{\bar{G}(x_i; \boldsymbol{\xi})} \right\} - 2\alpha \left\{ \frac{G(x_i; \boldsymbol{\xi})^{\alpha-1} [\dot{G}(x_i; \boldsymbol{\xi})]_{\omega_j} - [\bar{G}(x_i; \boldsymbol{\xi})]^{\alpha-1} [\dot{\bar{G}}(x_i; \boldsymbol{\xi})]_{\omega_j}}{G(x_i; \boldsymbol{\xi})^{\alpha} + [\bar{G}(x_i; \boldsymbol{\xi})]^{\alpha}} \right\} \right\} - \alpha \sum_{i \in C} \left\{ \bar{G}(x_i; \boldsymbol{\xi}) \left[\frac{1}{\bar{G}(x_i; \boldsymbol{\xi})} + \frac{G(x_i; \boldsymbol{\xi})^{\alpha-1} - \bar{G}(x_i; \boldsymbol{\xi})^{\alpha-1}}{G(x_i; \boldsymbol{\xi})^{\alpha} + \bar{G}(x_i; \boldsymbol{\xi})^{\alpha}} \right] \right\},$$

em que $[\dot{g}(x_i; \boldsymbol{\xi})]_{\omega_j} = \frac{\partial g(x_i; \boldsymbol{\xi})}{\partial \omega_j}$, $[\dot{G}(x_i; \boldsymbol{\xi})]_{\omega_j} = \frac{\partial G(x_i; \boldsymbol{\xi})}{\partial \omega_j}$ e $\bar{G}(x_i; \boldsymbol{\xi}) = 1 - G(x_i; \boldsymbol{\xi})$ para $j = 1, \dots, p$.

Para estimação intervalar e testes de hipóteses sobre os parâmetros do modelo, necessita-se da matriz de informação observada $J = J(\boldsymbol{\tau})$ com dimensões $(p+1) \times (p+1)$, cujos elementos podem ser obtidos mediante solicitação ao autor. Sob condições normais de regularidade, podemos construir regiões de confiança aproximadas para os parâmetros baseadas na distribuição Normal multivariada $N_{p+1}(0, J(\hat{\boldsymbol{\tau}})^{-1})$. Uma vez calculados os valores máximos da log-verossimilhança do modelo base e do modelo geral, a estatística da razão de verossimilhança (RV) pode ser usada para comparar o modelo OLL-G com alguns de seus sub-modelos. Por exemplo, o teste de $H_0 : \alpha = 1$ versus $H : H_0 \text{ não é verdadeira}$, é equivalente a comparar as distribuições OLL-G e G e a estatística RV reduz-se a $W = 2\{\ell(\hat{\alpha}, \hat{\boldsymbol{\xi}}) - \ell(1, \tilde{\boldsymbol{\xi}})\}$, em que $\hat{\alpha}$ e $\hat{\boldsymbol{\xi}}$ são as estimativas de máxima verossimilhança sob H , $\tilde{\boldsymbol{\xi}}$ são as estimativas sob H_0 e W segue uma distribuição de Qui-quadrado com 1 grau de liberdade.

2.11.2 Dados de sobrevivência com censura intervalar

Em muitos estudos, a resposta de sobrevivência pode ser censurada em intervalos, de forma que o evento de interesse não é observado exatamente, mas é somente conhecido a ocorrência dentro de intervalos de tempo que podem sobrepor-se ou variarem em comprimento. A literatura apresenta muitas aplicações de modelos de sobrevida para dados com censura intervalar, com respeito a família de distribuições Weibull (LAWLESS, 2003). Esta família é muito apropriada em situações nas quais as funções de risco são constante ou monótona.

Em algumas situações, os tempos dos eventos de interesse X (em que X_i é uma variável aleatória que segue (2.2) com vetor de parâmetros $\boldsymbol{\tau} = (\alpha, \boldsymbol{\xi}^T)^T$ somente são conhecidos como tendo ocorrido dentro de um intervalo de tempo $[U, V]$, em que $U \leq X \leq V$. Isso pode ocorrer em ensaios clínicos, por exemplo, quando pacientes são avaliados somente em visitas pré-agendadas, de onde ocorre que o evento X não foi observado como tendo ocorrido em uma visita no tempo U mas ocorreu em uma visita seguinte no tempo V . Dessa forma, só é sabido que T ocorreu em algum ponto no intervalo $[U, V]$. Portanto, esse é um exemplo de dados com censura intervalar. Note que as censuras à direita e à esquerda são casos especiais da censura intervalar. Assim, observa-se que $U = V$ caracteriza os dados sem censura (tempos exatos), $V = \infty$ para dados com censura à direita, e $U = 0$ para dados com censura à esquerda. Para dados com censura intervalar, observa-se que consistem de um intervalo (u_i, v_i) para cada indivíduo (observação), e tais intervalos sabe-se que incluem x_i com probabilidade um. Em outras palavras, $P[u_i \leq x_i \leq v_i] = 1$. Se $v_i = \infty$, tem-se um tempo censurado à direita para x_i . Este modelo é referido como família OLL-G para dados com censura intervalar. Seja $(u_1, v_1), \dots, (u_n, v_n)$ um conjunto de n observações com censura intervalar, em que u_i e v_i são os dados observados. O logaritmo da função de verossimilhança observada para o vetor de parâmetros $\boldsymbol{\tau} = (\alpha, \boldsymbol{\xi}^T)^T$ é dada por

$$\begin{aligned} l(\boldsymbol{\tau}) &= \sum_{i \in F^*} \log \left[\frac{G(u_i; \boldsymbol{\xi})^\alpha}{G(u_i; \boldsymbol{\xi})^\alpha + \bar{G}(u_i; \boldsymbol{\xi})^\alpha} - \frac{G(v_i; \boldsymbol{\xi})^\alpha}{G(v_i; \boldsymbol{\xi})^\alpha + \bar{G}(v_i; \boldsymbol{\xi})^\alpha} \right] \\ &\quad + \sum_{i \in C^*} \log \left[\frac{G(u_i; \boldsymbol{\xi})^\alpha}{G(u_i; \boldsymbol{\xi})^\alpha + \bar{G}(u_i; \boldsymbol{\xi})^\alpha} \right]. \end{aligned}$$

Aqui, F^* representa o conjunto de indivíduos com censura intervalar, isto é, $x_i \in (u_i, v_i)$. e C^* representa o conjunto de indivíduos com censura à direita, ou seja, $x_i \in (u_i, +\infty)$. As estimativas de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\tau}$ podem ser obtidas maximizando o logaritmo da função de verossimilhança. Nesse trabalho foi utilizada a função `optim` do R para calcular as estimativas de máxima verossimilhança. A estimativa da matriz de covariâncias dos estimadores de máxima verossimilhança $\hat{\boldsymbol{\tau}}$, sob condições padrões

de regularidade, pode ser obtida a partir da matriz hessiana. Intervalos de confiança e testes de hipóteses podem ser conduzidos usando uma distribuição amostral grande dos estimadores de máxima verossimilhança, ou seja, baseada na aproximação da distribuição Normal multivariada pela matriz de covariâncias dada pela inversa da matriz de informação observada.

Estudo de simulação

Aqui, um estudo de simulação de Monte Carlo foi realizado para avaliar o comportamento das estimativas de máxima verossimilhança de α , σ e λ . Os resultados foram obtidos a partir de 1.000 simulações de Monte Carlo realizadas no R. Para cada réplica, uma amostra de tamanho n é obtida a partir da distribuição OLLFr $(\alpha, \sigma, \lambda)$ e os parâmetros são obtidos pelo método de máxima verossimilhança. Desta forma, para cada combinação de n , α , σ e λ , 1.000 amostras foram geradas e ajustou-se o modelo OLLFr $(\alpha, \lambda, \sigma)$.

Neste estudo de simulação, considerou-se tamanhos amostrais iguais a $n = 50, 100, 200$ e 500 e $\alpha = 1, 5$, $\sigma = 2$ e $\lambda = 1$.

A Tabela 2.1 revela que os erros quadráticos médio (EQM) e os vieses diminuem à medida que o tamanho da amostra aumenta, como esperado.

Tabela 2.1 - Média, viés e Erro Quadrático Médio (EQM) das estimativas de máxima verossimilhança dos parâmetros do modelo OLLFr

n	Parâmetro	Média	Viés	E.Q.M.	n	Média	Viés	E.Q.M.
50	$\alpha = 1, 5$	2,1485	0,6485	9,0059	100	1,7513	0,2513	1,9769
	$\sigma = 2$	1,9113	-0,0887	0,2669		1,9563	-0,0437	0,1218
	$\lambda = 1$	1,1582	0,1582	0,3347		1,0563	0,0563	0,1269
n	Parâmetro	Média	Viés	E.Q.M.	n	Média	Viés	E.Q.M.
200	$\alpha = 1, 5$	1,5557	0,0557	0,3796	500	1,5280	0,0280	0,0668
	$\sigma = 2$	1,9866	-0,0134	0,0406		1,9938	-0,0062	0,0120
	$\lambda = 1$	1,0421	0,0421	0,0622		1,0051	0,0051	0,0201

2.12 Aplicações

2.12.1 Mortalidade de pessoas aposentadas

Aqui, um conjunto de dados analisados por Balakrishnan et al.(2009) foi utilizado para estudar o gerador da OLL aplicado as distribuições de base Weibull, Fréchet e half-Normal generalizada (HNG). Em cada caso, as estimativas dos parâmetros de cada distribuição são obtidas pelo método de máxima verossimilhança com a subrotina `optim` do R. Primeiro, descreveu-se o conjunto de dados. Posteriormente, são apresentados os EMV (e os correspondentes erros-padrão dados em parênteses) e os valores das estatísticas AIC (Akaike Information Criterion), CAIC (Consistent Akaike Information Criterion) e BIC (Bayesian Information Criterion). Os menores valores desses critérios define o melhor ajuste.

É importante para o Instituto mexicano de Seguro Social (IMSS) estudar o comportamento distribucional da mortalidade de pessoas aposentadas por invalidez, pois permite o cálculo da estimativa financeira de longo e curto prazo, tais como a avaliação das reservas obrigatórias para pagar as “pensões mínimas”. Os dados a serem utilizados representam o tempo de vida (em anos) de mulheres aposentadas, com incapacidades temporárias, que foram incorporadas no sistema público de seguro Mexicano e morreram em 2004.

As EMV e as estatísticas AIC, CAIC e BIC para os modelos Weibull, HNG e Fréchet e os correspondentes modelos OLL são listados na Tabela 2.2. Nota-se que, em comparação com os modelos de referência, aqueles gerados produzem valores menores das estatísticas AIC, CAIC e BIC. O histograma dos dados e funções densidades ajustadas são mostradas na Figura 2.16.

Testes formais para a assimetria dos parâmetros das distribuições geradas podem ser feitos baseados nas estatísticas de Razão de Verossimilhança (RV). Essas estatísticas são apresentadas na Tabela 2.3. A hipótese nula é rejeitada nos três testes a favor dos modelos OLLW, OLLHNG e OLLFr, respectivamente. A rejeição é altamente significativa e dá uma clara evidência do potencial do parâmetro extra de assimetria na modelagem de dados reais.

Tabela 2.2 - EMV (erros-padrão em parênteses) e as estatísticas AIC, CAIC e BIC

Modelo	Estimativa			AIC	CAIC	BIC
OLLW(α, λ, γ)	1,465 (0,2739)	52,979 (0,9731)	3,700 (0,6217)	2.113,694	2.113,781	2.124,599
Weibull(λ, γ)		51,928 (0,6566)	4,991 (0,2205)	2.117,746	2.117,79	2.125,016
OLLHNG(α, θ, λ)	2,052 (0,3939)	57,577 (1,7904)	2,176 (0,3876)	2.114,862	2.114,949	2.125,767
HNG(θ, λ)		54,066 (0,6977)	3,623 (0,1571)	2.142,454	2.142,498	2.149,724
OLLFr(α, σ, λ)	54,0295 (6,6042)	1,3039 (0,5336)	0,1021 (0,0117)	2.133,567	2.133,654	2.144,471
Fréchet(σ, λ)		41,327 (0,6586)	3,987 (0,1653)	2.208,645	2.208,689	2.215,915

Tabela 2.3 - Testes de RV

Modelos	Hipóteses	Estatística w	Valor- p
OLLW vs Weibull	$H_0 : \alpha = 1$ vs $H_1 : H_0$ se falsa	6,052	0,0139
OLLHNG vs HNG	$H_0 : \alpha = 1$ vs $H_1 : H_0$ se falsa	29,592	<0,0001
OLLFr vs Fréchet	$H_0 : \alpha = 1$ vs $H_1 : H_0$ se falsa	77,08	<0,0001

Figura 2.16 - (a) Densidades estimadas dos modelos Weibull e OLLW. (b) Densidades estimadas dos modelos HNG e OLLHNG. (c) Densidades estimadas dos modelos Fréchet e OLLFr

2.12.2 Longevidade de moscas adultas

Os dados usados vêm de um estudo realizado no departamento de Entomologia da Escola de Agricultura Luiz de Queiroz, Universidade de São Paulo, com o objetivo de avaliar a longevidade da mosca da fruta ou mosca do mediterrâneo (*ceratitis capitata*).

A necessidade dessa mosca de procurar comida, apenas depois da fase larval, tem permitido o uso de iscas tóxicas para seu controle em pomares brasileiros, por pelo menos 50 anos. Esta técnica de controle de pragas consiste na utilização de pequenas porções de alimentos contaminados com um inseticida, geralmente um organofosforado, que mata rapidamente as moscas, em vez de utilizar um inseticida sozinho. A variável resposta no experimento é o tempo de vida (em dias) de moscas adultas após serem expostas aos tratamentos. O período experimental foi de 51 dias, de modo que o número de larvas que sobreviveram, para além desse período, foram considerados como dados censurados. O tamanho total da amostra foi $n = 72$, uma vez que quatro observações foram perdidas. Para mais detalhes, ver Silva et al. (2013). Portanto, a variável aleatória usada neste estudo é o tempo de vida (em dias), X_i , de *ceratitis capitata* adultas. Do total de $n = 72$ observações, 9 são censuradas.

Tabela 2.4 - EMV (erros-padrão em parênteses) e estatísticas AIC, CAIC e BIC, considerando censura à direita, para os dados de *ceratitis capitata*

Modelo		Estimativa		AIC	CAIC	BIC
OLLW(α, λ, γ)	2,1325	29,3459	0,9505	1.279,55	1.279,693	1.288,993
	(0,7330)	(3,7855)	(0,3114)			
Weibull(λ, γ)		24,954	1,797	1.286,097	1.286,168	1.292,392
		(1,1215)	(0,1109)			
OLLGHN(α, θ, λ)	2,584	36,985	0,644	1.276,914	1.277,057	1.286,357
	(0,7840)	(6,5165)	(0,1872)			
GHN(θ, λ)		27,725	1,366	1.294,249	1.294,32	1.300,544
		(1,2161)	(0,0862)			
OLLFr(α, σ, λ)	28,5486	0,0950	0,0687	1.293,359	1.293,502	1.302,801
	(8,0559)	(0,1387)	(0,0188)			
Fréchet(σ, λ)		12,2011	0,9536	1.418,021	1.418,092	1.424,316
		(1,0413)	(0,0473)			

Testes formais para os parâmetros extras de assimetria nas distribuições geradas podem ser realizados com base na estatística de RV. Estas estatísticas para comparar os modelos ajustados são dadas na Tabela 2.5. Observa-se que a hipótese nula foi rejeitada nos três testes, o que favoreceu os modelos OLLW, OLLHNG e OLLFr.

Tabela 2.5 - Testes RV

Modelos	Hipóteses	Estatística w	Valor- p
OLLW vs Weibull	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	8,5466	0,0035
OLLGHN vs GHN	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	19,3344	<0,0001
OLLFr vs Fréchet	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	126,6620	<0,0001

Nas Figuras 2.17 (a), (b) e (c) são apresentadas as funções de sobrevivência empírica e estimadas dos modelos Weibull e OLLW, HNG e OLLHNG, Fréchet e OLLFr, respectivamente. Observa-se que as funções de sobrevivência dos modelos OLL estão mais próximos da função de sobrevivência empírica, evidenciando os resultados dos testes de RV listados na Tabela 2.5.

Figura 2.17 - Funções de sobrevivência estimadas dos modelos (a) Weibull e OLLW. (b) HNG e OLLHNG. (c) Fréchet e OLLFr

2.12.3 Estudo Signal-Tandmobiel

O conjunto de dados é referente a um estudo prospectivo longitudinal sobre saúde bucal, conduzido na cidade de Flanders (Bélgica) de 1996 à 2001.

O projeto Signal Tandmobiel foi um esforço colaborativo entre centros universitários (departamento de saúde bucal, departamento de saúde da juventude, centro de bioestatística), centros de saúde escolar, associação dental e da indústria. Os principais objetivos deste projeto foram o rastreamento das condições de saúde bucal de crianças de Flanders, com idades entre 7 e 12 anos, promoção de saúde bucal e educação no ambiente escolar e utilização de dados recolhidos para fins epidemiológicos.

Neste estudo, uma coorte de 4.430 estudantes que cursavam o primeiro ano da escola primária foi amostrada (2.297 meninos e 2.133 meninas). No início, os dentes de cada

criança foram anualmente examinados por um dos 16 dentistas treinados. O conjunto original de dados é composto por, no máximo, seis observações dentárias para cada criança, incluindo o tempo de aparecimento dos dentes, experiência de cáries e dados sobre hábitos alimentares e de higiene bucal. Para mais detalhes do delineamento do estudo e métodos de pesquisa empregados, ver Vanobbergen et al.(2000).

Em Gómez et al. (2009), a análise foi restringida para a distribuição dos tempos de emergência dos primeiros molares esquerdos (24 dentes na notação dentária Européia) e também foram excluídos 44 casos, para os quais a variável estado do dente predecessor da dentição principal era faltante.

Neste estudo, considerou-se a variável aleatória $X_i =$ idade (em anos) do aparecimento de 24 dentes permanentes. Das 4.386 observações consideradas, 1.611 são observações censuradas.

Testes formais para o parâmetro extra de assimetria nas distribuições geradas podem ser realizados com base na estatística de RV. Estas estatísticas, usadas para comparar os modelos ajustados, são dados na Tabela 2.7. A hipótese nula foi rejeitada nos três testes a favor dos modelos OLLW, OLLHNG e OLLLN.

Tabela 2.6 - EMV (erros-padrão em parênteses) e estatísticas AIC, CAIC e BIC para os dados de Signal-Tandmobiel

Modelo	Estimativa			AIC	CAIC	BIC
OLLW (α, λ, γ)	2,8225 (0,8440)	6,6710 (0,3834)	1,8843 (0,5744)	11.165,53	11.165,53	11.184,68
Weibull(λ, γ)		5,9698 (0,0217)	5,2954 (0,0843)	11.197,97	11.197,98	11.210,75
OLLGHN (α, θ, λ)	4,3165 (0,0696)	8,0342 (4,8671)	1,0131 (0,6137)	11.168,76	11.168,77	11.187,92
GHN(θ, λ)		6,1710 (0,0207)	4,3370 (0,0715)	11.296,38	11.296,38	11.309,15
OLLLN(α, μ, σ)	4,6360 (3,1496)	1,6970 (0,9252)	1,0450 (0,6137)	11.179,35	11.179,36	11.198,51
Lognormal(μ, σ)		1,6911 (0,0043)	0,2468 (0,0036)	11.222,55	11.222,55	11.235,32

Tabela 2.7 - Testes RV

Modelos	Hipóteses	Estatística w	Valor- p
OLLW vs Weibull	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	34,44	<0,0001
OLLGHN vs HNG	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	129,62	<0,0001
OLLLN vs Lognormal	$H_0 : \alpha = 1$ vs $H_1 : H_0$ é falsa	45,2	<0,0001

2.13 Conclusões finais

Distribuições para modelagem de dados, com qualquer suporte na literatura estatística, são numerosas. Neste trabalho, foi definida a família *odd log-Logística* a fim de proporcionar uma maior flexibilidade para qualquer distribuição contínua por meio da adição de apenas um parâmetro de forma. Alguns modelos especiais são brevemente discutidos. Demonstrou-se que a densidade da nova família pode ser expressa como uma combinação linear das densidades exponenciadas definidas a partir do modelo base. Foram investigadas algumas propriedades estruturais gerais da nova família, incluindo momentos ordinários e incompletos, função quantílica, entropias de Rényi e de Shannon e estatísticas de ordem. Uma série de potência para a função quantílica gerada foi obtida. Definiu-se uma extensão bivariada. Os parâmetros dos modelos foram estimados pelo método de máxima verossimilhança. A potencialidade da família é ilustrada por meio de três aplicações a dados reais.

Referências

- AHMAD, M.I.; SINCLAIR, C.D.; WERRITTY, A. Log-Logistic flood frequency Analysis. **Journal of Hydrology**, New York, v.98, p.205-224, 1988.
- ASHKAR, F.; MAHDI, S. Fitting the log-Logistic distribution by generalized moments. **Journal of Hydrology**, New York, v.328, p.694-703, 2006.
- BALAKRISHNAN, N.; LEIVA, V.; SANHUEZA, A.; CABRERA, E. Mixture inverse Gaussian distributions and its transformations, moments and applications. **Statistics**, Berlin, v.43, p.91-104, 2009.
- COLLETT, D. **Modelling survival data in medical research**. 2nd ed. London: Chapman & Hall, 2003. 391p.
- COORAY, K.; ANANDA, M.M.A. A generalization of the half-Normal distribution with applications to lifetime data. **Communications in Statistics - Theory and Methods**, New York, v.37, p.1323-1337, 2008.

DOORNIK, J. **Ox: Object-oriented matrix programming using Ox**. London: Timberlake Consultants Ltd, 2007. 461p.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-Normal distribution and its applications. **Communications in Statistics - Theory and Methods**, London, v.31, p.497-512, 2002.

GRADSHTEYN, I.S.; RYZHIK, I.M. **Table of integrals, series, and products**. San Diego: Academic Press, 2007. 315p.

GRAHAM, R.L.; KNUTH, D.E.; PATASHNIK, O. **Concrete mathematics: A foundation for computer science**, 2nd ed. California: Addison-Wesley, 1994. 625p.

GÓMEZ, G.; CALLE, M.L.; OLLER, R.; LANGOHR, K. Tutorial on methods for interval-censored data and their implementation in R. **Statistical Modelling**, New York, v.9, p.259-297, 2009.

GUPTA, R.D.; GUPTA, P.L.; GUPTA, R.D. Modeling failure time data by Lehman Alternatives. **Communications in Statistics - Theory and Methods**, London, v.27, p.887-904, 1998.

GUPTA, R.D.; KUNDU, D. Generalized Exponential distributions. **Australian and New Zealand Journal of Statistics**, Oxford, v.41, p.173-188, 1999.

KLEIBER, C.; KOTZ, S. **Statistical size distributions in economics and actuarial sciences**. New York: John Wiley, 2003, 327p.

LAWLESS, J.F. **Statistical models and methods for lifetime data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 664p.

LEADBETTER, M.R.; LINDGREN, G.; ROOTZÉN, H. **Extremes and related properties of random sequences and processes**, New York: Springer, 1983. 336p.

MUDHOLKAR, G.S.; SRIVASTAVA, D.K. Exponentiated Weibull family for analyzing bathtub failure-rate data. **IEEE Transactions on Reliability**, New York, v.42, p.299-302, 1993.

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

NADARAJAH, S. The exponentiated Gumbel distribution with climate application. **Environmetrics**, New York, 17, p.13-23, 2005.

NADARAJAH, S.; KOTZ, S. The exponential type distributions. **Acta applicandae mathematicae: an international survey journal on applying mathematics and mathematical applications**, Amsterdam, v.92, p.97-111, 2006.

NADARAJAH, S.; GUPTA, A.K. The exponentiated Gamma distribution with application to drought data. **Calcutta Statistical Association Bulletin**, Amsterdam, v.59, p.29-54, 2007.

PRUDNIKOV, A.P.; BRYCHKOV, Y.A.; MARICHEV, O.I. **Integrals and series**. New York: Gordon and Breach Science Publishers, 1986. p.527.

R, DEVELOPMENT, C. T. R: Language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2009. Disponível em: <http://www.R-project.org/>. Acesso em: 24 mar. 2014.

RÉNYI, A. On measures of entropy and information. **Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability**, Berkeley, v.1, p.547-561, 1961.

SHANNON, C.E. Prediction and entropy of printed english. **The Bell System Technical Journal**, London, v.30, p.50-64, 1951.

SHIRKE, D.T; KAKADE, C.S. On exponentiated log-Normal distribution. **International Journal of Agricultural and Statistical Sciences**, New Delhi, v.2, p.319-326, 2006.

SILVA, M.A.; BEZERRA-SILVA, G.C.D.; VENDRAMIN, J.D.; MASTRANGELO, T. Sublethal effect of neem extract on Mediterranean fruit fly adults. **Revista Brasileira de Fruticultura**, Jaboticabal, v.35, p.93-101, 2013.

VANOBERGEN, J.; MARTENS, L.; LESAFFRE, E.; DECLERCK, D. The Signal Tandmobiel[®] project - a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. **European Journal of Paediatric Dentistry**, Roma, v.2, p.87-96, 2000.

3 O MODELO DE REGRESSÃO LOG-ODD LOG-LOGÍSTICO WEIBULL: MODELAGEM, ESTIMAÇÃO, ANÁLISE DE DIAGNÓSTICOS DE INFLUÊNCIA E RESIDUAL

Resumo

Em aplicações de técnicas de análise de sobrevivência, a função taxa de falha pode, frequentemente, apresentar uma forma unimodal. Em tais casos, distribuições como a Lognormal ou log-Logística são usadas. Neste trabalho, preocupou-se somente com formas paramétricas e assim, um modelo de regressão localização-escala baseado na distribuição odd log-logística Weibull, é proposto para modelar dados com funções de taxa de falha decrescente, crescente, unimodal e banheira, como um modelo alternativo ao modelo de regressão log-Weibull. Assumindo dados censurados, foi considerado um estimador da análise clássica para os parâmetros do modelo proposto, o estimador de máxima verossimilhança. Derivou-se as matrizes apropriadas para avaliar influências locais sobre as estimativas dos parâmetros sob diferentes esquemas de perturbação. Além disso, para diferentes configurações dos parâmetros, tamanhos amostrais e porcentagens de censura, vários estudos de simulação foram realizados. Em adição, a distribuição empírica de alguns resíduos modificados são apresentadas e comparada com a distribuição Normal padrão. Esses estudos sugerem que a análise residual, geralmente realizada em modelos de regressão linear Normal, podem ser estendidos para um resíduo deviance modificado no modelo de regressão proposto, aplicado à dados censurados. Analisou-se um conjunto de dados reais usando o modelo de regressão log-odd log-Logístico Weibull.

Palavras-chave: Análise de sensibilidade; distribuição log-odd log-logística Weibull; Envelope simulado; Modelo de regressão localização-escala; Resíduos deviance modificado

Abstract

In technical applications of survival analysis, the failure rate function may frequently present a unimodal shape. In such case, the log-Weibull, Lognormal or Log-logistic distributions are used. This work, was concerned only with parametric forms, so a location-scale regression model based on the odd log-logistic Weibull distribution is proposed for modeling data with a decreasing, increasing, unimodal and bathtub failure rate functions as an alternative to the log-Weibull regression model. Assuming censored data, was considered a classic analysis estimator for the parameters of the proposed model, the likelihood maximum estimator. Were derived the appropriate matrices for assessing local influences on the parameter estimates under different perturbation schemes. Further, for different parameter settings, sample sizes and censoring percentages, various simulations were performed. In addition, the empirical distribution of some modified residuals are displayed and compared with the standard Normal distribution. These studies suggest that the residual analysis usually performed in Normal linear regression models can be extended to a modified deviance residual in the proposed regression model applied to censored data. Were analyzed a real data set using the log-odd log-Logistic Weibull regression model.

Keywords: Sensitivity Analysis; log-odd log-Logistic Weibull distribution; Simulated envelope; Location-scale regression model; Modified deviance residuals

3.1 Introdução

Os modelos de regressão podem ser propostos de diferentes formas em análise de sobrevivência. Dentre eles, o modelo de regressão locação-escala (LAWLESS, 2003) é diferenciado e frequentemente usado em ensaios clínicos. Neste trabalho foi proposto um modelo de regressão locação-escala, usando o logaritmo da distribuição odd log-Logística Weibull (OLLW), chamado modelo de regressão log-odd log-Logístico Weibull (LOLLW), como uma alternativa ao modelo de regressão log-Logístico para análise de tempos de sobrevida. Considerando que os modelos de regressão log-Exponencial e log-Weibull são modelos encaixados, o teste da razão de verossimilhança, por exemplo, pode ser usado para discernir entre estes modelos.

No presente estudo, uma análise clássica para o modelo de regressão LOLW foi considerada. A parte inferencial foi realizada utilizando a distribuição assintótica dos estimadores de máxima verossimilhança (EMV) que, em situações em que a amostra é pequena, podem apresentar resultados difíceis de serem justificados. Estudos, via simulação Monte Carlo, foram realizados para avaliar o desempenho do modelo de regressão LOLW, por meio da variância e do erro quadrático médio.

Após a modelagem, é importante verificar as suposições do modelo, bem como realizar um estudo de robustez, a fim de detectar as observações influentes ou extremas que podem causar distorções nos resultados da análise. Várias abordagens têm sido propostas na literatura para detectar observações influentes ou outliers. Uma eficiente maneira de detectar essas observações foi proposta por Cook (1986). Ele sugeriu que mais confiança pode ser depositada em um modelo que é relativamente estável a pequenas modificações.

Por outro lado, quando o caso deleção é usado, toda informação de um único indivíduo é deletada e, portanto, é difícil dizer se este indivíduo tem alguma influência sobre um aspecto específico do modelo. Uma solução para o problema anterior, pode ser encontrada em um paradigma completamente diferente, sendo uma abordagem de influência local em que novamente se investiga como os resultados de uma análise são alterados sob pequenas perturbações no modelo e em que estas perturbações podem ser interpretações específicas.

Recentemente, alguns autores têm investigado a avaliação da influência local em diferentes modelos de regressão com a presença e ausência de dados censurados. Por exemplo, Lemonte e Patriota (2011) consideraram o problema de avaliar influência local em modelos de regressão não-linear Birnbaum-Saunders, Rondon, Penagos e Cysneiros (2012) adaptaram métodos de influência local para modelos de regressão não-linear Birnbaum-Saunders. Matos et al. (2013) investigaram influência local em modelos lineares e não-lineares de efeitos mistos com dados censurados, Paula (2013) realizou cálculos de curvatura sob

vários esquemas de perturbação em modelos lineares generalizados. Fachini, Ortega e Cordeiro (2014) investigaram influência local em modelo de locação-escala para tempos de sobrevivida bivariado, baseado no modelo de cópula, objetivando modelar a dependência de dados de sobrevivência bivariados com fração de cura. Aqui, foi desenvolvida uma metodologia para detectar observações influentes no modelo de regressão LOLLW com dados censurados. Além disso, utilizou-se uma análise de resíduos para avaliar as pressuposições do modelo.

Na Seção 3.2, um breve estudo da distribuição OLLW é apresentado junto a parte inferencial deste modelo. Na Seção 3.3 é apresentada a distribuição LOLLW. Na Seção 3.4, sugeriu-se um modelo de regressão LOLLW na forma de locação-escala em adição aos estimadores de máxima verossimilhança (EMV) e resultados de vários estudos de simulação na Seção 3.5. Na Seção 3.6, usou-se medidas de diagnóstico considerando o esquema de perturbação baseado no caso deleção no modelo de regressão LOLLW com observações censuradas. Na Seção 3.7, um tipo de resíduo deviance foi proposto para avaliar características subjacentes da distribuição LOLLW e para detectar outliers. Também alguns estudos de simulação são discutidos na Seção 3.8. Finalmente, na Seção 3.9, um conjunto de dados reais é analisado e algumas conclusões são apresentadas na Seção 3.10.

3.2 A distribuição LOLLW

A literatura estatística está preenchida com centenas de distribuições univariadas contínuas. Desenvolvimentos recentes concentram-se em novas técnicas para a construção de modelos significativos. Mais recentemente, muitos métodos de introduzir um ou mais parâmetros para gerar novas distribuições têm sido propostos. A distribuição de Weibull de dois parâmetros tem sido o modelo mais popular para modelar dados de tempo de vida. No entanto, sua maior fraqueza é a incapacidade para acomodar taxas de risco não-monótonas (em particular, taxas de risco na forma de banheira). Isso tem levado a busca de generalizações dessa distribuição. Com base na função de distribuição acumulada (fda) $G(t; \gamma, \lambda) = 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right]$, em que $\gamma > 0$ é o parâmetro de forma e $\lambda > 0$ é o parâmetro de escala, a fda da distribuição OLLW, com um parâmetro adicional de forma $\alpha > 0$, é definida por

$$F(t; \gamma, \lambda, \alpha) = \int_0^{\frac{G(t; \gamma, \lambda)}{1 - G(t; \gamma, \lambda)}} \frac{\alpha t^{\alpha-1}}{(1 + t^\alpha)^2} dx = \frac{G(t; \gamma, \lambda)^\alpha}{G(t; \gamma, \lambda)^\alpha + \bar{G}(t; \gamma, \lambda)^\alpha}. \quad (3.1)$$

Ainda, pode-se escrever

$$\alpha = \frac{\log \left[\frac{F(t)}{F(t)} \right]}{\log \left[\frac{G(t)}{\bar{G}(t)} \right]} \quad \text{e} \quad \bar{G}(t; \gamma, \lambda) = 1 - G(t; \gamma, \lambda).$$

Deste modo, o parâmetro α representa o quociente do log da razão de chances para as distribuições gerada e de base. Note que não há uma função complicada na equação (3.1), em contraste com a família Beta generalizada (EUGENE; LEE; FAMOYE, 2002), que inclui a função beta incompleta e dois parâmetros extras de forma. A distribuição de Weibull $G(x; \gamma, \lambda)$ é claramente um caso particular de (3.1) quando $\alpha = 1$. Se $G(t; \gamma, \lambda) = t/(1 + t)$, tem-se a distribuição Log-logística.

Doravante, assume-se que a variável aleatória T tem uma distribuição OLLW com parâmetros $(\alpha, \gamma, \lambda)^T$. A função densidade de probabilidade (fdp) de T é dada por

$$f(t; \alpha, \gamma, \lambda) = \frac{\alpha \gamma t^{\gamma-1} \left\{ \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha \left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^{\alpha-1}}{\lambda^\gamma \left[\left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha + \left\{ \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha \right]^2}, \quad t > 0, \quad (3.2)$$

em que $\alpha > 0$ é o novo parâmetro de forma.

A função de sobrevivência de T é dada por

$$S(t; \alpha, \gamma, \lambda) = 1 - \frac{\left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha}{\left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha + \left\{ \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha}, \quad (3.3)$$

em que $t > 0$, $\alpha > 0$ e $\gamma > 0$ são os parâmetros de forma, e $\lambda > 0$ é um parâmetro de escala. A função taxa de falha (ftf) é dada por $h(t; \alpha, \gamma, \lambda) = f(t; \alpha, \gamma, \lambda)/S(t; \alpha, \gamma, \lambda)$. A grande flexibilidade deste modelo, para dados de tempo de vida, é devido as diferentes formas que a ftf pode assumir: (i) se $\alpha = 1$, então a ftf toma a forma da função de risco da distribuição de Weibull; (ii) para alguns valores de α no intervalo $(0, 1)$ e certos valores de λ e γ , a ftf pode ter a forma de banheira; (iii) para alguns valores de $\alpha > 1$ e combinações especiais de λ e γ , a ftf é unimodal.

Figura 3.1 - Gráficos da ftf da distribuição OLLW para alguns valores dos parâmetros

A variável aleatória $Y = \log(T)$ tem uma distribuição LOLLW, com função de densidade parametrizada em termos de $\gamma = \sigma^{-1}$ e $\mu = \log(\lambda)$, dada por

$$f(y; \alpha, \sigma, \mu) = \frac{\alpha \exp\left(\frac{y-\mu}{\sigma}\right) \left\{\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha \left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma \left\{\left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha + \left\{\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha\right\}^2}, \quad (3.4)$$

em que $-\infty < y < \infty$, $\alpha > 0$, $\sigma > 0$ e $-\infty < \mu < \infty$.

Portanto, se $T \sim \text{OLLW}(\alpha, \gamma, \lambda)$ então $Y = \log(T) \sim \text{LOLLW}(\alpha, \sigma, \mu)$. A função de sobrevivência de Y é dada por

$$S(y; \alpha, \sigma, \mu) = 1 - \frac{\left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha}{\left\{1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha + \left\{\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^\alpha}. \quad (3.5)$$

Gráficos da função de densidade (3.4), para valores dos parâmetros selecionados, são exibidos na Figura 3.2. Estes gráficos indicam grande flexibilidade para diferentes valores de α , μ e σ .

Figura 3.2 - Gráficos da fdp da LOLLW para alguns valores dos parâmetros. (a) Para $\mu = 2$ e $\sigma = 2$. (b) Para $\sigma = 2$. (c) Para diferentes valores de α , μ e σ

3.3 Propriedades matemáticas

Considere a variável aleatória padronizada $Z = (Y - \mu)/\sigma$, com função de densidade dada por

$$\pi(z) = \alpha \frac{\exp(z) \left\{\exp[-\exp(z)]\right\}^\alpha \left\{1 - \exp[-\exp(z)]\right\}^{\alpha-1}}{\left\{\left\{1 - \exp[-\exp(z)]\right\}^\alpha + \left\{\exp[-\exp(z)]\right\}^\alpha\right\}^2}, \quad -\infty < z < \infty. \quad (3.6)$$

A distribuição valor-extremo padrão é um caso especial de (3.6) quando $\alpha = 1$.

A seguir, será apresentada uma expressão geral que pode fornecer muitas propriedades estruturais para Z por uma simples transformação linear para a distribuição LOLIW.

Inicialmente, são apresentados três resultados matemáticos que serão usados nesta Seção. Primeiro, para $a > 0$ real, obtém-se a seguinte expansão binomial

$$(1 - u)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} u^j, \quad (3.7)$$

em que o coeficiente binomial é definido para qualquer real a .

Segundo, a seguinte expansão é válida para qualquer real $c > 0$

$$u^c = \sum_{r=0}^{\infty} p_r(c) u^r, \quad (3.8)$$

em que $p_r(c) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{c}{j} \binom{j}{r}.$

Terceiro, usando uma equação de Gradshteyn e Ryzhik (2000, Seção 0.314) para uma série de potências elevada a um número positivo i dada por

$$\left(\sum_{j=0}^{\infty} a_j v^j \right)^i = \sum_{j=0}^{\infty} c_{i,j} v^j, \quad (3.9)$$

em que os coeficientes $c_{i,j}$ (para $j = 1, 2, \dots$) são obtidos da equação de recorrência (para $j \geq 1$)

$$c_{i,j} = (ja_0)^{-1} \sum_{m=1}^j [m(i+1) - j] a_m c_{i,j-m} \quad (3.10)$$

e $c_{i,0} = a_0^i$. Consequentemente, os coeficientes $c_{i,j}$ podem ser calculados diretamente de $c_{i,0}, c_{i,1}, \dots, c_{i,j-1}$ e, portanto, de $a_0, a_1, \dots, a_j$. Eles podem ser apresentados explicitamente em termos de a_j 's, embora isso não seja necessário para programar numericamente essas expansões em qualquer *software* numérico ou algébrico.

Considerando $e^{-e^z} = u$, pode-se escrever de (3.6) que

$$E[W(Z)] = \int_{-\infty}^{\infty} W(z)\pi(z)dz = \alpha \int_0^1 W(\log[-\log(u)]) \frac{u^{\alpha-1}(1-u)^{\alpha-1}}{[u^\alpha + (1-u)^\alpha]^2}, \quad (3.11)$$

em que $W(\cdot)$ é a função geradora usada para obter algumas propriedades matemáticas de Z . Por exemplo, $W(z) = z^n$ e $W(z) = e^{tZ}$ correspondem aos momentos e a função geradora de momentos (fgm) de Z , respectivamente.

A partir das equações (3.7) e (3.8), obtém-se uma expansão para $u^\alpha + (1-u)^\alpha$, como segue

$$u^\alpha + (1-u)^\alpha = \sum_{j=0}^{\infty} t_j u^j,$$

em que $t_j = t_j(\alpha) = p_j(\alpha) \binom{\alpha}{j} + (-1)^j \binom{\alpha}{j}$. Então, usando as equações (3.9) e (3.10), obtém-se

$$[u^\alpha + (1-u)^\alpha]^2 = \sum_{j=0}^{\infty} f_j u^j,$$

em que $f_j = (jt_0)^{-1} \sum_{m=1}^j [3m-j] t_m f_{j-m}$ (para $j \geq 1$) e $f_0 = t_0^2$.

Além disso, pode-se obter

$$\frac{(1-u)^{\alpha-1}}{[u^\alpha + (1-u)^\alpha]^2} = \frac{\sum_{j=0}^{\infty} m_j u^j}{\sum_{j=0}^{\infty} f_j u^j},$$

em que $m_j = (-1)^j \binom{\alpha-1}{j}$. Com base no quociente entre duas séries de potências, tem-se

$$\frac{u^{\alpha-1}(1-u)^{\alpha-1}}{[u^\alpha + (1-u)^\alpha]^2} = \sum_{j=0}^{\infty} q_j u^{\alpha+j-1}, \quad (3.12)$$

em que q_j (para cada j) pode ser determinado recursivamente como

$$q_j = \frac{1}{f_0} \left(m_j - \sum_{r=1}^j f_r q_{j-r} \right).$$

Portanto, de (3.12) obtém-se

$$E[W(Z)] = J(W(u)) = \alpha \sum_{j=0}^{\infty} q_j \int_0^1 u^{\alpha+j-1} W(\log[-\log(u)]) du, \quad (3.13)$$

em que a integral pode ser obtida para cada propriedade desejável de Z . A equação (3.13) é o principal resultado desta Subseção. A seguir, são apresentadas três propriedades gerais: momentos ordinários, incompletos e a fgm de Z .

O n -ésimo momento ordinário de Z (para $n = 1, 2, \dots$) é dado por

$$\mu'_n = E(Z^n) = J(\log[-\log(u)]) = \alpha \sum_{j=0}^{\infty} q_j \int_0^1 u^{\alpha+j-1} \{\log[-\log(u)]\}^n. \quad (3.14)$$

Expandindo $\{\log[-\log(u)]\}$ em séries de Taylor em $u = 2$, obtém-se usando Mathematica $\{\log[-\log(u)]\} = \sum_{k=0}^{\infty} p_k (u-2)^k$, em que $p_0 = -0,366513$, $p_1 = 1/\log(4)$, $p_2 = -[1 + \log(2)]/[8 \log^2(2)]$, $p_3 = [2 + 3 \log(2) + 2 \log^2(2)]/[48 \log^3(2)]$ e etc. Então, usando novamente (3.9) e (3.10), tem-se

$$\{\log[-\log(u)]\}^n = \left(\sum_{k=0}^{\infty} p_k (u-2)^k \right)^n = \sum_{k=0}^{\infty} v_{n,k} (u-2)^k,$$

em que $v_{n,0} = p_0^n$ e $v_{n,k} = (kp_0)^{-1} \sum_{m=1}^k [m(n+1) - k] p_m v_{n,k-m}$, para $n \geq 1$. Assim, combinando os últimos resultados e (3.14)

$$\mu'_n = \alpha \sum_{j,k=0}^{\infty} q_j v_{n,k} \int_0^1 u^{\alpha+j-1} (u-2)^k du. \quad (3.15)$$

A integral em (3.15) pode ser calculada no Maple em termos da função hiper-

geométrica generalizada como

$$\mu'_n = E(Z^n) = \alpha \sum_{j,k=0}^{\infty} \frac{(-1)^k 2^k q_j v_{n,k}}{(\alpha + j)} {}_2F_1(-k, \alpha + j; \alpha + j + 1; 1/2), \quad (3.16)$$

em que ${}_2F_1(p, q; r; y) = \sum_{j=0}^{\infty} (p)_j (q)_j y^j / [(r)_j j!]$ é a função hipergeométrica generalizada, $(p)_j$ é o símbolo de Pochhammer, definido por $(p)_j = p(p+1) \cdots (p+j-1) = \Gamma(p+j)/\Gamma(p) = (-1)^j \Gamma(1-p)/\Gamma(1-p-j)$ e $\Gamma(\cdot)$ é a função gama. Essa função ${}_2F_1(p, q; r; y)$ pode ser calculada no MATHEMATICA e MAPLE como `HypergeometricPFQ[{p,q},{r},y]` e `Hypergeometric([p,q],[r],y)`. Os momentos centrais (μ_s) e cumulantes (κ_s) de Z podem ser determinados de (3.16) como

$$\mu_s = \sum_{k=0}^p \binom{s}{k} (-1)^k \mu'_1{}^s \mu'_{s-k} \quad \text{e} \quad \kappa_s = \mu'_s - \sum_{k=1}^{s-1} \binom{s-1}{k-1} \kappa_k \mu'_{s-k},$$

respectivamente, em que $\kappa_1 = \mu'_1$. Portanto, $\kappa_2 = \mu'_2 - \mu_1'^2$, $\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3$, $\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu_2'^2 + 12\mu'_2\mu_1'^2 - 6\mu_1'^4$ e etc.

A assimetria $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ e curtose $\gamma_2 = \kappa_4/\kappa_2^2$ podem ser calculados a partir do terceiro e quarto cumulantes padronizados. Gráficos de assimetria e curtose, para algumas escolhas do parâmetro σ como função de α , são exibidos nas Figuras 3.3 (a) e 3.3 (b), respectivamente.

Figura 3.3 - Assimetria e curtose da distribuição LOLLW como uma função de α para valores selecionados de σ e para $\mu = 2$

O n -ésimo momento incompleto de Z , diga-se $m_n(z) = \int_0^q z^n \pi(z) dz$, é determinado de (3.15) como

$$m_n(z) = \alpha \sum_{j,k=0}^{\infty} q_j v_{n,k} \int_0^z u^{\alpha+j-1} (u-2)^k du.$$

Por expansão binomial, obtém-se (para $0 < z < 1$)

$$m_n(z) = \alpha \sum_{j,k=0}^{\infty} \sum_{r=0}^k \frac{(-2)^{k-r} q_j v_{n,k}}{(\alpha+j+r)} \binom{k}{r} z^{\alpha+j+r-1}. \quad (3.17)$$

Os momentos incompletos podem ser usados para calcular os desvios médios e as curvas de Bonferroni e Lorenz. A quantidade de dispersão em uma população é, evidentemente até certo ponto, medida pela totalidade dos desvios da média e mediana. Estes são conhecidos como desvios médios em torno da média e da mediana dados por $\delta_1 = \int_{-\infty}^{\infty} |z - \mu'_1| \pi(z) dz$ e $\delta_2 = \int_{-\infty}^{\infty} |z - M| \pi(z) dz$, respectivamente, em que $\mu'_1 = E(Z)$ é a média e $M = Q(0,5)$ é a mediana. Essas medidas podem ser determinadas de

$$\delta_1 = -2\mu'_1 S(\mu'_1) \quad \text{e} \quad \delta_2 = \mu'_1 - 2m_1(M),$$

respectivamente, em que $S(M)$ e $S(\mu'_1)$ vem de (3.5) (com $\mu = 0$ e $\sigma = 1$), a mediana M de Z é dada por $M = F^{-1}(1/2)$, em que $F(z) = 1 - S(z)$ e $m_1(q)$ é o primeiro momento incompleto de Z calculado de (3.17) com $n = 1$.

Os desvios médios podem ser usados para construir os gráficos das curvas de Lorenz e Bonferroni. Essas curvas são utilizadas em diversas áreas como confiabilidade, demografia, seguros e medicina. Para uma dada probabilidade π , elas são definidas por $L(\pi) = m_1(q)/\mu'_1$ e $B(\pi) = m_1(q)/(\pi \mu'_1)$, respectivamente, em que $q = F^{-1}(\pi)$ pode ser obtido invertendo a fda de Z numericamente.

Finalmente, a fgm de Z pode ser obtida a partir da equação (3.13) como

$$M(t) = E(e^{tZ}) = J(e^{tu}) = \alpha \sum_{j=0}^{\infty} q_j \int_0^1 u^{\alpha+j-1} [-\log(u)]^t du.$$

Se $p > -1$ e $v > -1$, a integral $H(p, v) = \int_0^1 u^p [-\log(u)]^v du$ é determinada usando o software *Mathematica* como $H(p, v) = \Gamma(v+1)/(1+p)^{v+1}$. Para $t > -1$, a última equação pode ser expressa como

$$M(t) = E(e^{tZ}) = \alpha \Gamma(t+1) \sum_{j=0}^{\infty} \frac{q_j}{(\alpha+j)^{t+1}}. \quad (3.18)$$

3.4 Modelo de regressão log-odd log-Logístico Weibull

Em muitas aplicações práticas, os tempos de vida são afetados por variáveis que são referidas como variáveis explanatórias ou covariáveis, tais como, o nível de colesterol, a pressão sanguínea, a idade e muitas outras. Assim, é importante explorar a relação entre o tempo de vida e variáveis explanatórias. Uma abordagem baseada em um modelo de regressão pode ser usada.

Seja $\mathbf{x} = (x_1, \dots, x_p)^T$ um vetor de variáveis explanatórias associadas à variável resposta Y . Com base na densidade LOLLW, pode-se construir um modelo de regressão linear (log-linear para T), ligando a variável resposta y_i e o vetor de variáveis explanatórias $\mathbf{x}_i$, como segue:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n. \quad (3.19)$$

O erro aleatório z_i tem densidade (3.6), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\sigma > 0$ e $\alpha > 0$ são os parâmetros desconhecidos do modelo e $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})^T$ é o vetor de variáveis explanatórias que modela o parâmetro de locação μ_i . Note que os parâmetros σ e μ controlam a escala e a forma da distribuição, respectivamente. Assim, o vetor de parâmetros de locação $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ do modelo de regressão LOLLW pode ser expresso como um modelo linear $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, em que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ é uma matriz conhecida do modelo.

Usando o modelo log-linear em (3.19), a função de sobrevivência de $Y_i|\mathbf{x}$ é dada por

$$S(y_i|\mathbf{x}) = 1 - \frac{\left\{1 - \exp \left[- \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}^\alpha}{\left\{1 - \exp \left[- \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}^\alpha + \left\{ \exp \left[- \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\}^\alpha}. \quad (3.20)$$

Para o modelo (3.19), tem-se os seguintes casos especiais:

- $\sigma = 1$, então o modelo (3.19) reduz-se ao modelo de regressão log-odd log-Logístico Exponencial;
- $\alpha = 1$, tem-se o modelo de regressão log-Weibull (LW);
- $\alpha = \sigma = 1$, obtém-se o modelo de regressão log-Exponencial.

3.4.1 Estimação por máxima verossimilhança

Considere uma amostra aleatória $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \dots, (y_n, \mathbf{x}_n)$ de n observações, em que $y_i = \min\{\log(T_i), \log(C_i)\}$ e $\mathbf{x}_i$ é o vetor de variáveis explanatórias, associado com o i -ésimo indivíduo e $\boldsymbol{\theta} = (\alpha, \sigma, \boldsymbol{\beta}^T)^T$ é o vetor de parâmetros. O logaritmo da função de verossimilhança assumindo censura à direita é dado por

$$l(\boldsymbol{\theta}) = \sum_{i \in F} l_1(\alpha, z_i) + \sum_{i \in C} l_2(\alpha, z_i), \quad (3.21)$$

em que

$$l_1(\alpha, z_i) = \log \left[\frac{\alpha \exp(z_i) \{\exp[-\exp(z_i)]\}^\alpha \{1 - \exp[-\exp(z_i)]\}^{\alpha-1}}{\sigma \{ \{1 - \exp[-\exp(z_i)]\}^\alpha + \{\exp[-\exp(z_i)]\}^\alpha \}^2} \right],$$

$$l_2(\alpha, z_i) = \log \left[1 - \frac{\{1 - \exp[-\exp(z_i)]\}^\alpha}{\{1 - \exp[-\exp(z_i)]\}^\alpha + \{\exp[-\exp(z_i)]\}^\alpha} \right],$$

$z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, F representa o conjunto de indivíduos não-censurados e C representa o conjunto de indivíduos com censura à direita.

O logaritmo da função de verossimilhança pode ser maximizado diretamente usando o SAS (proc NLMixed), o R (optim) ou a rotina MaxBFGS do programa de linguagem matricial Ox (ver DOORNIK, 2007). Os valores iniciais para σ e $\boldsymbol{\beta}$ podem ser tomados do ajuste do modelo de regressão LW com $\alpha = 1$. O ajuste do modelo de regressão LOLLW produz a função de sobrevivência estimada para y_i dada por

$$\hat{S}(y_i; \hat{\alpha}, \hat{\sigma}, \hat{\boldsymbol{\beta}}^T) = 1 - \frac{\{1 - \exp[-\exp(z_i)]\}^{\hat{\alpha}}}{\{1 - \exp[-\exp(z_i)]\}^{\hat{\alpha}} + \{\exp[-\exp(z_i)]\}^{\hat{\alpha}}}, \quad (3.22)$$

em que $\hat{z}_i = \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}$.

Para obtenção de intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo, necessita-se da matriz de informação observada $J = J(\boldsymbol{\theta})$ de ordem $(p+3) \times (p+3)$. Sob certas condições de regularidade para os parâmetros, a distribuição assintótica de $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ é $N_{p+3}(0, I(\boldsymbol{\theta})^{-1})$, em que $I(\boldsymbol{\theta})$ é a matriz de informação esperada. Na prática, a matriz de informação $I(\boldsymbol{\theta})$ pode ser substituída pela matriz de informação observada

avaliada em $\hat{\boldsymbol{\theta}}$ (diga-se $J(\hat{\boldsymbol{\theta}})$).

A matriz de informação observada pode ser expressa como

$$-\ddot{\mathbf{J}}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\mathbf{J}}_{\alpha\alpha} & \ddot{\mathbf{J}}_{\alpha\sigma} & \ddot{\mathbf{J}}_{\alpha\beta_j} \\ . & \ddot{\mathbf{J}}_{\sigma\sigma} & \ddot{\mathbf{J}}_{\sigma\beta_j} \\ . & . & \ddot{\mathbf{J}}_{\beta_j\beta_s} \end{pmatrix},$$

em que as submatrizes podem ser obtidas a partir do autor, a pedido. As regiões de confiança aproximadas para os parâmetros podem ser construídas baseadas na distribuição Normal multivariada $N_{p+3}(0, J(\hat{\boldsymbol{\theta}})^{-1})$. Além disso, a estatística da razão de verossimilhança (RV), pode ser usada para comparar o modelo LOLIW com seus submodelos. Para realizar essas comparações, pode-se calcular os valores máximos do logaritmo das verossimilhanças do modelo base e do modelo geral e construir a estatística de RV. Por exemplo, o teste de $H_0 : \alpha = 1$ *versus* $H : H_0$ não é verdadeira, é equivalente a comparar os modelos de regressão LOLIW e LW, e a estatística de RV reduz-se a

$$W = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\},$$

em que $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são os EMV sob H e H_0 , respectivamente. Para grandes amostras, W tem aproximadamente uma distribuição Qui-quadrado com 1 grau de liberdade.

Para interpretação dos coeficientes estimados, uma possível proposta é baseada na razão dos tempos medianos. Portanto, quando a variável explanatória é binária (0 ou 1) e considerando a razão dos tempos medianos com $x = 1$ no numerador, se $\hat{\beta}$ é negativo (positivo), isso implica que os indivíduos com $x = 1$ apresentam decréscimo (acrécimo) do tempo de sobrevivência mediano $[\exp(\hat{\beta}) \times 100\%]$ com relação aos indivíduos do grupo com $x = 0$, desconsiderando os efeitos das outras variáveis explanatórias. Esta interpretação pode ser estendida para variáveis contínuas ou categóricas.

3.4.2 Estudo de simulação

Um estudo de simulação de Monte Carlo foi conduzido para avaliar o comportamento sobre amostra finita dos EMV de α , σ e $\boldsymbol{\beta}$. Obteve-se resultados de 1.000 simulações de Monte Carlo, realizadas usando a linguagem de programação Ox (ver DOORNIK, 2007). Em cada réplica, uma amostra aleatória de tamanho n é obtida a partir da distribuição LOLIW($\alpha, \sigma, \boldsymbol{\beta}$) e os parâmetros são estimados pelo método da máxima verossimilhança. Os logaritmos dos tempos de vida, denotados por $\log(t_1), \dots, \log(t_n)$, foram gerados do modelo de regressão LOLIW (3.19), em que $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_i$, com x_i gerada de uma distribuição Uniforme no intervalo $[0, 1]$. Os tempos de censura $c_1, \dots, c_n$ foram gerados de uma distribuição Uniforme (n, τ) , em que τ indica as porcentagens de censura. Para cada

configuração de n , α , σ , β_0 , β_1 e porcentagens de censura, 1000 amostras foram geradas. Para cada conjunto de dados gerados, ajustou-se o modelo de regressão LOLIW (3.19), em que $\mu_i = \beta_0 + \beta_1 x_i$.

No presente trabalho, realizou-se um estudo de simulação considerando $n = 100, 200$ e 500 , $\alpha = \sigma = 0, 5$, $\beta_0 = 2$, $\beta_1 = 3$ e porcentagens de censura iguais a 0%, 10% e 30%.

A partir da Tabela 3.1, observa-se que o aumento das porcentagens de censura provoca o aumento dos erros quadrático médio. Também, os erros quadrático médio decrescem à medida que o tamanho da amostra aumenta, como esperado. Percebe-se que há um viés amostral pequeno na estimação dos parâmetros dos modelos de regressão LOLIW.

Tabela 3.1 - Médias e Erros Quadrático Médio (E.Q.M.) dos EMV dos parâmetros da distribuição LOLIW

n	Parâmetro	0%	10%	30%
100	α	0,6237(0,0265)	0,6294(0,0288)	0,6353(0,0357)
	σ	0,5937(0,0161)	0,5972(0,0175)	0,5991(0,0232)
	β_0	2,0080(0,0144)	2,0044(0,0164)	2,0001(0,0237)
	β_1	2,9948(0,0305)	2,9993(0,0392)	3,0040(0,0611)
200	α	0,5836(0,0109)	0,5870(0,0121)	0,5978(0,0166)
	σ	0,5668(0,0071)	0,5693(0,0078)	0,5787(0,0112)
	β_0	1,9989(0,0065)	2,0009(0,0006)	2,0030(0,0086)
	β_1	2,9993(0,0124)	2,9987(0,0125)	2,9931(0,0213)
500	α	0,5528(0,0040)	0,5566(0,0048)	0,5629(0,0057)
	σ	0,5443(0,0028)	0,5474(0,0032)	0,5535(0,0042)
	β_0	2,0015(0,0016)	1,9999(0,0018)	1,9999(0,0024)
	β_1	2,9966(0,0026)	2,9966(0,0032)	2,9943(0,0050)

3.4.3 Probabilidade de cobertura

Para verificação das probabilidades de cobertura foram construídos intervalos de confiança para os parâmetros. O estudo de simulação é realizado para $n = 100, 200$ e 300 com $\alpha = 1, 2$, $\sigma = 1, 5$, $\beta_0 = 2$, $\beta_1 = 3$ e porcentagens de censura 0%, 10% e 30%. Foram determinados 3.000 intervalos de confiança, com um nível de 95%, baseados nas estimativas de máxima verossimilhança obtidas para cada configuração de n e porcentagens de observações censuradas. Ainda, calculou-se quantos intervalos continham os verdadeiros valores dos parâmetros para cada configuração especificada. As porcentagens de cobertura são dadas por: $((3.000 - \text{n}^\circ \text{ de intervalos que não contem o verdadeiro valor}) / 3.000) \times 100$.

do parâmetro)/3.000) $\times$ 100%.

Para cada configuração apresentadas na Tabela 3.2, conclui-se que:

- Se consideradas observações não censuradas, as probabilidades de coberturas tendem a aumentar à medida que o tamanho amostral aumenta, como esperado;
- Se a porcentagem de observações censuradas mudam, observa-se que há um pequeno efeito sobre as probabilidades de cobertura.

Tabela 3.2 - Porcentagens de cobertura dos EMVs dos parâmetros do modelo de regressão LOLIW

n	Parâmetro	0%	10%	30%
100	α	95,87%	95,47%	94,13%
	σ	94,83%	94,43%	93,17%
	β_0	95,53%	95,20%	95,50%
	β_1	93,77%	94,50%	93,67%
200	α	95,63%	96,83%	96,70%
	σ	95,57%	97,20%	96,20%
	β_0	95,50%	95,77%	95,73%
	β_1	95,43%	94,97%	95,20%
300	α	95,53%	97,77%	97,27%
	σ	95,43%	97,47%	96,90%
	β_0	95,23%	95,90%	96,03%
	β_1	95,33%	95,83%	95,20%

3.5 Estudo de má-especificação do modelo

Para avaliar o comportamento dos EMV dos parâmetros do modelo de regressão LOLIW quando este é mal especificado, foi realizado um estudo de simulação de Monte Carlo considerando 1.000 replicações e usando a linguagem de programação Ox (DOORNIK, 2007). O logaritmo dos tempos de vida foram gerados de um modelo de regressão Lognormal com parâmetros selecionados σ e $\beta = (\beta_0, \beta_1)^T$ e a covariável x_i gerada de uma distribuição Uniforme no intervalo $(0, 1)$. Os tempos censurados foram gerados de uma distribuição Uniforme (n, τ) , em que τ denota a porcentagem de censura. Ainda, foi considerada a configuração com tamanho amostral $n = 30$, $\sigma = 0,5$, $\beta_0 = 2$, $\beta_1 = 3$ e porcentagens de censura 0%, 10% e 30% para gerar as amostras. Para cada conjunto gerado, foi ajustado o modelo de regressão LOLIW. Os resultados deste estudo são mostrados na Tabela 3.3, em que é possível observar que o aumento na porcentagem de censura,

em geral, implica no aumento do Erro Quadrático Médio (E.Q.M.). Há um pequeno viés amostral na estimação dos parâmetros deste modelo de regressão. Portanto, o modelo de regressão LOLW fornece EMV consistentes, mesmo quando os dados são gerados de um modelo diferente.

Tabela 3.3 - Estimativas médias e Erros Quadrático Médio (E.Q.M.) dos EMV dos parâmetros do modelo de regressão LOLW

n	Parâmetro	0%	10%	30%
30	α	1,5419(-)	1,5120(-)	1,3986(-)
	σ	0,6096(0,0704)	0,5888(0,0681)	0,5449(0,0615)
	β_0	2,2707(0,1201)	2,2710(0,1251)	2,2609(0,1355)
	β_1	3,0323(0,1284)	3,0240(0,1472)	3,0169(0,2005)

3.6 Influência local

A realização de diagnósticos de influência é um importante passo na análise de dados, uma vez que podem fornecer uma indicação de mau ajuste do modelo ou observações influentes. Como os modelos de regressão são sensíveis aos pressupostos do modelo subjacente, geralmente realizar uma análise de sensibilidade é fortemente aconselhável. Cook(1986) usa essa ideia para motivar a sua avaliação da análise de influência. Ele sugere que mais confiança pode ser depositada em um modelo, o qual é relativamente estável a pequenas modificações. Outra abordagem sugerida por Cook (1986), é ponderar as observações, em vez de removê-las. O cálculo de influências locais pode ser realizado para o modelo (3.19). Se a verossimilhança deslocada $LD(\omega) = 2\{l(\hat{\theta}) - l(\hat{\theta}_\omega)\}$ é usada, em que $\hat{\theta}_\omega$ denota a EMV sob o modelo perturbado, então a curvatura normal para θ na direção $\mathbf{d}$, $\|\mathbf{d}\| = 1$, é dada por $C_{\mathbf{d}}(\theta) = 2|\mathbf{d}^T \Delta^T [\ddot{\mathbf{L}}(\theta)]^{-1} \Delta \mathbf{d}|$, em que Δ é uma matriz $(p+3) \times n$ que depende do esquema de perturbação. Os elementos deste esquema são dados por $\Delta_{vi} = \partial^2 l(\theta|\omega) / \partial \theta_v \partial \omega_i$, $i = 1, 2, \dots, n$ e $v = 1, 2, \dots, p+3$ avaliados em $\hat{\theta}$ e ω_0 ; ω_0 é o vetor não-perturbado. Para o modelo de regressão LOLW, os elementos de $\ddot{\mathbf{L}}(\theta)$ podem ser obtidos a partir do autor, mediante solicitação. As curvaturas normais $C_{\mathbf{d}}(\lambda)$, $C_{\mathbf{d}}(\sigma)$ e $C_{\mathbf{d}}(\beta)$ também podem ser calculadas para realizar vários gráficos de índices, incluindo por exemplo, o gráfico de índices $\mathbf{d}_{max}$, o autovetor correspondente a $C_{\mathbf{d}_{max}}$, o maior autovalor da matriz $\mathbf{B} = -\Delta^T [\ddot{\mathbf{L}}(\theta)]^{-1} \Delta$, e o gráfico de índices de $C_{\mathbf{d}_i}(\lambda)$, $C_{\mathbf{d}_i}(\sigma)$ e $C_{\mathbf{d}_i}(\beta)$, o que juntos são denotados por influência local total. Ver, por exemplo, Lesaffre e Verbeke (1998), em que $\mathbf{d}_i$ denota um vetor de zeros $n \times 1$ com 1 na i -ésima posição. Portanto, a curvatura na direção $\mathbf{d}_i$ toma a forma $C_i = 2|\Delta_i^T [\ddot{\mathbf{L}}(\theta)]^{-1} \Delta_i|$, em que Δ_i^T denota a i -ésima linha de Δ . É comum apontar casos para os quais $C_i \geq 2\bar{C}$, em que $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$.

Em seguida, para três regimes de perturbação, foi calculada a seguinte matriz

$$\Delta = (\Delta_{vi})_{(p+3) \times n} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_v} \right)_{(p+3) \times n}, \quad v = 1, \dots, p+3 \quad \text{e} \quad i = 1, \dots, n.$$

Para os passos seguintes foram considerados o modelo (3.19), o logaritmo da função de verossimilhança dada por (3.21) e o vetor de pesos $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$.

• Perturbação de casos

Neste caso, o logaritmo da função de verossimilhança tem a forma

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

em que $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ e $l_1(\cdot)$ e $l_2(\cdot)$ são definidos em (3.21). Ainda, $\Delta = (\Delta_\alpha^T, \Delta_\sigma^T, \Delta_\beta^T)^T$ pode ser calculado numericamente.

• Perturbação da variável resposta

Agora, considera-se que cada y_i é perturbado como $y_{iw} = y_i + \omega_i S_y$, em que S_y é um fator de escala que pode ser estimado pelo desvio padrão da variável resposta observada y e $\omega_i \in \mathbb{R}$. O logaritmo da função de verossimilhança pode ser expresso por

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

em que $z_i^* = [(y_i + \omega_i S_y) - \mathbf{x}_i^T \boldsymbol{\beta}] / \sigma$, $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$ e $l_1(\cdot)$ e $l_2(\cdot)$ são definidas em (3.21). A matriz $\Delta = (\Delta_\alpha^T, \Delta_\sigma^T, \Delta_\beta^T)^T$ é obtida numericamente.

• Perturbação de uma variável explanatória

Considere agora uma perturbação aditiva sobre uma variável explanatória contínua particular, diga-se X_q , definindo $x_{iq\omega} = x_{iq} + \omega_i S_q$, em que S_q é um fator de escala e $\omega_i \in \mathbb{R}$. O logaritmo da função de verossimilhança perturbada tem a forma

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

em que $z_i^{**} = (y_i - \mathbf{x}_i^{*T} \boldsymbol{\beta}) / \sigma$, $\mathbf{x}_i^{*T} \boldsymbol{\beta} = \beta_1 + \beta_2 x_{i2} + \dots + \beta_q (x_{iq} + \omega_i S_q) + \dots + \beta_p x_{ip}$, $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$ e $l_1(\cdot)$ e $l_2(\cdot)$ são definidos em (3.21). A matriz $\Delta = (\Delta_\alpha^T, \Delta_\sigma^T, \Delta_\beta^T)^T$ pode ser obtida numericamente.

Trabalhos anteriores, sobre curvaturas de influência local em modelos de regressão

para dados censurados, podem ser encontrados em Escobar e Meeker (1992), Ortega, Bolfarine e Paula (2003), Ortega, Cancho e Paula (2009), Ortega, Cordeiro e Hashimoto (2011), Silva et al. (2008) e Hashimoto et al. (2010).

Em ordem, para determinar se a i -ésima observação é possivelmente influente, Labra et al. (2012), com base no trabalho de Poon e Poon (1999), assumiram a i -ésima observação como possivelmente influente se $M(0)_i = \frac{Cd_i(\theta)}{tr(2\mathbf{B})}$ for maior que $\overline{M(0)} + c^*SM(0)$, em que $\overline{M(0)} = 1/q$, $SM(0)$ é o erro padrão amostral $\{M(0)_k, k = 1, \dots, q\}$ e c^* é uma constante qualquer selecionada de acordo com a aplicação real.

3.7 Análise de resíduos

O objetivo da análise de resíduos é avaliar a adequacidade do modelo para um conjunto de dados particular. Este propósito inclui a verificação de características intrínsecas dos dados. Em ordem para checar algumas destas características, por exemplo, outliers, muitas abordagens têm sido propostas por Cox e Snell (1968), Cook e Weisberg (1982), Collet (2003) e Ortega, Paula e Bolfarine (2008).

De acordo com Cox e Snell (1968), as técnicas mais formais da análise de resíduos apresentam testes de significância sensíveis, que indicarão a natureza de um afastamento do modelo inicial, mas não explicitamente como estender ou substituir o modelo.

3.7.1 Resíduos martingale

Os resíduos martingale usados em processos de contagem (FLEMING; HARRINGTON, 1991) são dados por $r_{M_i} = \delta_i + \log[S(y_i; \hat{\theta})]$, em que δ_i é o indicador de censura, ou seja, $\delta_i = 0$ ($\delta_i = 1$) denota uma observação censurada (não-censurada) e $S(y_i; \hat{\theta})$ denota a função de sobrevivência estimada apresentada em (3.22). Assim, o resíduo martingale para o modelo de regressão LOLLW, considerando censura à direita, é dado por

$$r_{M_i} = \begin{cases} 1 + \log \left\{ 1 - \frac{[1 - \exp[-\exp(z_i)]]^{\hat{\alpha}}}{[1 - \exp[-\exp(z_i)]]^{\hat{\alpha}} + [\exp[-\exp(z_i)]]^{\hat{\alpha}}} \right\} & \text{se } \delta_i = 1, \\ \log \left\{ 1 - \frac{[1 - \exp[-\exp(z_i)]]^{\hat{\alpha}}}{[1 - \exp[-\exp(z_i)]]^{\hat{\alpha}} + [\exp[-\exp(z_i)]]^{\hat{\alpha}}} \right\} & \text{se } \delta_i = 0, \end{cases} \quad (3.23)$$

em que $z_i = \frac{\log(t_i) - \mathbf{x}_i^T \hat{\beta}}{\hat{\sigma}}$.

3.7.2 Resíduo deviance modificado

O resíduo deviance modificado é um refinamento do resíduo martingale para torná-lo mais simétrico em torno do zero, o que facilita em geral, a detecção de pontos atípicos (outliers) (THERNEAU; GRAMBSCH; FLEMING, 1990; COLLETT, 2003).

O resíduo deviance modificado para modelos de regressão paramétrica é dado por

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

O resíduo deviance modificado aplicado ao modelo de regressão LOLLW, considerando censura à direita, pode ser expresso por

$$r_{D_i} = \begin{cases} \text{sinal}(\hat{r}_{M_i}) \left\{ -2 \left\{ 1 + \log \left[1 - \frac{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}}{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}} + [\exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}} \right] + \right. \\ \left. \log \left\{ -\log \left[1 - \frac{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}}{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}} + [\exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}} \right] \right\} \right\}^{1/2} & \text{if } \delta_i = 1, \\ \text{sinal}(\hat{r}_{M_i}) \left\{ -2 \left\{ \log \left[1 - \frac{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}}{[1 - \exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}} + [\exp[-\exp(\hat{z}_i)]]^{\hat{\alpha}}} \right] \right\} \right\}^{1/2} & \text{if } \delta_i = 0, \end{cases}$$

em que $\hat{r}_{M_i}$ é o resíduo martingale dado por (3.23) e $\hat{z}_i = \frac{\log(t_i) - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}$.

Se o modelo é apropriado, os resíduos martingale e deviance modificado, devem apresentar um comportamento aleatório em torno do zero. Os gráficos dos resíduos, martingale ou deviance modificado *versus* os valores ajustados, fornecem uma maneira para verificar a adequacidade do modelo e auxiliam na detecção de observações atípicas.

3.8 Estudos de simulação

Para investigar a distribuição empírica dos resíduos r_{D_i} para valores de $n = 100, 200$ e 500 , $\alpha = 0, 5$ e $1, 2$, $\sigma = 0, 5$, $\beta_0 = 2$, $\beta_1 = 4$ e 3 , e porcentagens de censura 0% , 10% e 30% , foi realizado um estudo de simulação descrito a seguir. O logaritmo dos tempos de vida, denotados por $\log(t_1), \dots, \log(t_n)$ foram gerados do modelo de regressão LOLLW (3.19), em que $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_i$ e x_i foi gerada a partir de uma distribuição Uniforme no intervalo $(0, 1)$. Os tempos de censura $c_1, \dots, c_n$, foram gerados de uma distribuição Uniforme (n, τ) , em que τ denota as porcentagens de censura. Para cada configuração

de n , α , σ , β_0 , β_1 e porcentagens de censura, 1.000 amostras foram geradas. Para cada conjunto de dados gerados, ajustou-se o modelo de regressão LOLLW (3.19), em que $\mu_i = \beta_0 + \beta_1 x_i$, e calculou-se os resíduos r_{D_i} . Posteriormente, os resíduos r_{D_i} foram ordenados e apresentados *versus* os quantis esperados da distribuição Normal padrão em um gráfico. Estes gráficos podem ser usados para verificar a normalidade assintótica dos resíduos.

As Figuras 3.4, 3.5, 3.7 e 3.8 levam às seguintes conclusões: a distribuição empírica dos resíduos rM_i é assimétrica em torno do zero e apresenta em geral acentuada curtose, a distribuição empírica dos resíduos rD_i apresenta uma boa concordância com a distribuição Normal padrão e finalmente, quando a porcentagem de censura diminui ou o tamanho da amostra aumenta, a distribuição empírica dos resíduos rD_i apresenta melhor aproximação da distribuição Normal padrão, como esperado em ambas as situações. Portanto, pode-se usar o gráfico Normal de probabilidade para os resíduos rD_i com envelope simulado para o modelo, como sugerido por Atkinson (1985).

As Figuras 3.6 e 3.9 exibem os histogramas dos resíduos deviance modificado, em que, à medida que aumenta a contaminação da amostra pela censura, começa a surgir uma suave assimetria à direita.

Figura 3.4 - Gráficos Normais de probabilidade para os resíduos r_{M_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura = 0%, 10% e 30% ($\alpha = 0,5$, $\sigma = 0,5$, $\beta_0 = 2$, $\beta_1 = 4$)

Figura 3.5 - Gráficos Normais de probabilidade para os resíduos r_{D_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura = 0%, 10% e 30% ($\alpha = 0,5$, $\sigma = 0,5$, $\beta_0 = 2$, $\beta_1 = 4$)

Figura 3.6 - Histogramas para os resíduos r_{D_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura = 0%, 10% e 30% ($\alpha = 0,5$, $\sigma = 0,5$, $\beta_0 = 2$, $\beta_1 = 4$)

Figura 3.7 - Gráficos Normais de probabilidade para os resíduos r_{M_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura=0%, 10% e 30% ($\alpha = 1, 2, \sigma = 0, 5, \beta_0 = 2, \beta_1 = 3$)

Figura 3.8 - Gráficos Normais de probabilidade para os resíduos r_{D_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura = 0%, 10% e 30% ($\alpha = 1, 2$, $\sigma = 0, 5$, $\beta_0 = 2$, $\beta_1 = 3$)

Figura 3.9 - Histogramas para os resíduos r_{D_i} . Tamanho de amostra $n = 100$, $n = 200$ e $n = 500$, porcentagens de censura = 0%, 10% e 30% ($\alpha = 1, 2, \sigma = 0, 5, \beta_0 = 2, \beta_1 = 3$)

3.9 Aplicação

Aqui, foi usado um conjunto de dados analisado por Kalbfleisch e Prentice (1980) para estudar o modelo de regressão LOLLW com censura à direita. Os parâmetros do modelo foram estimados pelo método da máxima verossimilhança com a subrotina `optim` do R.

Primeiro, é descrito o conjunto de dados. Então, são fornecidos os EMV (e os correspondentes erros-padrão em parênteses) e os valores das estatísticas AIC (Akaike Information Criterion), CAIC (Consistent Akaike Information Criterion) e BIC (Bayesian Information Criterion). Quanto menor os valores destes critérios melhor o ajuste.

3.9.1 Transplante de coração em Stanford

Estes dados estão disponíveis no livro de Kalbfleisch e Prentice (1980, apêndice I, pgs. 230-232). Dentre os 103 pacientes, 69 receberam transplante de coração. O número de mortes registradas foi de 75 e os 28 pacientes restantes contribuíram com tempos de sobrevivência censurados. As variáveis consideradas foram t_i = tempos de sobrevivência (tempo desde aceitação para o programa de transplantação para receber um transplante de coração até a morte) da entrada do paciente no estudo (em dias) e um vetor de variáveis explanatórias, incluindo x_{i1} = idade de aceitação, x_{i2} = cirurgia prévia (cirurgia: código como 1 = sim e 0 = não) e x_{i3} = transplante (código como 1 = sim; 0 = não), para $i = 1, \dots, 103$.

Em seguida, são fornecidos os resultados da análise exploratória desses dados. As porcentagens de observações censuradas para os tempos de sobrevivência é de 27,18%. Primeiro, considerou-se os tempos de vida e a censura para uma análise preliminar. As estimativas dos parâmetros α , σ e μ foram obtidas por máxima verossimilhança. Utilizou-se os EMV de μ e σ para a distribuição log-Weibull como valores iniciais para o procedimento iterativo. Os cálculos foram realizados usando o procedimento NLMixed do SAS. A Tabela 3.4 lista os EMV e os correspondentes erros-padrão (E.P.) dos parâmetros do modelo e os valores das estatísticas AIC, BIC e CAIC. Os resultados indicam que o modelo LOLLW tem os menores valores de AIC, BIC e CAIC e, portanto, pode ser escolhido como o melhor modelo. Além disso, calculou-se os máximos dos logaritmos das funções de verossimilhança do modelo base e do modelo geral e a estatística de RV para testar alguns submodelos. Por exemplo, a estatística de RV para testar as hipóteses $H_0: \alpha = 1$ versus $H_1: H_0$ não é verdadeira, ou seja, para comparar os modelos LOLLW e log-Weibull, é igual a $W = 2\{-194,10 - (-198,05)\} = 7,90$ (valor- $p = 0,0049$), o qual dá indicação favorável ao modelo LOLLW.

Tabela 3.4 - EMVs dos parâmetros do modelo para os dados de transplante de coração, os correspondentes E.P. (dados em parênteses) e as estatísticas AIC, CAIC e BIC

Modelo	μ	σ	α	AIC	CAIC	BIC
LOLLW	19,5669 (23,1601)	39,9024 (63,2153)	21,3898 (33,7730)	394,2	394,1	402,1
log-Weibull	5,8922 (0,2193)	1,8995 (0,1792)	1 (-)	400,1	400,3	405,4

A Figura 3.10 exibem os gráficos das funções de sobrevivência empírica e estimadas das distribuições LOLW e log-Weibull. Obteve-se um bom ajuste do modelo LOLW para esses dados.

Figura 3.10 - Funções de sobrevivência estimadas para as distribuições LOLW e log-Weibull e função de sobrevivência empírica para os dados de transplante de coração

O modelo de regressão LOLW ajustado é dado por

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma z_i,$$

em que os erros $z_1, \dots, z_{103}$ são variáveis aleatórias independentes com função de densidade

(3.6).

Os EMVs e as estatísticas AIC, BIC e CAIC para os modelos de regressão LOLIW e log-Weibull são listados na Tabela 3.5. Observa-se que o modelo de regressão LOLIW, quando comparado com o modelo de regressão log-Weibull, produz menores valores das estatísticas AIC, BIC e CAIC.

A estatística de RV para testar a hipótese $H_0: \alpha = 1$ versus $H_1: H_0$ não é verdadeira, ou seja, para comparar os modelos de regressão LOLIW e log-Weibull, é $w = 2\{-167, 79 - (-171, 71)\} = 7,84$ (valor- $p=0,0051$), o que indica melhor ajuste para o modelo de regressão LOLIW.

Cox (1972) propôs um modelo de regressão muito útil para analisar dados de sobrevivência, em que a variável aleatória de interesse representa o tempo de falha e estes tempos são assumidos identicamente distribuídos em alguma distribuição especificada. Ainda, notou que se a pressuposição de riscos proporcionais é verificada, sendo possível estimar o efeito do(s) parâmetro(s) sem nenhuma consideração da função de risco (abordagem não-paramétrica). Esta abordagem para dados de sobrevivência é chamada modelos de riscos proporcionais.

O modelo de Cox pode ser utilizado se uma razão existe para assumir que a função de risco de base tem uma forma paramétrica. Neste caso, a função de risco de base pode ser substituída por uma densidade paramétrica. Os EMV $\hat{\beta}$ podem ser obtidos maximizando a função de verossimilhança usando a linguagem de programação SAS.

A Tabela 3.6 fornece as estimativas, erros-padrão correspondentes e valores- p para o ajuste do modelo de regressão de Cox. Duas variáveis explanatórias são significativas ao nível de 5% de significância, o qual é o mesmo resultado para o modelo de regressão LOLIW. Ainda, as Tabelas 3.5 e 3.6 listam os valores do AIC, CAIC e BIC para comparar os modelos de regressão LOLIW, log-Weibull e de riscos proporcionais de Cox. O modelo de regressão LOLIW apresenta menores valores destes critérios com relação aos demais modelos, o que indica que pode ser usado efetivamente na análise destes dados. Portanto, o modelo proposto é uma grande alternativa na modelagem de dados de sobrevivência.

Tabela 3.5 - EMVs, Erros-padrão (E.P.), valor- p e as estatísticas AIC, CAIC e BIC para os modelos de regressão log-Weibull e LOLIW ajustados aos dados de transplante cardíaco

θ	Log-Weibull			LOLIW		
	Estimativa	E.P.	Valor- p	Estimativa	E.P.	Valor- p
α	-	-	-	4,6283	3,5307	-
σ	1,4658	0,1315	-	6,2032	4,6851	-
β_0	7,9742	0,9340	$< 0,001$	8,7448	1,7603	$< 0,001$
β_1	-0,0924	0,0206	$< 0,001$	-0,0769	0,0199	$< 0,001$
β_2	1,2143	0,6470	0,063	1,4055	0,5745	0,016
β_3	2,5375	0,3734	$< 0,001$	2,5920	0,3884	$< 0,001$
	AIC	CAIC	BIC	AIC	CAIC	BIC
	353,4208	354,0393	366,5944	347,5957	348,4707	363,4041

Tabela 3.6 - Estimativas para o modelo de regressão de Cox ajustado aos dados de transplante cardíaco

Parâmetro	Estimativa	E.P.	Valor- p	I.C. 95%
β_1	0,0589	0,0149	$< 0,0001$	(0,0208, 0,0792)
β_2	-0,7421	0,4422	0,0933	(-1,6088, 0,1246)
β_3	-1,6528	0,2761	$< 0,0001$	(-2,1939, -1,1116)
	AIC	CAIC	BIC	
	555.20	555.80	563.10	

3.9.2 Análises de sensibilidade e residual

Na tentativa de detectar possíveis observações influentes no modelo de regressão ajustado aos dados presentes, calculou-se influência local usando a linguagem matricial Ox versão 6.01. Na Figura 3.11, são apresentados os gráficos de influência local ($M(0)$ *versus* índices), considerando esquemas de perturbação da covariável idade ($C_{dmax} = 1,7664$), do logaritmo do tempo ($C_{dmax} = 2,3595$) e a perturbação ponderação de casos (perturbação da verossimilhança) ($C_{dmax} = 2,1443$). Portanto, com base nos gráficos, identificou-se os pontos #29, #67, #73, #83, #93, #94, e #99 como possíveis observações influentes. Para detectar possíveis outliers no modelo de regressão LOLIW, os gráficos dos resíduos martingale *versus* valores ajustados e dos resíduos deviance modificado (r_{D_i}) *versus* valores ajustados foram construídos e exibidos na Figuras 3.12 (a) e 3.12 (b). Pela

análise do gráfico dos resíduos deviance modificado observou-se, em uma análise de sensibilidade, que a observação #99 aparece como possível outlier. Desta forma, a análise de sensibilidade (local) e residual detectaram como possíveis observações influentes, as observações #67 e #99. Elas são as que aparecem mais frequentemente. O próximo passo a ser realizado é verificar se estas observações têm algum impacto sobre as estimativas dos parâmetros. O modelo foi então reajustado eliminando cada observação individualmente, e posteriormente, as duas.

Figura 3.11 - Gráfico de $M(0)$ *versus* índices para os parâmetros do modelo de regressão LOLLW sob os esquemas de perturbação: (a) da covariável idade, (b) da variável resposta e (c) da verossimilhança (ponderação de casos) nos dados “Stanford”

Figura 3.12 - (a) Gráfico dos resíduos martingale *versus* valores ajustados e (b) Gráfico dos resíduos deviance modificado *versus* valores ajustados

Na Tabela 3.7, são apresentadas as mudanças relativas (em parênteses) de cada parâmetro estimado, definida por $\mathbf{RC}_{\theta_j} = \left[(\hat{\theta}_j - \hat{\theta}_{j(i)}) / \hat{\theta}_j \right] \times 100$, em que $\hat{\theta}_{j(i)}$ é o EMV sem a i -ésima observação. Nesta Tabela, verifica-se a robustez dos parâmetros do modelo de regressão LOLW, uma vez que as covariáveis x_1 (idade), x_2 (cirurgia : 1 = sim, 0 = não) e x_3 (transplante: 1 = sim, 0 = não) permaneceram significativas no modelo.

Portanto, não houveram mudanças inferencias após a remoção das observações consideradas como possíveis influentes nos gráficos de diagnóstico.

A Figura 3.13 exibe o gráfico Normal de probabilidade para os resíduos deviance modificado rD_i com envelope simulado (ATKINSON, 1985), que foi usado para verificar a qualidade do ajuste do modelo de regressão LOLW. Como não houveram mudanças na inferência, mesmo após a eliminação dos possíveis pontos influentes, o gráfico Normal de probabilidade com envelope simulado foi construído considerando todo o conjunto de dados. Observe que todos os pontos estão dentro do envelope, o que traz evidências de um bom ajuste do modelo de regressão LOLW para os dados apresentados.

Tabela 3.7 - Mudanças relativas [$\mathbf{RC}$ - em %], estimativas dos parâmetros e valor- p em parênteses quando as observações #67, #99 e (#67 e #99) são eliminadas, respectivamente

Exclusão	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Nenhuma	-	-	-	-	-	-
	4,6283	6,2032	8,7448	-0,0769	1,4055	2,5929
	(-)	(-)	(< 0,001)	(< 0,001)	0,016	(< 0,001)
#67	[21,9433]	[21,2439]	[6,3249]	[3,1209]	[-2,7677]	[1,2191]
	3,6127	4,8854	8,1917	-0,0745	1,4444	2,5604
	(-)	(-)	(< 0,001)	(< 0,001)	(0,0155)	(< 0,001)
#99	[24,7348]	[26,9683]	[14,6304]	[16,9051]	[-0,5051]	[-2,1026]
	3,4835	4,5303	7,4654	-0,0639	1,4126	2,6465
	(-)	(-)	(< 0,001)	(< 0,001)	(0,0140)	(< 0,001)
#67 e #99	[16,9025]	[18,8354]	[14,2027]	[20,6762]	[-0,8894]	[-2,6543]
	3,8460	5,0348	7,5028	-0,0610	1,4180	2,6608
	(-)	(-)	(< 0,001)	(< 0,001)	0,0141	(< 0,001)

Figura 3.13 - Gráfico Normal de probabilidade com envelope simulado

3.9.3 Modelo final

Com base em toda análise realizada, conclui-se que o modelo de regressão LOLLW mais apropriado para ajustar estes dados é

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma z_i, \quad i = 1, \dots, 103.$$

Os EMVs dos parâmetros no modelo final são dados na Tabela 3.8. Pode-se interpretar as estimativas do modelo final da seguinte forma: o tempo de sobrevida mediano deve decrescer aproximadamente 92% ($e^{-0.0769} \times 100\%$) quando a idade de aceitação aumenta uma unidade.

Tabela 3.8 - EMVs dos parâmetros do modelo de regressão LOLLW ajustado aos dados de transplante de coração - modelo final

Parâmetro	Estimativa	E.P.	Valor- p	I.C. 95%
α	4,6283	3,5307	-	(0,0000, 11,5483)
σ	6,2033	4,6851	-	(0,0000, 15,3859)
β_0	8,7449	1,7603	<0,001	(5,2948, 12,1949)
β_1	-0,0769	0,0199	<0,001	(-0,1159, -0,0379)
β_2	1,4055	0,5745	0,0016	(0,2794, 2,5315)
β_3	2,5920	0,3884	<0,001	(1,8307, 3,3532)

Os gráficos das funções de sobrevivência correspondentes as variáveis explanatórias para o modelo de regressão LOLLW são apresentadas na Figura 3.14.

A partir da Figura 3.14 (a), verifica-se que $\hat{S}(4 | \text{idade} = 8) = 0,8631$, significando que cerca de 86% dos pacientes com idade igual a 8 anos estarão vivos em $y = 4$ (≈ 55 dias). Por outro lado, para pacientes com idades de 44 e 64 anos $\hat{S}(4 | \text{idade} = 44) = 0,2649$ e $\hat{S}(4 | \text{idade} = 64) = 0,0620$, ou seja, as porcentagens de pacientes que estarão vivos em $y = 4$ (≈ 55 dias) caem para 26% e 6%, respectivamente. Estes resultados indicam que a sobrevivência estimada de pacientes diminui, à medida que a idade aumenta. Analogamente, as Figuras 3.14 (b) e 3.14 (c) indicam que pacientes sem cirurgia prévia ou que não receberam transplante, aproximadamente 62% estarão vivos em $y = 6$ (≈ 403 dias). Em contrapartida, cerca de 86% dos pacientes que passaram por uma cirurgia prévia, estarão vivos em $y = 6$ (≈ 403 dias), enquanto que para os pacientes que receberam transplantes, esta porcentagem de sobrevivência em $y = 6$ aumenta para 95%. Portanto, pode-se afirmar que o paciente que se submete a cirurgia aumenta o tempo de sobrevivência, assim como receber um transplante de coração.

Figura 3.14 - Funções de sobrevivência ajustadas usando o modelo de regressão LOLLW para os dados de transplante cardíaco. Figuras (a) para $x_1 =$ idade de aceitação, (b) para $x_2 =$ cirurgia prévia e (c) para $x_3 =$ transplante

3.10 Conclusões

Neste trabalho, foi proposto um modelo de regressão log-odd log-Logístico Weibull para dados censurados à direita. Algumas propriedades matemáticas da distribuição log-odd log-Logística Weibull foram estudadas. As estimativas dos parâmetros foram obtidas

pelo método da máxima verossimilhança. Desenvolveu-se vários estudos de simulação para mostrar que a distribuição empírica dos resíduos deviance modificado se aproxima da distribuição Normal padrão. Aplicou-se procedimentos de análise de sensibilidade a um conjunto de dados da área médica, com o objetivo de verificar a sensibilidade do modelo quando sujeito a esquemas de perturbações. Essa análise detectou algumas possíveis observações influentes e, usando o caso de deleção, foi observado que as significâncias das variáveis explanatórias não se alteraram. Para averiguar a qualidade do ajuste, construiu-se o gráfico Normal de probabilidade com envelope simulado, de onde observou-se que todos os pontos estavam dentro do envelope, indicando que o modelo de regressão LOLLW se ajusta bem aos dados de transplante de coração.

Referências

- ATKINSON, A.C. **Plots, transformations, and regression**. Oxford: University Press, 1985. 282p.
- COLLETT, D. **Modelling survival data in medical research**. 2nd ed. London: Chapman & Hall, 2003. 391p.
- COOK, R.D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman and Hall, 1982. 230p.
- COOK, R.D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.
- COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society - Series B**, Oxford, v.30, p.248-275, 1968.
- DOORNIK, J. Ox: **Object-oriented matrix programming using Ox**. London: Timberlake Consultants Ltd, 2007. 461p.
- ESCOBAR, L.A.; MEEKER, W.Q. Assessing influence in regression analysis with censored data. **Biometrics**, Washington, v.48, p.507-528, 1992.
- EUGENE, N; LEE, C; FAMOYE, F. Beta-Normal distribution and its applications. **Communications in Statistics - Theory and Methods**, London, v.31, p.497-512, 2002
- FACHINI, J.B.; ORTEGA; E.M.M.; CORDEIRO, G.M. A bivariate regression model with cure fraction. **Journal of Statistical Computation and Simulation**, New York, v.84, p.1580-1595, 2014.

FLEMING, T.T.; HARRINGTON, D.P. **Counting process and survival analysis**. New York: Wiley, 1991. 448p.

HASHIMOTO, E.M.; ORTEGA, E.M.M.; CANCHO, V.G.; CORDEIRO, G.M. The log-exponentiated Weibull regression model for interval-censored data. **Computational Statistics and Data Analysis**, Amsterdam, v.54, p.1017-1035, 2010.

KALBFLEISH, J.D.; PRENTICE, R.L. **The statistical analysis of failure time data**. New York: John Wiley, 2002. 439p.

LAWLESS, J.F. **Statistical models and methods for lifetime data**. 2nd ed. New Jersey: John Wiley & Sons, 2003. 664p.

LABRA, F.V.; GARAY, A.M.; LACHOS, V.H.; ORTEGA, E.M.M. Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-Normal distributions. **Journal of Statistical Planning and Inference**, Amsterdam, v.142, p.2149-2165, 2012.

LEMONTE, A.J.; PATRIOTA, A. Influence diagnostics in Birnbaum-Saunders nonlinear regression models. **Journal of Applied Statistics**, Abingdon, v.38, p.871-884, 2011.

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

MATOS, L.A.; LACHOS, V.H.; BALAKRISHNAN, N.; LABRA, F.V. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. **Computational Statistics and Data Analysis**, Amsterdam, v.57, p.450-464, 2013.

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.; PAULA, G.A. Generalized log-Gamma regression models with cure fraction. **Lifetime Data Analysis**, New York, v.15, p.79-106, 2009.

ORTEGA, E.M.M.; CORDEIRO, G.M.; HASHIMOTO, E.M. A log-linear regression model for the Beta-Weibull distribution. **Communications in Statistics - Simulation and Computation**, London, v.40, p.1206-1235, 2011.

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

PAULA, G.A. (2013). On diagnostics in double generalized linear models. **Computational Statistics and Data Analysis**, Amsterdam, v.68, p.44-51, 2013.

POON, W.; 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.

RONDON, L.M.; PENAGOS, L.H.V.; CYSNEIROS, F.J.A. Diagnostic procedures in Birnbaum-Saunders nonlinear regression models. **Computational Statistics and Data Analysis**, Amsterdam, v.56, p.1662-1680, 2012.

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

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

4 O MODELO DE REGRESSÃO LOG-ODD LOGÍSTICA WEIBULL COM CENSURA INFORMATIVA

Resumo

Uma das principais características dos dados em análise de sobrevivência, deve-se ao fato de que a variável de interesse nem sempre pode ser observada, o que faz com que algumas observações sejam censuradas. As metodologias usuais assumem que estas observações censuradas não carregam informações a respeito da distribuição que a variável resposta assume (censura não-informativa), ou seja, admite-se que uma determinada observação é censurada simplesmente pelo fato do evento de interesse, falha ou morte, não ter ocorrido durante o período de acompanhamento do estudo. Contudo, em muitas situações, o tempo de sobrevivência claramente é perturbado pelo mecanismo de censura e, portanto, este efeito produzido deve ser incorporado na análise. A questão é que, uma vez assumida censura informativa como não-informativa, os resultados da análise podem mascarar vieses, comprometendo o poder de predição do modelo. Assim, este trabalho tem por objetivo considerar o mecanismo de censura informativo no modelo de regressão log-odd-Logístico Weibull, baseado na metodologia descrita em Huang e Wolfe (2002), para analisar as variações que ocorrem nas estimativas dos parâmetros quando considerado o mecanismo informativo.

Palavras-chave: Distribuição Gama; Distribuição log-odd log-Logística weibull; Fragilidade; Mecanismo de censura informativa

Abstract

One of the main characteristics of data from survival analysis is that the variable of interest is not always observed, so that some observations are censored. The usual methods assume that these observations do not carry information on the distribution of the response variable (non-informative censoring). In other words, it is considered that an observation is censored simply by the fact that the event of interest (failure or death) did not occur during the period of study. However, in many situations, the survival time is clearly perturbed by the censoring mechanism, so the effect produced must be included in the analysis. The question is that once informative censoring is assumed to be non-informative, the results of the analysis can mask biases and thus weakening the model's predictive power. Therefore, one of the objectives of this paper is to consider the informative censoring mechanism in the log-odd-Logistic Weibull regression model, based on the method described in Huang and Wolfe (2002), to analyze the variations that occur in estimating the model parameters. We obtain maximum likelihood estimates for the model parameters by considering censored data and evaluate local influence on the estimates of the parameters by taking different perturbation schemes. In addition, we define martingale and deviance residuals to detect outliers and evaluate the model assumptions. We demonstrate that our extended regression model is very useful to the analysis of real data and may give more realistic fits than other special regression models.

Keywords: Gamma distribution; Log-odd log-logistic Weibull distribution; Fragility; Informative censoring mechanism

4.1 Introdução

De modo geral, os dados em análise de sobrevivência diferem de outros tipos de dados encontrados em problemas de cunho estatístico, no que se refere, em particular, ao fato de que a variável de interesse nem sempre pode ser observada. Uma das principais características em dados dessa natureza é a presença da censura. A censura pode então, ser entendida como uma observação parcial da variável resposta, sendo que muitas causas podem caracterizá-la, a citar: término do estudo sem ocorrência do evento, morte do indivíduo por uma causa diferente da estudada, abandono do paciente ao tratamento, etc.

Nos modelos de análise de sobrevivência, geralmente, é assumido independência entre os tempos de falha e censura (censura não-informativa). Contudo, há muitas outras causas e tipos de censura do que apenas o término pré-determinado de um período de observação (censura à direita). Em ensaios clínicos e outros estudos médicos e biológicos, por exemplo, várias razões podem caracterizar a censura como suscetível a ser informativa. De acordo com Lagakos (1979), estas são:

- Quando os indivíduos saem de um ensaio clínico por razões que podem estar relacionadas com a terapia;
- Quando os indivíduos são removidos de um ensaio clínico pelo delineamento e não mais acompanhados para tempo de sobrevivência se já experienciaram um evento crítico específico;
- Quando os indivíduos em um estudo experienciam uma falha de uma causa de interesse secundário o qual censura o tempo de falha da causa de interesse primário.

Em todos estes casos, torna-se necessário verificar se estas observações parciais trazem informações sobre a distribuição dos tempos de falha. Como relatado em Lagakos (1979), situações de não adaptação do indivíduo a determinado tratamento, pondo fim ao período de acompanhamento pode ser considerado como um caso de censura informativa, visto que o motivo da censura está relacionado com o tempo de sobrevivência posterior do indivíduo.

O problema é que uma vez assumido, erroneamente, censura não-informativa em um conjunto de dados, sendo ela informativa, os resultados da análise podem ser viesados e os métodos usuais irão superestimar (subestimar) a função de sobrevivência (STAPLIN; KIMBER; COLLETT; RODERICK, 2014). Ainda, a magnitude do viés tenderá a aumentar à medida que o número de censurados aumenta.

Desta forma, alguns autores formularam propostas para modelagem de dados de sobrevivência sob a pressuposição de censura informativa. Lagakos e Williams (1978) propuseram um modelo que apresenta um parâmetro escalar θ que diz respeito ao mecanismo de censura oculto. Para $\theta = 1$, a função de verossimilhança, depende somente da distribuição de sobrevivência. Para $0 \leq \theta < 1$, os efeitos da censura tornam-se cada vez mais importantes na estimação dos parâmetros do modelo.

Wang, Qin e Chiang (2001) modelaram a ocorrência de eventos recorrentes com uso de um processo de Poisson não-estacionário e por meio de uma variável latente, considerando os parâmetros da distribuição da censura (informativa) e das variáveis latentes como parâmetros *nuisance*. Em seguida, foi considerado um modelo de intensidade multiplicativa para estimação não-paramétrica da função de risco.

Rotnitzky et al. (2007) obtiveram estimadores da função de sobrevivência na presença de mecanismos de censura à direita. Neste mesmo trabalho, foi realizada uma análise de sensibilidade sobre o estimador que representa a censura potencialmente informativa, assumindo que depois do ajuste para todos os fatores de prognóstico, o tempo de falha e censura eram independentes. Nesta mesma linha de raciocínio, Scharfstein e Robins (2002) apresentaram um método que além do ajuste de modelos considerando censura informativa, permite quantificar simultaneamente a sensibilidade da inferência para dependência residual entre a falha e censura devido a fatores não controlados. Contudo, este modelo de Scharfstein e Robins (2002) não permite mais de um mecanismo de censura, o que significa que toda censura em um conjunto de dados deve ser considerada como informativa. Desta forma, o modelo de Rotnitzky et al. (2007) é uma extensão deste modelo por permitir múltiplas causas de censura.

Em todos esses trabalhos a maior dificuldade é identificar nos dados o mecanismo de censura dependente (TSIATIS, 1975). Nesta linha, vários autores propuseram uma análise de sensibilidade a fim de avaliar os efeitos da censura dependente sobre as estimativas dos parâmetros da distribuição dos tempos de falha (SIANNIS, 2004; SIANNIS; COPAS; LU, 2005; ZHANG; HEITJAN, 2006; HUANG; ZHANG, 2008). Na abordagem de Siannis (2004) e Siannis, Copas e Lu (2005) foi utilizada uma estrutura de riscos proporcionais, permitindo a estimação das mudanças individuais das estimativas dos parâmetros, bem como, um preditor linear. Uma abordagem mais abrangente de análise de sensibilidade para modelos na presença de censura informativa foi proposta por Siannis (2011). Esta nova abordagem faz uso dos modelos de riscos proporcionais de Cox, a qual é mais flexível que os modelos de sobrevivência paramétricos padrões.

No contexto de riscos competitivos, Lu e Tsiatis (2011) utilizaram covariáveis auxiliares para obter estimadores que quantificam a censura informativa e são mais eficientes que aqueles que desconsideram essas covariáveis.

Freitas e Rodrigues (2013) consideraram o modelo de taxa de cura padrão exponencial com censura informativa e investigaram, por meio de um estudo de simulação, o impacto causado pela censura informativa nas probabilidades de cobertura e nos comprimentos de intervalos de confiança assintóticos dos parâmetros de interesse.

Os objetivos deste trabalho são ajustar modelos paramétricos a dados reais de sobrevivência, sob a suposição de que os tempos de falha e censura são condicionalmente independentes dada uma fragilidade (mecanismo de censura informativo) e analisar as variações nas estimativas dos parâmetros, quando consideradas as metodologias usual e censura informativa.

Para avaliar a adequacidade do modelo, foram desenvolvidos estudos de diagnóstico para detectar possíveis observações influentes ou extremas que podem causar distorções nos resultados da análise. Finalmente, foram comparados dois resíduos para avaliar desvios das pressuposições do erro, bem como detectar possíveis outliers no modelo de regressão log-odd log-Logística Weibull na presença de censura informativa.

O trabalho está organizado como segue: na Seção 4.2, é definido o novo modelo de regressão log-odd log-Logístico Weibull. Na Seção 4.3, é estudado o mecanismo de censura informativa no modelo de regressão locação-escala. Na Seção 4.4, são adotadas algumas medidas de diagnósticos sobre três esquemas de perturbações no modelo de regressão proposto com censura informativa. Na Seção 4.5, são definidos dois tipos de resíduos dos modelos ajustados para avaliar características pressupostas da distribuição do erro e para detectar outliers. Na Seção 4.6, um conjunto de dados reais é analisado, o qual mostra a flexibilidade, relevância prática e aplicabilidade do modelo de regressão proposto. Finalmente, algumas conclusões são apresentadas na Seção 4.7.

4.2 Modelo de regressão log-odd log-Logístico Weibull

Muitas generalizações da distribuição Weibull têm sido propostas na literatura relacionada a área de confiabilidade, para fornecerem melhores ajustes a certos conjuntos de dados quando comparados com os modelos Weibull com dois ou três parâmetros. Ver, por exemplo, as distribuições listadas e discutidas nas Tabelas I e II dadas em Pham e Lai (2007). Mais recentemente, Cruz et al. (2014) introduziram uma extensão do modelo Weibull com três parâmetros, chamada distribuição odd log-Logística Weibull(OLLW), que tem a função densidade de probabilidade (fdp) dada por

$$f(t; \alpha, \gamma, \lambda) = \frac{\alpha \gamma t^{\gamma-1} \left\{ \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha \left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^{\alpha-1}}{\lambda \gamma \left[\left\{ 1 - \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha + \left\{ \exp \left[- \left(\frac{t}{\lambda} \right)^\gamma \right] \right\}^\alpha \right]^2}, \quad t > 0, \quad (4.1)$$

em que $\lambda > 0$ é um parâmetro de escala e $\alpha > 0$ e $\gamma > 0$ são parâmetros de forma.

Doravante, T é denotada como uma variável aleatória com fdp (4.1). A função de sobrevivência de T é dada por

$$S(t; \alpha, \gamma, \lambda) = 1 - \frac{\{1 - \exp[-(\frac{t}{\lambda})^\gamma]\}^\alpha}{\{1 - \exp[-(\frac{t}{\lambda})^\gamma]\}^\alpha + \{\exp[-(\frac{t}{\lambda})^\gamma]\}^\alpha}.$$

Então, a função de risco ou função taxa de falha (ftf) de T vem a ser $h(t; \alpha, \gamma, \lambda) = f(t; \alpha, \gamma, \lambda)/S(t; \alpha, \gamma, \lambda)$. A grande flexibilidade deste modelo, para ajustar dados de tempo de vida, está nas diferentes formas que a ftf pode tomar: (i) se $\alpha = 1$, a ftf toma a forma da função de risco da distribuição Weibull; (ii) para alguns valores de α no intervalo $(0, 1)$ e certos valores de λ e γ , a ftf pode tomar a forma de banheira; (iii) para alguns valores de $\alpha > 1$ e combinações especiais de λ e γ , a ftf é unimodal.

Seja $Y = \log(T)$ uma variável aleatória com distribuição log-odd log-Logística Weibull (LOLLW). Recentemente, Cruz, Ortega e Cordeiro (2015) definiram o modelo de regressão LOLW por

$$y_i = \log(t_i) = \mu_i + \sigma z_i, \quad i = 1, \dots, n, \quad (4.2)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\sigma > 0$ e $\lambda > 0$ são os parâmetros desconhecidos e $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ é o vetor de variáveis explanatórias que modela o preditor linear $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. O vetor de preditores lineares considerados $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ do modelo de regressão LOLW é simplesmente $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, em que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ é uma matriz conhecida do modelo. O modelo (4.2) é também referido como um modelo log-locação-escala ou modelo de tempo de falha acelerado.

Além disso, após a transformação adequada, definimos a variável aleatória padronizada $Z = (Y - \mu)/\sigma$. A função densidade de probabilidade para a resposta Y_i pode ser expressa como

$$f(y_i|\mathbf{x}) = \frac{\alpha \exp(z_i) \{\exp[-\exp(z_i)]\}^\alpha \{1 - \exp[-\exp(z_i)]\}^{\alpha-1}}{\sigma [\{1 - \exp[-\exp(z_i)]\}^\alpha + \{\exp[-\exp(z_i)]\}^\alpha]^2}, \quad (4.3)$$

em que $z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$, $y_i \in \mathbb{R}$, $\alpha > 0$ e $\sigma > 0$.

A função de sobrevivência, ftf e ftf acumulada de Y_i são dadas por

$$S(y_i|\mathbf{x}) = 1 - \frac{\{1 - \exp[-\exp(z_i)]\}^\alpha}{\{1 - \exp[-\exp(z_i)]\}^\alpha + \{\exp[-\exp(z_i)]\}^\alpha},$$

$$h(y_i|\mathbf{x}) = \frac{\alpha \exp(z_i) \{1 - \exp[-\exp(z_i)]\}^{\alpha-1}}{\sigma \{ \{1 - \exp[-\exp(z_i)]\}^\alpha + \{\exp[-\exp(z_i)]\}^\alpha \}}$$

e

$$H(y_i|\mathbf{x}) = \alpha \exp(z_i) + \log \left\{ \{1 - \exp[-\exp(z_i)]\}^\alpha \{\exp[-\exp(z_i)]\}^\alpha \right\},$$

respectivamente.

4.3 Censura informativa no modelo de regressão

Nesta seção é descrita a construção da função de verossimilhança usual considerando censura informativa. Então, é detalhada a função de verossimilhança marginal considerando que os tempos censurados carregam informações a respeito dos tempos de falha (censura informativa ou dependente).

Para construir a função de verossimilhança marginal é utilizado o método descrito por Huang e Wolfe (2002) recorrendo a f.d.p. da Gama para construir a distribuição de fragilidade e ainda supondo que a mesma atua de forma multiplicativa (SANTOS JR., 2012).

Sob o mecanismo de censura à direita, considere T uma variável aleatória representando o tempo de falha de uma observação (tempo até a ocorrência de um evento de interesse), e C , outra variável aleatória, a qual representa o tempo de censura associado a esta observação. Para $i = 1, \dots, n$, os dados observados consistirão de pares (y_i, δ_i) , em que $y_i = \min(T_i^*, C_i^*)$ é o logaritmo dos tempos observados para a i -ésima observação ($T_i^* = \log(T_i)$, $C_i^* = \log(C_i)$) e δ_i é a variável indicadora de falha, ou seja, $\delta_i = 1$ se $T_i^* \leq C_i^*$ ou $\delta_i = 0$ se $T_i^* > C_i^*$. Também, as funções de densidade e de sobrevivência para T_i e C_i são denotadas por $f_{T^*}(y_i, \boldsymbol{\kappa})$, $S_{T^*}(y_i, \boldsymbol{\kappa})$, $f_{C^*}(c_i^*, \boldsymbol{\nu})$ e $S_{C^*}(c_i^*, \boldsymbol{\nu})$, respectivamente, em que $\boldsymbol{\kappa}$ é o vetor de parâmetros associados com a distribuição dos tempos de falha e $\boldsymbol{\nu}$ é o vetor de parâmetros associados com a censura.

Se uma associação existe entre os tempos de falha e censura, os métodos usuais (censura não-informativa) utilizados para analisar dados censurados não são robustos, causando a necessidade de formular uma estrutura capaz de incorporar aquela dependência. Para este propósito, baseado no trabalho de Huang e Wolfe (2002), assumiu-se que as variáveis aleatórias T^* e C^* tornam-se independentes quando individualmente são condicionadas a uma fragilidade W (efeito aleatório). Portanto, como apresentado em Santos Jr. (2012), considerou-se que $W \sim \text{Gama}(\phi, \phi)$. Embora outras distribuições possam ser usadas, tal como as distribuições Uniforme, Weibull e log-Normal (VAUPEL; YASHIN, 1983), a distribuição Gama como fragilidade em modelos semiparamétricos é mais ampla-

mente utilizada, basicamente devido a conveniência algébrica. A escolha da distribuição de fragilidade com média e variância iguais a ϕ é para evitar problemas de identificabilidade do modelo.

Considerando o modelo no qual a fragilidade atua de forma multiplicativa, a função de risco condicional para o logaritmo dos tempos de falha é definida por

$$h_{T^*}(y|\boldsymbol{\kappa}, W) = h_0^{(T^*)}(y|\boldsymbol{\kappa}, W), \quad (4.4)$$

em que $h_0^{(T^*)}(y|\boldsymbol{\kappa})$ é a função de risco base para o logaritmo dos tempos de falha. A função de sobrevivência condicional é dada por

$$S_{T^*}(y|\boldsymbol{\kappa}, W) = \exp(-WH_0^{(T^*)}(y|\boldsymbol{\kappa})), \quad (4.5)$$

em que $H_0^{(T^*)}(y|\boldsymbol{\kappa})$ é a função de risco base acumulada para T^* . Do mesmo modo, as funções de risco e sobrevivência podem ser definidos para os tempos de censura. Assim, a função de risco condicional para os tempos de censura podem ser expressas como

$$h_{C^*}(y|\boldsymbol{\nu}, W) = h_0^{(C^*)}(y|\boldsymbol{\nu})W, \quad (4.6)$$

em que $h_0^{(C^*)}(y|\boldsymbol{\nu})$ é a função de risco base para o logaritmo dos tempos censurados. A função de sobrevivência condicional torna-se

$$S_{C^*}(y|\boldsymbol{\nu}, W) = \exp(-WH_0^{(C^*)}(y|\boldsymbol{\nu})), \quad (4.7)$$

em que $H_0^{(C^*)}(y|\boldsymbol{\nu})$ função de risco base acumulada para C^* .

Também, a função de sobrevivência marginal pode ser expressa como

$$S(y|\phi) = \int_W S(y, W)dW = \int_W S(y|W)f(W)dW.$$

Então, considerando que $W \sim \text{Gama}(\phi, \phi)$ e $S(y|W) = \exp(-WH_0(y))$, tem-se

$$S(y|\phi) = \left(\frac{\phi}{\phi + H_0(y)} \right)^\phi. \quad (4.8)$$

Desta forma, quando considerados os logaritmos dos tempos de falha ou censura, a função de sobrevivência marginal mudará somente com respeito a função de risco acumulada, a qual está associada com a distribuição considerada para os tempos em questão. Os estimadores de máxima verossimilhança (EMVs), sob a pressuposição que T^* e C^* são condicionalmente independentes dadas a fragilidade W , são obtidos a partir da verossi-

milhança marginal. Assumindo a fragilidade W na distribuição conjunta de T^* e C^* , a função de verossimilhança marginal é dada por

$$\begin{aligned}
L(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi | D) &= \prod_{i=1}^n \int_{W_i} \{P[T_i^* = y_i, C_i^* > y_i, W_i]\}^{\delta_i} \{P[C_i^* = y_i, T_i^* > y_i, W_i]\}^{1-\delta_i} dW_i \\
&= \prod_{i=1}^n \int_{W_i} \{P[(T_i^* = y_i, C_i^* > y_i) | W_i]\}^{\delta_i} \{f(W_i)\}^{\delta_i} \{P[C_i^* = y_i, T_i^* > y_i | W_i]\}^{1-\delta_i} \{f(W_i)\}^{1-\delta_i} dW_i \\
&= \prod_{i=1}^n \int_{W_i} \{P[(T_i^* = y_i, C_i^* > y_i) | W_i]\}^{\delta_i} \{f(W_i)\}^{\delta_i} \{P[C_i^* = y_i, T_i^* > y_i | W_i]\}^{1-\delta_i} f(W_i) dW_i.
\end{aligned} \tag{4.9}$$

Assumindo que T^* e C^* são condicionalmente independentes dada a fragilidade W , a função de verossimilhança marginal (4.9) pode ser reescrita como

$$\begin{aligned}
L(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi | D) &= \prod_{i=1}^n \int_{W_i} \{P[T_i^* = y_i | W_i]\}^{\delta_i} \{P[C_i^* > y_i | W_i]\}^{\delta_i} \\
&\quad \{P[C_i^* = y_i | W_i]\}^{1-\delta_i} \{P[T_i^* > y_i | W_i]\}^{1-\delta_i} f(W_i) dW_i.
\end{aligned} \tag{4.10}$$

Substituindo as funções (4.4)-(4.7) na equação (4.10) e considerando as relações

$$f_{T^*}(y_i | \boldsymbol{\kappa}, W_i) = h_{T^*}(y_i | \boldsymbol{\kappa}, W_i) S_{T^*}(y_i | \boldsymbol{\kappa}, W_i) \text{ e } f_{C^*}(y_i | \boldsymbol{\nu}, W_i) = h_{C^*}(y_i | \boldsymbol{\nu}, W_i) S_{C^*}(y_i | \boldsymbol{\nu}, W_i),$$

a função de verossimilhança reduz-se a

$$L(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi | D) = \prod_{i=1}^n \int_{W_i} \{h_{T^*}(y_i | \boldsymbol{\kappa}, W_i)\}^{\delta_i} S_{T^*}(y_i | \boldsymbol{\kappa}, W_i) h_{C^*}(y_i | \boldsymbol{\nu}, W_i)^{1-\delta_i} S_{C^*}(y_i | \boldsymbol{\nu}, W_i) f(W_i) dW_i.$$

Substituindo a função de densidade da fragilidade $W \sim Gama(\phi, \phi)$, obtém-se

$$\begin{aligned}
L(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi | D) &= \prod_{i=1}^n \int_{W_i} \{h_{T^*}(y_i | \boldsymbol{\kappa}, W_i)\}^{\delta_i} S_{T^*}(y_i | \boldsymbol{\kappa}, W_i) h_{C^*}(y_i | \boldsymbol{\nu}, W_i)^{1-\delta_i} S_{C^*}(y_i | \boldsymbol{\nu}, W_i) \\
&\quad \times \frac{\phi}{\Gamma(\phi)} W_i^{\phi-1} \exp(-\phi W_i) dW_i \\
&= \prod_{i=1}^n h_0^{(T^*)}(y_i | \boldsymbol{\kappa})^{\delta_i} h_0^{(C^*)}(y_i | \boldsymbol{\nu})^{1-\delta_i} \int_{W_i} W_i^{\phi} \\
&\quad \exp\{-W_i[\phi + H_0^{(T^*)}(y_i | \boldsymbol{\kappa}) + H_0^{(C^*)}(y_i | \boldsymbol{\nu})]\} dW_i.
\end{aligned}$$

O núcleo de uma distribuição Gama aparece na integral, ou seja,

$$W_i \sim \text{Gama} \left(\phi + 1; \phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right).$$

Assim, a função de verossimilhança reduz-se a

$$L(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi|D) = \prod_{i=1}^n \left\{ \frac{h_0^{(T^*)}(y_i|\boldsymbol{\kappa})^{\delta_i} h_0^{(C^*)}(y_i|\boldsymbol{\nu})^{1-\delta_i} \phi^\phi \Gamma(\phi + 1)}{\Gamma(\phi) [\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{(C^*)}(y_i|\boldsymbol{\nu})]^{\phi+1}} \right\}. \quad (4.11)$$

Nota-se que a função de verossimilhança marginal (4.11) terá forma fechada quando a função de risco base $h_0(y)$ e a função de risco acumulada $H_0(y)$ tiverem formas fechadas. Desta forma, o logaritmo da função de verossimilhança tem a seguinte forma

$$\begin{aligned} l(\boldsymbol{\kappa}, \boldsymbol{\nu}, \phi|D) &= \sum_{i=1}^n \left\{ \delta_i \log[h_0^{(T^*)}(y_i|\boldsymbol{\kappa})] + (1 - \delta_i) \log[h_0^{(C^*)}(y_i|\boldsymbol{\nu})] + \phi \log(\phi) + \log[\Gamma(\phi + 1)] \right\} \\ &\quad - \log[\Gamma(\phi)] - (\phi + 1) \log \left[\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right]. \end{aligned}$$

Considere uma amostra $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ de n observações independentes, em que a resposta aleatória é definida por $y_i = \min\{T_i^*, C_i^*\}$. Considere ainda censura informativa e a distribuição LOLLW para o logaritmo dos tempos de falha e a distribuição log-Weibull para o logaritmo dos tempos censurados definidas na Seção 4.2. O logaritmo da função de verossimilhança para o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\kappa}^\top, \boldsymbol{\nu}^\top, \phi)^\top$, em que $\boldsymbol{\kappa} = (\alpha^{(T^*)}, \sigma^{(T^*)}, \boldsymbol{\beta}^{(T^*)})^\top$, $\boldsymbol{\nu} = (\alpha^{(C^*)}, \sigma^{(C^*)}, \mu^{(C^*)})^\top$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, é obtido dos modelos (4.4) e (4.6) como

$$\begin{aligned} l(\boldsymbol{\theta}|D) &= r\phi \log(\phi) + r \log[\Gamma(\phi + 1)] - r \log[\Gamma(\phi)] + \sum_{i=1}^n \delta_i \log \left[h_0^{(T^*)}(y_i|\boldsymbol{\kappa}) \right] \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \log \left[h_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right] \\ &\quad - (\phi + 1) \sum_{i=1}^n \log \left[\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right], \end{aligned} \quad (4.12)$$

em que r é o número de observações não censuradas (falhas),

$$h_0^{(T^*)}(y_i|\boldsymbol{\kappa}) = \frac{\alpha^{(T^*)} \exp(z_i^{(T^*)}) (1 - u_i^{(T^*)})^{\alpha^{(T^*)}-1}}{\sigma^{(T^*)} \left\{ (1 - u_i^{(T^*)})^{\alpha^{(T^*)}} + [u_i^{(T^*)}]^{\alpha^{(T^*)}} \right\}},$$

$$h_0^{(C^*)}(y_i|\boldsymbol{\nu}) = \frac{\alpha^{(C^*)} \exp(z_i^{(C^*)})(1 - u_i^{(C^*)})^{\alpha^{(C^*)}-1}}{\sigma^{(C^*)} \left\{ (1 - u_i^{(C^*)})^{\alpha^{(C^*)}} + [u_i^{(C^*)}]^{\alpha^{(C^*)}} \right\}},$$

$$H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) = \log \left\{ (1 - u_i^{(T^*)})^{\alpha^{(T^*)}} + [u_i^{(T^*)}]^{\alpha^{(T^*)}} \right\} - \alpha^{(T^*)} \log(u_i^{(T^*)}),$$

$$H_0^{(C^*)}(y_i|\boldsymbol{\nu}) = \log \left\{ (1 - u_i^{(C^*)})^{\alpha^{(C^*)}} + [u_i^{(C^*)}]^{\alpha^{(C^*)}} \right\} - \alpha^{(C^*)} \log(u_i^{(C^*)}),$$

$$u_i^{(T^*)} = \exp\{-\exp[z_i^{(T^*)}]\}, \quad u_i^{(C^*)} = \exp\{-\exp[z_i^{(C^*)}]\},$$

$$z_i^{(T^*)} = \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(T^*)}}{\sigma^{(T^*)}} \quad \text{e} \quad z_i^{(C^*)} = \frac{y_i - \mu^{(C^*)}}{\sigma^{(C^*)}}.$$

Note que a estrutura do modelo de regressão foi considerada somente nos log-tempos de falha, o qual é o interesse. Pesquisas futuras podem ser desenvolvidas para colocar a estrutura do modelo de regressão, simultaneamente no log-tempos de falha e log-censura.

O logaritmo da função de verossimilhança pode ser maximizada, quer diretamente usando o SAS (proc NLMixed), o R (optim) ou a rotina MaxBFGS no programa de linguagem matricial Ox (DOORNIK, 2007), ou resolvendo as equações não lineares de verossimilhança obtidas derivando (4.12). Valores iniciais $\sigma^{(T^*)}$, $\boldsymbol{\beta}^{(T^*)}$, $\sigma^{(C^*)}$, $\mu^{(C^*)}$ e ϕ podem ser tomados do ajuste do modelo de regressão log-Weibull com censura informativa para $\alpha^{(T^*)} = \alpha^{(C^*)} = 1$. Os componentes do vetor escore $U(\boldsymbol{\theta})$ são dados no Apêndice B.

Seja $\widehat{\boldsymbol{\theta}}$ EMV de $\boldsymbol{\theta}$. Os intervalos de confiança aproximados e testes de hipóteses sobre os parâmetros do modelo necessita da matriz de informação observada total $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$ $(p+6) \times (p+6)$. Sob condições que são cumpridas para os parâmetros no interior do espaço paramétrico mas não sobre a borda, a distribuição assintótica de $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ é $N_{p+6}(0, I(\boldsymbol{\theta})^{-1})$, em que $I(\boldsymbol{\theta})$ é a matriz de informação esperada. Na prática, substitui $I(\boldsymbol{\theta})$ por $-\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})$, ou seja, a matriz de informação observada avaliada em $\widehat{\boldsymbol{\theta}}$. A matriz de informação observada é expressa como

$$-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\mathbf{L}}_{\alpha^{(T^*)}\alpha^{(T^*)}} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\sigma^{(T^*)}} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\alpha^{(C^*)}} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\sigma^{(C^*)}} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\phi} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\alpha^{(T^*)}\mu^{(C^*)}} \\ . & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\sigma^{(T^*)}} & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\alpha^{(C^*)}} & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\sigma^{(C^*)}} & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\phi} & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\sigma^{(T^*)}\mu^{(C^*)}} \\ . & . & \ddot{\mathbf{L}}_{\alpha^{(C^*)}\alpha^{(C^*)}} & \ddot{\mathbf{L}}_{\alpha^{(C^*)}\sigma^{(C^*)}} & \ddot{\mathbf{L}}_{\alpha^{(C^*)}\phi} & \ddot{\mathbf{L}}_{\alpha^{(C^*)}\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\alpha^{(C^*)}\mu^{(C^*)}} \\ . & . & . & \ddot{\mathbf{L}}_{\sigma^{(C^*)}\sigma^{(C^*)}} & \ddot{\mathbf{L}}_{\sigma^{(C^*)}\phi} & \ddot{\mathbf{L}}_{\sigma^{(C^*)}\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\sigma^{(C^*)}\mu^{(C^*)}} \\ . & . & . & . & \ddot{\mathbf{L}}_{\phi\phi} & \ddot{\mathbf{L}}_{\phi\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\phi\mu^{(C^*)}} \\ . & . & . & . & . & \ddot{\mathbf{L}}_{\sigma\beta_j^{(T^*)}} & \ddot{\mathbf{L}}_{\mu^{(C^*)}\mu^{(C^*)}} \end{pmatrix},$$

em que as sub-matrizes podem ser obtidas do autor mediante solicitação. Os intervalos de confiança aproximados para os parâmetros. Foram construídos baseados na distribuição Normal multivariada $N_{p+6}(0, -\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}})^{-1})$. Também, a estatística da razão de verossimilhança (RV) pode ser usada para comparar o modelo de regressão LOLIW com censura informativa e alguns de seus submodelos. Pode-se calcular os valores máximos dos logaritmos das verossimilhanças do modelo base e do modelo geral para construir a estatística RV para testar algumas sub-modelos do modelo de regressão LOLIW com censura informativa.

Por exemplo, o teste de $H_0 : \alpha^{(T^*)} = \alpha^{(C^*)} = 1$ *versus* $H : H_0$ não é verdadeira é equivalente a comparar os modelos de regressão LOLIW e log-Weibull com censura informativa e a estatística RV é dada por $W = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\}$, em que $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são os EMVs sob H e H_0 , respectivamente. Para grandes amostras, W tem aproximadamente uma distribuição Qui-quadrado com 1 grau de liberdade.

4.4 Análise de sensibilidade

Após a modelagem, um passo importante na análise de dados é fazer a análise de sensibilidade pois ela fornecerá uma indicação da inadequacidade do modelo ajustado ou de observações influentes. Sendo os modelos de regressão sensíveis aos pressupostos do modelo adjacente, a realização de uma análise de sensibilidade é amplamente recomendável. Cook (1986) usou essa ideia para impulsionar a avaliação da análise de influência. Ele sugere que mais preconiza que mais confiança pode ser depositada em um modelo relativamente estável sob pequenas perturbações. Outra abordagem proposta por Cook (1986), é ponderar observações em vez de removê-las. Trabalhos anteriores sob curvaturas de influência local em modelos de regressão para dados censurados podem ser encontrados em Escobar e Meeker (1992), Ortega, Bolfarine e Paula (2003), Ortega, Cancho e Paula (2009), Ortega, Cordeiro e Hashimoto (2011), Silva, Ortega e Cordeiro (2008), Silva et al. (2009) e Hashimoto et al. (2010). Os cálculos de influências locais podem ser realizadas para o modelo (4.2) considerando censura informativa. Se a verossi-

milhança deslocada $LD(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ é usada, em que $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denota o EMV sob o modelo perturbado, então a curvatura normal para $\boldsymbol{\theta}$ na direção $\mathbf{d}$, $\|\mathbf{d}\| = 1$, é dada por $C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^T \boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta} \mathbf{d}|$, em que $\boldsymbol{\Delta}$ é uma matriz $(p+6) \times n$ que depende do esquema de perturbação. Os elementos desta matriz são dados por $\Delta_{vi} = \partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \theta_v \partial \omega_i$, $i = 1, 2, \dots, n$ e $v = 1, 2, \dots, p+6$, avaliados em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$; $\boldsymbol{\omega}_0$ é o vetor não-perturbado. Para o modelo de regressão LOLLW com censura informativa, os elementos de $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ podem ser obtidos a partir do autor, mediante solicitação. As curvaturas normais $C_{\mathbf{d}}(\boldsymbol{\kappa})$, $C_{\mathbf{d}}(\boldsymbol{\nu})$ e $C_{\mathbf{d}}(\phi)$ também podem ser calculadas construir vários gráficos de índices, incluindo, por exemplo, o gráfico de índices de $\mathbf{d}_{max}$, o autovetor correspondente a $C_{\mathbf{d}_{max}}$, o maior autovetor associado à matriz $\mathbf{B} = -\boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}$, e o gráfico de índices $C_{\mathbf{d}_i}(\boldsymbol{\kappa})$, $C_{\mathbf{d}_i}(\boldsymbol{\nu})$ e $C_{\mathbf{d}_i}(\phi)$, os quais juntos são denotados por influência local total. Lesaffre e Verbeke (1998) sugeriram o cálculo da estatística $C_{\mathbf{d}_i}$, em que $\mathbf{d}_i$ denota um vetor $n \times 1$ de zeros com o valor um na i -ésima posição. Assim, a curvatura na direção $\mathbf{d}_i$ assume a forma $C_i = 2|\boldsymbol{\Delta}_i^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}_i|$, em que $\boldsymbol{\Delta}_i^T$ denota a i -ésima linha de $\boldsymbol{\Delta}$. Desta forma, um método utilizado para classificar a i -ésima observação como possível observação influente é quando $C_i \geq 2\bar{C}$, em que $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$.

Em seguida, para três esquemas de perturbação, foi calculada a seguinte matriz:

$$\boldsymbol{\Delta} = (\Delta_{vi})_{(p+6) \times n} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_v} \right)_{(p+6) \times n}, \quad v = 1, \dots, p+6 \quad \text{e} \quad i = 1, \dots, n.$$

O modelo (4.2) com censura informativa e o logaritmo da função de verossimilhança dada por (4.12) foram considerados. O vetor de pesos foi denotado por $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$.

• Perturbação de casos

Neste caso, o logaritmo da função de verossimilhança assume a forma

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{\omega}) &= \{r\phi \log(\phi) + r \log[\Gamma(\phi+1)] - r \log[\Gamma(\phi)]\} \sum_{i=1}^n \omega_i + \sum_{i=1}^n \delta_i \omega_i \log \left[h_0^{(T^*)}(y_i|\boldsymbol{\kappa}) \right] \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \omega_i \log \left[h_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right] \\ &\quad - (\phi+1) \sum_{i=1}^n \omega_i \log \left[\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{(C^*)}(y_i|\boldsymbol{\nu}) \right], \end{aligned}$$

em que $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ e $h_0^{(T^*)}$, $h_0^{(C^*)}$, $H_0^{(T^*)}$ e $H_0^{(C^*)}$ são definidas em (4.12). Ainda, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\kappa}}^T, \boldsymbol{\Delta}_{\boldsymbol{\nu}}^T, \boldsymbol{\Delta}_{\phi}^T)^T$ podem ser calculadas numericamente.

• Perturbação da variável resposta

Agora, considera-se que cada observação y_i é perturbada como $y_{iw} = y_i + \omega_i S_y$, em

que S_y é um fator de escala que pode ser estimado pelo desvio padrão da variável resposta observada Y e $\omega_i \in \mathbf{R}$. O logaritmo da função de verossimilhança se reduz-a

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{\omega}) &= r\phi \log(\phi) + r \log[\Gamma(\phi + 1)] - r \log[\Gamma(\phi)] + \sum_{i=1}^n \delta_i \log \left[h_0^{\dagger(T^*)}(y_i|\boldsymbol{\kappa}) \right] \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \log \left[h_0^{\dagger(C^*)}(y_i|\boldsymbol{\nu}) \right] \\ &\quad - (\phi + 1) \sum_{i=1}^n \log \left[\phi + H_0^{\dagger(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{\dagger(C^*)}(y_i|\boldsymbol{\nu}) \right], \end{aligned}$$

em que

$$h_0^{\dagger(T^*)}(y_i|\boldsymbol{\kappa}) = \frac{\alpha^{(T^*)} \exp(z_i^{\dagger(T^*)})(1 - u_i^{\dagger(T^*)})^{\alpha^{(T^*)}-1}}{\sigma^{(T^*)} \left\{ (1 - u_i^{\dagger(T^*)})^{\alpha^{(T^*)}} + [u_i^{\dagger(T^*)}]^{\alpha^{(T^*)}} \right\}},$$

$$h_0^{\dagger(C^*)}(y_i|\boldsymbol{\nu}) = \frac{\alpha^{(C^*)} \exp(z_i^{\dagger(C^*)})(1 - u_i^{\dagger(C^*)})^{\alpha^{(C^*)}-1}}{\sigma^{(C^*)} \left\{ (1 - u_i^{\dagger(C^*)})^{\alpha^{(C^*)}} + [u_i^{\dagger(C^*)}]^{\alpha^{(C^*)}} \right\}},$$

$$H_0^{\dagger(T^*)}(y_i|\boldsymbol{\kappa}) = \log \left\{ (1 - u_i^{\dagger(T^*)})^{\alpha^{(T^*)}} + [u_i^{\dagger(T^*)}]^{\alpha^{(T^*)}} \right\} - \alpha^{(T^*)} \log(u_i^{\dagger(T^*)}),$$

$$H_0^{\dagger(C^*)}(y_i|\boldsymbol{\nu}) = \log \left\{ (1 - u_i^{\dagger(C^*)})^{\alpha^{(C^*)}} + [u_i^{\dagger(C^*)}]^{\alpha^{(C^*)}} \right\} - \alpha^{(C^*)} \log(u_i^{\dagger(C^*)}),$$

$$u_i^{\dagger(T^*)} = \exp\{-\exp[z_i^{\dagger(T^*)}]\}, \quad u_i^{\dagger(C^*)} = \exp\{-\exp[z_i^{\dagger(C^*)}]\},$$

$$z_i^{\dagger(T^*)} = \frac{[(y_i + \omega_i S_y) - \mathbf{x}_i^\top \boldsymbol{\beta}^{(T^*)}]}{\sigma^{(T^*)}}, \quad z_i^{\dagger(C^*)} = \frac{[(y_i + \omega_i S_y) - \mu^{(C^*)}]}{\sigma^{(C^*)}},$$

e $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$. A matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\kappa}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\nu}}^\top, \boldsymbol{\Delta}_{\phi}^\top)^\top$ é obtida numericamente.

• Perturbação de uma variável explanatória

Agora, considere uma perturbação aditiva sob uma particular variável explanatória contínua, diga-se X_q , definindo $x_{iq\omega} = x_{iq} + \omega_i S_q$, em que S_q é um fator escala dado pelo desvio padrão da variável explanatória X_q e $\omega_i \in \mathfrak{R}$. O logaritmo da função de

verossimilhança perturbada tem a forma

$$\begin{aligned}
 l(\boldsymbol{\theta}|\boldsymbol{\omega}) &= r\phi \log(\phi) + r \log[\Gamma(\phi + 1)] - r \log[\Gamma(\phi)] + \sum_{i=1}^n \delta_i \log \left[h_0^{\dagger\dagger(T^*)}(y_i|\boldsymbol{\kappa}) \right] \\
 &\quad + \sum_{i=1}^n (1 - \delta_i) \log \left[h_0^{\dagger\dagger(C^*)}(y_i|\boldsymbol{\nu}) \right] \\
 &\quad - (\phi + 1) \sum_{i=1}^n \log \left[\phi + H_0^{\dagger\dagger(T^*)}(y_i|\boldsymbol{\kappa}) + H_0^{\dagger\dagger(C^*)}(y_i|\boldsymbol{\nu}) \right],
 \end{aligned}$$

em que

$$h_0^{\dagger\dagger(T^*)}(y_i|\boldsymbol{\kappa}) = \frac{\alpha^{(T^*)} \exp(z_i^{\dagger\dagger(T^*)})(1 - u_i^{\dagger\dagger(T^*)})^{\alpha^{(T^*)}-1}}{\sigma^{(T^*)} \left\{ (1 - u_i^{\dagger\dagger(T^*)})^{\alpha^{(T^*)}} + [u_i^{\dagger\dagger(T^*)}]^{\alpha^{(T^*)}} \right\}},$$

$$h_0^{\dagger\dagger(C^*)}(y_i|\boldsymbol{\nu}) = \frac{\alpha^{(C^*)} \exp(z_i^{\dagger\dagger(C^*)})(1 - u_i^{\dagger\dagger(C^*)})^{\alpha^{(C^*)}-1}}{\sigma^{(C^*)} \left\{ (1 - u_i^{\dagger\dagger(C^*)})^{\alpha^{(C^*)}} + [u_i^{\dagger\dagger(C^*)}]^{\alpha^{(C^*)}} \right\}},$$

$$H_0^{\dagger\dagger(T^*)}(y_i|\boldsymbol{\kappa}) = \log \left\{ (1 - u_i^{\dagger\dagger(T^*)})^{\alpha^{(T^*)}} + [u_i^{\dagger\dagger(T^*)}]^{\alpha^{(T^*)}} \right\} - \alpha^{(T^*)} \log(u_i^{\dagger\dagger(T^*)}),$$

$$H_0^{\dagger\dagger(C^*)}(y_i|\boldsymbol{\nu}) = \log \left\{ (1 - u_i^{\dagger\dagger(C^*)})^{\alpha^{(C^*)}} + [u_i^{\dagger\dagger(C^*)}]^{\alpha^{(C^*)}} \right\} - \alpha^{(C^*)} \log(u_i^{\dagger\dagger(C^*)}),$$

$$u_i^{\dagger\dagger(T^*)} = \exp\{-\exp[z_i^{\dagger\dagger(T^*)}]\}, \quad u_i^{\dagger\dagger(C^*)} = \exp\{-\exp[z_i^{\dagger\dagger(C^*)}]\},$$

$$z_i^{\dagger\dagger(T^*)} = \frac{y_i - \mathbf{x}_i^{\dagger\dagger\top} \boldsymbol{\beta}^{(T^*)}}{\sigma^{(T^*)}}, \quad z_i^{\dagger\dagger(C^*)} = \frac{y_i - \mu^{(C^*)}}{\sigma^{(C^*)}},$$

$\mathbf{x}_i^{\dagger\dagger\top} \boldsymbol{\beta} = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_q (x_{iq} + \omega_i S_q) + \cdots + \beta_p x_{ip}$ e $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$. A matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\kappa}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\nu}}^\top, \boldsymbol{\Delta}_{\phi}^\top)^\top$ é obtida numericamente.

4.5 Análise residual

A avaliação do modelo ajustado é uma importante etapa da análise de dados, particularmente, em modelos de regressão, a análise residual é uma ferramenta muito útil para validar o modelo ajustado. O exame dos resíduos pode ser usado, em instância, para detectar a presença de outliers, a ausência de componentes na parte sistemática do modelo e afastamentos das pressuposições assumidas para o erro e variância. Contudo,

encontrar resíduos apropriados em modelos de regressão não-normais tem sido um importante tópico de pesquisa, particularmente sob censura. Para mais detalhes, ver por exemplo, Ortega, Cancho e Paula (2009), Hashimoto et al. (2010) e Silva, Ortega e Paula (2011).

• Resíduo martingale

Em modelos de tempo de vida paramétricos com censura informativa, o resíduo martingale pode ser expresso por

$$rm_i = \delta_i \{1 + \log[S^{(T^*)}(y_i|\boldsymbol{\kappa}, \phi)]\} + (1 - \delta_i) \{\log[S^{(C^*)}(y_i|\boldsymbol{\nu}, \phi)]\}$$

com $\delta_i = 1$ indicando que a observação é não-censurada e $\delta_i = 0$ indicando que observação é censurada, $S^{(T^*)}(y|\boldsymbol{\kappa}, \phi)$ e $S^{(C^*)}(y|\boldsymbol{\nu}, \phi)$ são as funções de sobrevivência para os tempos de falha e censura, respectivamente, calculadas a partir da equação (4.8).

Desta maneira, o resíduo martingale para o modelo de regressão LOLIW com censura informativa assume a seguinte forma

$$r_{M_i} = \begin{cases} 1 + \phi \log \left\{ \frac{\phi}{\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa})} \right\} & \text{if } \delta_i = 1, \\ \phi \log \left\{ \frac{\phi}{\phi + H_0^{(C^*)}(y_i|\boldsymbol{\nu})} \right\} & \text{if } \delta_i = 0, \end{cases} \quad (4.13)$$

em que $H_0^{(T^*)}(y_i|\boldsymbol{\kappa})$ e $H_0^{(C^*)}(y_i|\boldsymbol{\nu})$ são definidas na Seção 4.3.

• Resíduo deviance modificado

Outra possibilidade de análise residual, é usar uma transformação do resíduo martingale baseada na componente deviance residual para o modelo de riscos proporcionais de Cox, sem variáveis explanatórias dependentes no tempo introduzida por Therneau, Grambsch e Fleming (1990), e definida por

$$r_{D_i} = \text{sign}(r_{M_i}) \left\{ -2 \left[r_{M_i} + \delta_i \log \left(\delta_i - r_{M_i} \right) \right] \right\}^{1/2},$$

em que r_{M_i} é o resíduo martingale apresentado em (4.13). Por conseguinte, o resíduo deviance modificado para o modelo de regressão LOLIW com censura informativa pode ser expresso por

$$r_{D_i} = \begin{cases} \text{sign}(\hat{r}_{M_i}) \left\{ -2 \left[1 + \log \left(\frac{\phi}{\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa})} \right)^\phi + \log \left\{ -\phi \log \left(\frac{\phi}{\phi + H_0^{(T^*)}(y_i|\boldsymbol{\kappa})} \right) \right\} \right] \right\}^{1/2} \\ \text{se } \delta_i = 1, \\ \\ \text{sin}(\hat{r}_{M_i}) \left\{ -2\phi \log \left(\frac{\phi}{\phi + H_0^{(C^*)}(y_i|\boldsymbol{\nu})} \right) \right\}^{1/2} \\ \text{if } \delta_i = 0, \end{cases}$$

em que $H_0^{(T^*)}(y_i|\boldsymbol{\kappa})$ e $H_0^{(C^*)}(y_i|\boldsymbol{\nu})$ são definidas na Seção 4.3.

Essa transformação é usada em ordem para obter um novo resíduo distribuído simetricamente em torno do zero.

4.6 Aplicação

Nesta seção, é apresentada uma aplicação para analisar o mecanismo de censura informativa. Os dados utilizados são obtidos do livro “Modelling Survival Data in Medical Research” (COLLETT, 2003). Durante o tratamento da leucemia, os pacientes muitas vezes são submetidos a um transplante de médula óssea, em ordem para ajudar a trazer de volta seus glóbulos a um nível normal. Um efeito colateral, potencialmente fatal, é a doença do enxerto-*versus*-hospedeiro, em que as células transplantadas atacam as células hospedeiras. Os dados aqui considerados, constam de $n = 37$ pacientes que estavam em remissão da leucemia mielóide aguda (LMA) ou leucemia linfócita aguda (LLA) ou em fase de leucemia mielóide crônica (LMC) e receberam um transplante de médula óssea alogênica não-empobrecido. Nesta aplicação, as variáveis consideradas foram

- t_i : tempo de sobrevivência (em dias) de pacientes que estavam em remissão completa da leucemia mielóide aguda (LMA) ou leucemia linfócita aguda (LLA) ou na fase de leucemia mielóide crônica (LMC) e receberam um transplante de medula óssea alogênico não-empobrecido;
- δ_i : indicadora de censura;
- x_{i1} : gravidez do doador (0=não, 1=sim);
- x_{i2} : doença do doador-*versus*-hospedeiro (0=não, 1=sim).

Análise descritiva

Das 37 observações dos dados, o evento de interesse ocorreu em 17, ou seja, cerca de 46% das observações falharam e 56% foram censuradas. A Tabela 4.1 fornece algumas medidas descritivas para os dados de transplante de medula óssea.

Tabela 4.1 - Estatísticas descritivas para os dados de transplante de medula óssea

Média	Mediana	DP	Variância	Assimetria	Curtose	Mín.	Máx.
Tempos de falha							
271,82	142	306,91	94.193,15	1,76	2,18	41	1.181
Tempos de censura							
1.008,05	1.006	319,16	10.1865,84	0,1161	-1,52	572	1.504

A Figura 4.1 apresenta os histogramas para todos os tempos observados, apenas para os tempos de falha e apenas para os tempos de censura. Para os três casos fica difícil assumir uma distribuição, visto que, graficamente, não apresentam formas conhecidas.

Figura 4.1 - Histogramas para os tempos observados (a), para os tempos de falha (b) e para os tempos de censura (c)

Na Tabela 4.2 são apresentadas os EMVs, os correspondentes erros-padrão (EP, em parênteses) e os valores das estatísticas AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) e CAIC (Consistent Akaike Information Criterion). Observe-se que para os tempos censurados, os valores das estatísticas AIC, CAIC e BIC foram menores para a distribuição Weibull, indicando um melhor ajuste que a distribuição OLLW.

No entanto, para os tempos de falha, obteve-se menores valores dessas estatísticas para o modelo OLLW.

Tabela 4.2 - EMVs, EP (em parênteses) e estatísticas AIC, CAIC e BIC para os tempos de censura e de falha dos dados de transplante de médula óssea

Modelo	γ	λ	α	AIC	CAIC	BIC
Tempos de censura						
Weibull	3,6468 (0,6518)	1121,06 (72,6025)	1 (-)	289,8	290,1	293
OLLW	7,5219 (4,1264)	1152,68 (84,0535)	0,4348 (0,2704)	290,4	291,1	295,5
Tempos de falha						
Weibull	1,0581 (0,1857)	278 (67,6267)	1 (-)	228,4716	229,3288	230,1381
OLLW	0,4829 (0,5308)	279,95 (237,1882)	3,9954 (4,4379)	219,6419	221,4881	222,1416

As Figuras 4.2(a) e 4.2(b) exibem os histogramas das densidades ajustadas Weibull e OLLW para os tempos censurados e de falha, respectivamente. Observa-se a partir da Figura 4.2(a) que a distribuição Weibull produz um melhor ajuste para os tempos censurados que o modelo OLLW. Por outro lado, a Figura 4.2(b) revela que a distribuição OLLW fornece um melhor ajuste para os tempos de falha.

Figura 4.2 - Densidades Weibull e OLLW estimadas para: (a) tempos censurados e (b) tempos de falha

O próximo passo será o ajuste do modelo de regressão LOLLW com censura informativa

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma z_i,$$

em que a variável Y_i tem a distribuição LOLLW dada em (4.3), para $i = 1, 2, \dots, 37$.

Na Tabela 4.3 são listadas as EMVs para o modelo de regressão LOLLW com censura não-informativa e informativa. Observa-se que as covariáveis x_{i2} = doador-*versus*-hospedeiro torna-se significativa quando se considera a censura informativa. Também, os erros-padrão são bem menores, indicando um melhor ajuste quando considerado o cenário de censura informativa.

Tabela 4.3 - EMVs, Erro-Padrão (E.P.) e estatísticas AIC, CAIC e BIC para o modelo de regressão LOLLW ajustado aos dados de transplante de medula óssea, considerando censura não-informativa e informativa, respectivamente

θ	Não-informativa			Informativa		
	Estimativa	E.P.	Valor- p	Estimativa	E.P.	valor- p
α	0,2084	0,2820	-	0,0973	0,0188	-
σ	0,2357	0,3153	-	0,1304	0,0098	-
μ_c	-	-	-	7,0042	0,0713	-
σ_c	-	-	-	0,2457	0,0434	-
ϕ	-	-	-	5,2310	3,9063	-
β_0	8,0774	0,5790	< 0,001	9,2274	0,0032	< 0,001
β_1	-1,0546	0,4662	0,0304	-0,9774	0,3430	< 0,001
β_2	-1,1286	0,6205	0,0780	-2,2651	0,3216	< 0,001

Comparando modelos não-encaixados

Note que o modelo de regressão LOLLW com censura informativa e o modelo de regressão sem LOLLW com censura não-informativa não são encaixados. Uma alternativa é utilizar a estatística de verossimilhança generalizada (VG) para discriminar modelos não-encaixados como é discutido no livro de Cameron e Trivedi (1998, p. 184).

Considere a escolha de um dentre dois modelos não-encaixados - modelo F_θ com função de densidade $f(y_i|\mathbf{x}_i, \theta)$ e o modelo G_γ com função de densidade $g(y_i|\mathbf{x}_i, \gamma)$. Esta estatística é uma distância entre os dois modelos medidos em termos do critério de in-

formação de Kullback-Liebler definido por

$$T_{LR,NN} = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \log \frac{f(y_i|\mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i, \hat{\boldsymbol{\gamma}})} \right\} \div \left\{ \frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i|\mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i, \hat{\boldsymbol{\gamma}})} \right)^2 - \left(\frac{1}{n} \sum_{i=1}^n \log \frac{f(y_i|\mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i, \hat{\boldsymbol{\gamma}})} \right)^2 \right\}.$$

Para modelos estritamente não-aninhados, a estatística $T_{LR,NN}$ converge em distribuição para uma distribuição Normal padrão sob a hipótese nula de equivalência dos modelos. Desta maneira, a hipótese nula não é rejeitada se $|T_{LR,NN}| \leq z_{\frac{\alpha}{2}}$. Por outro lado, rejeita-se (ao nível de significância α) a hipótese de equivalência dos modelos em favor do modelo F_θ ser melhor (ou pior) que o modelo G_γ se $T_{LR,NN} > z_\alpha$ (ou $T_{LR,NN} < -z_\alpha$).

Neste trabalho será usada (4.3) para representar a fdp sob censura informativa ($f(y_i|\mathbf{x}_i, \boldsymbol{\theta})$) e censura não-informativa ($g(y_i|\mathbf{x}_i, \boldsymbol{\gamma})$).

A estatística de teste VG produziu $TR_{LR,NN} = -22,6911$. Assim, como $TR_{LR,NN} < -1,96$, rejeita-se (ao nível de 5% de significância) a hipótese nula de equivalência entre os modelos LOLLW com censura informativa e com censura não-informativa. Ainda, o valor desta estatística indica que o modelo com censura informativa é o melhor modelo para os dados aqui utilizados.

4.7 Análise de influência local

Nesta seção, são analisadas influências locais com respeito aos dados de transplante de medula óssea usando o modelo de regressão LOLLW com censura informativa.

4.7.1 Perturbação de casos

Ao aplicar a metodologia de influência local, desenvolvida na Seção 4.4, no qual a perturbação de casos é usada, o valor $C_{\mathbf{d}_{max}} = 1,9204$ foi obtido como uma curvatura máxima. Na Figura 4.3a, é apresentado o gráfico de índices *versus* o autovetor correspondente a $\mathbf{d}_{max}$. O gráfico de influência local total C_i é apresentado na Figura 4.3b. As observações #1, #8, #25 e #28 encontram-se bastante “afastadas” das demais.

Figura 4.3 - Gráficos considerando o esquema de perturbação de casos para θ : (a) influência local $\mathbf{d}_{max}$ e (b) influência local total, baseado no modelo ajustado aos dados de transplante de médula óssea

4.7.2 Perturbação da variável resposta

Um outro esquema de perturbação utilizado, é aquele em que as influências das perturbações sobre os tempos de sobrevivência observados são analisados. O valor para a curvatura máxima, neste caso, é $C_{d_{max}} = 97,708$. A Figura 4.4a mostra o gráfico de $\mathbf{d}_{max}$ *versus* os índices das observações e indica que a observação #9 se resalta em relação as outras. A Figura 4.4b fornece os gráficos de influência local total (C_i), segundo o qual as observações #9 e #25 se destacam novamente.

Figura 4.4 - Gráficos considerando o esquema de perturbação da variável resposta para θ : (a) influência local $\mathbf{d}_{max}$ e (b) influência local total, baseado no modelo ajustado aos dados de transplante de médula óssea

4.7.3 Impacto das observações influentes detectadas

A partir da análise de diagnóstico detectou-se as observações #25 e #28 como possíveis pontos influentes. A observação #25 corresponde ao menor tempo de censura ($t_{25} = 572$) e o tempo de sobrevivência #28 é o maior tempo de falha com $x_{i1} = 1$ e $x_{i2} = 1$ ($t_{28} = 479$).

Em ordem para avaliar o impacto destas duas observações sobre as estimativas dos parâmetros, o modelo foi reajustado considerando algumas situações. Primeiro, foram eliminadas individualmente cada uma destas duas observações. Posteriormente, foi removido o conjunto “A” (conjunto dos dados originais) a totalidade das possíveis observações influentes.

Na Tabela 4.4 são apresentadas as mudanças relativas (em percentagens) de cada estimativa definida por $RC_{\theta_j} = [(\hat{\theta}_j - \hat{\theta}_j(A))/\hat{\theta}_j] \times 100$, e os correspondentes valores- p , em que $\hat{\theta}_j(A)$ é a EMV de θ_j depois do conjunto “A” de observações ser removido. A Tabela 4.4 fornece os seguintes conjuntos: $A_1 = \{\#25\}$, $A_2 = \{\#28\}$ e $A_3 = \{\#25, \#28\}$. Ela indica que as estimativas do modelo de regressão LOLLW com censura informativa não são altamente sensíveis sob a deleção das possíveis observações influentes. Em geral, a significância das estimativas dos parâmetros não mudam (ao nível de 5% de significância) após a remoção do conjunto I . Portanto, observa-se que não há mudanças inferenciais após a remoção das observações identificadas nos gráficos de diagnósticos.

Tabela 4.4 - Mudanças relativas [$-RC$ - em %], estimativas e seus valores- p (em parênteses) para o conjunto correspondente

Conjunto (I)	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\mu}_c$	$\hat{\sigma}_c$	$\hat{\phi}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
A	-	-	-	-	-	-	-	-
	0,0973	0,1304	7,0042	0,2457	5,2310	9,2274	-0,9774	-2,2651
	(-)	(-)	(-)	(-)	(-)	(< 0,0010)	(0,0075)	(< 0,0010)
A ₁	[-1,0637]	[-0,1992]	[-0,2986]	[1,3553]	[-2,4480]	[-0,8348]	[-12,5455]	[-8,2005]
	0,0983	0,1307	7,0251	0,2424	5,3590	9,3044	-1,1001	-2,4508
	(-)	(-)	(-)	(-)	(-)	(< 0,0010)	(0,0075)	(< 0,0010)
A ₂	[3,2922]	[-1,3677]	[0,2182]	[-20,3612]	[0,3638]	[-0,3187]	[11,6610]	[-1,5614]
	0,0943	0,1322	6,9889	0,2957	5,2119	9,2568	-0,8635	-2,3004
	(-)	(-)	(-)	(-)	(-)	(< 0,0010)	(0,0273)	(< 0,0010)
A ₃	[4,8551]	[4,6130]	[-0,5171]	[17,4850]	[22,2952]	[1,5168]	[-67,8432]	[10,8243]
	0,0925	0,1244	7,0405	0,2027	4,0647	9,0874	-1,6406	-2,0199
	(-)	(-)	(-)	(-)	(-)	(< 0,0010)	(< 0,0010)	(< 0,0010)

4.7.4 Análise residual

Para detectar possíveis observações atípicas no ajuste do modelo de regressão LOLLW para censura não-informativa e no modelo de regressão com censura informativa, construi-

se o gráfico de índices *versus* r_{D_i} (ver Figura 4.5). A Figura 4.5(a) indica que os resíduos não estão aleatoriamente distribuídos em torno do zero (censura não-informativa). Este gráfico também mostra que os resíduos revela a formação de dois grupos. Em contrapartida, a Figura 4.5b apresenta os resíduos melhor distribuídos em torno do zero quando o modelo de regressão LOLLW com censura informativa é considerado. Assim, a Figura 4.5b mostra que o modelo de regressão com censura informativa é mais apropriado para ajustar os dados, uma vez que não apresenta valores extremos.

Figura 4.5 - Gráfico do componente deviance residual *versus* índices para os dados de transplante de médula óssea: (a) modelo de regressão LOLLW sob censura não-informativa e (b) modelo de regressão LOLLW sob censura informativa

4.8 Conclusões

Neste trabalho foi introduzido um estudo da distribuição log-odd log-Logística Weibull (LOLLW) e construído o modelo de regressão LOLLW para investigar o mecanismo de censura informativa em um tipo de modelo de regressão locação-escala. O método de estimação por máxima verossimilhança foi utilizado para estimar os parâmetros do modelo. Algumas medidas de diagnóstico foram adotadas considerando três esquemas de perturbação no novo modelo de regressão com censura informativa. Ainda, foram definidos dois tipos de resíduos, a partir do modelo ajustado afim de avaliar características das pressuposições da distribuição do erro e observações atípicas. A flexibilidade, relevância prática e aplicabilidade do modelo de regressão proposto é ilustrado por meio de um conjunto de dados real. O modelo de regressão LOLLW ajustado com censura informativa é mais eficaz para os dados em estudo pois seu poder preditivo é melhor, como mostrado pelo erros-padrão menores das estimativas de máxima verossimilhança dos parâmetros do

modelo e também por produzir resíduos melhores quando comparados com o modelo com censura não-informativa.

Referências

- CAMERON, A.C., TRIVEDI, P.K. **Regression analysis of count data**. New York: Cambridge University Press, 1998. 489p.
- COLLETT, D. **Modelling survival data in medical research**. 2nd ed. London: Chapman & Hall, 2003. 391p.
- COOK, R.D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.
- CRUZ, J.N., ORTEGA, E.M.M., CORDEIRO, G.M., ALIZADEH, M. Distribuição Odd log-logística Weibull. In: 21º SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 2014, Natal. **Anais eletrônicos**. Natal: UFRN, 2014. Disponível em: <<http://www.ime.usp.br/~abe/sinape2014/trabalhos/anais#106>>. Acesso em: 11 nov. 2015.
- CRUZ, J.N.; ORTEGA, E.M.M.; CORDEIRO, G.M. The log-odd log-Logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. **Journal of Statistical Computation and Simulation**, London, v.1, p.1-23, 2015.
- DOORNIK, J. Ox: **Object-oriented matrix programming using Ox**. London: Timberlake Consultants Ltd, 2007. 461p.
- ESCOBAR, L.A.; MEEKER, W.Q. Assessing influence in regression analysis with censored data. **Biometrics**, Washington, v.48, p.507-528, 1992.
- FREITAS, L.A.; RODRIGUES, J. Standard Exponential cure rate model with informative censoring. **Communications in Statistics - Simulation and Computation**, London, v.42, p.8-23, 2013.
- HASHIMOTO, E.M.; ORTEGA, E.M.M.; CANCHO, V.G.; CORDEIRO, G.M. The log-exponentiated Weibull regression model for interval-censored data. **Computational Statistics and Data Analysis**, Amsterdam, v.54, p.1017-1035, 2010.
- HUANG, X.; WOLFE, R.A. A frailty model for informative censoring. **Biometrics**, Washington, v.58, p.510-520, 2002.
- HUANG, X.; ZHANG, N. Regression survival analysis with an assumed copula for dependent censoring: A sensitivity analysis approach. **Biometrics**, Washington, v.64, p.1090-1099, 2008.

- LAGAKOS, S. W. General right censoring and its impact on the analysis of survival data. **Biometrics**, Washington, v.35, p.139-156, 1979.
- LAGAKOS, S.W.; WILLIAMS, J.S. A cone class of variable-sum models. **Biometrika**, London, v.65, p.181-189, 1978.
- LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, Washington, v.54, p.570-582, 1998.
- LU, X.; TSIATIS, A.A. Semiparametric estimation of treatment effect with time-lagged response in the presence of informative censoring. **Lifetime Data Analysis**, New York, v.17, p.566-593, 2011.
- 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.; PAULA G.A. Generalized log-Gamma regression models with cure fraction. **Lifetime Data Analysis**, New York, v.15, p.79-106, 2009.
- ORTEGA, E. M. M.; CORDEIRO, G. M.; HASHIMOTO, E. M. A log-linear regression model for the Beta-Weibull distribution. **Communications in Statistics - Simulation and Computation**, London, v.40, p.1206-1235, 2011.
- PHAM, H.; LAI, CHIN-DIEW. On recent generalizations of the Weibull distribution. **IEEE Transactions on Reliability**, New York, v.56, p.454-458, 2007.
- ROTNITZKY, A.; FARALL, A.; BERGESIO, A.; SCHARFSTEIN, D. Analysis of failure time data under competing censoring mechanisms. **Journal of the Royal Statistical Society-Series B (Statistical Methodology)**, London, v.69, p.307-327, 2007.
- SANTOS JR., P.C. **Análise de sobrevivência na presença de censura informativa**. 2012. 73 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- SCHARFSTEIN, D.O.; ROBINS, J.M. Estimation of the failure time distribution in the presence of informative censoring. **Biometrika**, Cambridge, v.89, p.617-634, 2002.
- SIANNIS, F. Applications of a parametric model for informative censoring. **Biometrics**, Washington, v.60, p.704-714, 2004.
- SIANNIS, F. Sensitivity analysis for multiple right censoring processes: Investigating mortality in psoriatic arthritis. **Statistics in Medicine**, Chichester, v.6, p.77-91, 2011.

- SIANNIS, F.; COPAS, J.; LU, G. Sensitivity analysis for informative censoring in parametric survival models. **Biostatistics**, Oxford, v.6, p.77-91, 2005.
- SILVA, G.O.; ORTEGA, E. M.M.; CORDEIRO, G.M. A log-extended Weibull regression model. **Computational Statistics and Data Analysis**, Amsterdam, v.53, p.4482-4489, 2009.
- SILVA, G.O.; ORTEGA, E.M.M.; GARIBAY, V.C.; BARRETO, M.L. Log-Burr XII regression models with censored data. **Computational Statistics and Data Analysis**, Amsterdam, v.52, p.3820-3842, 2008.
- SILVA, G.O.; ORTEGA, E.M.M.; PAULA, G.A. Residuals for log-Burr XII regression models in survival analysis. **Journal of Applied Statistics**, Abingdon, v.38, p.1435-1445, 2011.
- STAPLIN, N.D.; KIMBER, A.C.; COLLETT, D.; RODERICK, P.J. Dependent censoring in piecewise Exponential survival models. **Statistical methods in medical research**, London, v.24, p.325-341, 2015.
- THERNEAU, T.M.; GRAMBSCH, P.M.; FLEMING, T.R. Martingale-based residuals for survival models. **Biometrika**, Cambridge, v.77, p.147-160, 1990.
- TSIATIS, A. A noindentifiability aspect of the problem of competing risks. **Proceedings of the National Academy of Sciences**, Philadelphia, v.72, p.20-22, 1975.
- VAUPEL, J.W.; YASHIN, A.I. **The deviant dynamics of death in heterogeneous populations**. Laxenburg, Áustria: International Institute for Applied Systems Analysis, 1983. 29p.
- WANG, M.C.; QIN, J.; CHIANG, C.T. Analyzing recurrent event data with informative censoring. **Journal of the American Statistical Association**, Boston, v.96, p.1057-1065, 2001.
- ZHANG, J.; HEITJAN, D.F. A simple local sensitivity analysis tool for nonignorable coarsening: application to dependent censoring. **Biometrics**, Washington, v.62, p.1260-1268, 2006.

5 MODELO DE REGRESSÃO LOG-ODD LOG-LOGÍSTICO WEIBULL PARA QUALIDADE DE VIDA DE SAÚDE BUCAL

Resumo

Neste trabalho foram estudados modelos de regressão bivariados com distribuições marginais arbitrárias e distribuições conjuntas obtidas pelas cópulas de Frank e de Clayton. Proposto por He e Lawless (2005), estes modelos são aqui usados para dados bivariados dependentes com variáveis explanatórias usando a distribuição log-odd log-Logística Weibull (CRUZ; ORTEGA; CORDEIRO, 2015). Com o objetivo de avaliar o desempenho do modelo, foram utilizados alguns métodos propostos por Cook (1986) para detectar observações aberrantes ou influentes. Para o modelo proposto, consideraram-se os procedimentos inferenciais baseados nos parâmetros da máxima verossimilhança restrita. Métodos de análise de sensibilidade, tais como influências local e total, são obtidos sobre três esquemas de perturbação. Esta metodologia é ilustrada pelo estudo que tem por objetivo avaliar mudanças na qualidade de vida da saúde bucal de estudantes (OHRQoL) em um exame de acompanhamento de três anos e o impacto da incidência de cárie na OHRQoL dos adolescentes.

Palavras-chave: Cópula de Frank; Distribuição de Weibull; Influência local; Modelo de regressão

Abstract

In this work were studied the bivariate response regression models with arbitrary marginal distributions and joint distributions obtained by the Frank and Clayton copulas. Proposed by He and Lawless (2005) this models are used here for fitting dependent bivariate data with explanatory variables using the log-odd log-logistic Weibull distribution. In order to evaluate the model's performance were used some methods proposed by Cook (1986) to detect aberrant or influential observations. For the proposed model, it was considered the inferential procedures based on the restricted maximum likelihood parameters. Sensitivity analysis methods such as local and total influences were derived under three perturbation schemes. The methodology is illustrated by study had the objective to assess changes on schoolchildren's oral health-related quality of life (OHRQoL) in a follow-up exam after three years and to evaluate the impact of caries incidence on the OHRQoL of adolescents.

Keywords: Frank copula; Weibull distribution; Local influence; Regression model

5.1 Introdução

Nesse trabalho foi considerado um conjunto de dados fornecido pelo Departamento da Comunidade dentista, Divisão da Educação da Saúde e Promoção da Saúde da Faculdade de Odontologia de Piracicaba, Universidade de Campinas-UNICAMP. O presente estudo teve como objetivos avaliar as mudanças da qualidade de vida relacionada a saúde bucal de alunos (OHRQoL) em um exame de acompanhamento de três anos e avaliar o impacto de incidências de cáries na OHRQoL desses adolescentes.

Na análise estatística, dados longitudinais bivariados podem representar a ocorrência de eventos sucessivos dentro do mesmo indivíduo. Provavelmente, há uma dependência entres estes eventos e, investigar essa dependência pode ser o principal interesse em um ensaio clínico. Modelos individuais para cada evento são baseados sobre suposições de independência e não permite inferências em uma possível associação. A utilização de modelos bivariados parece mais adequada e esta abordagem tem sido utilizada sob diferentes cenários e pode ser encontrada em Nuñez (2005) e Chatterjee e Shih (2001). Além do uso das distribuições paramétricas multivariadas clássicas, as cópulas podem ser usadas para unir modelos marginais em modelos multivariados. A flexibilidade proporcionada pelas cópulas, sobre a seleção de marginais, resulta em incontáveis distribuições com propriedades distintas. Neste trabalho, as Cópulas de Frank e de Clayton foram usadas para modelos marginais lineares de locação-escala, incluindo covariáveis na modelagem. Para as distribuições marginais foi usada a distribuição log-odd log-Logística Weibull (LOLLW), a qual é uma generalização recente da distribuição de Weibull.

A parte inferencial do trabalho foi conduzida usando a teoria de máxima verossimilhança e as propriedades assintóticas dos estimadores. Diagnósticos de influência é um importante passo na análise de um conjunto de dados, pois fornece uma indicação da falta de ajuste do modelo considerado ou de observações influentes. Cook (1986) propôs uma abordagem de diagnóstico chamada influência local para avaliar o efeito de pequenas perturbações no modelo e/ou dados sobre as estimativas dos parâmetros. Muitos autores têm aplicado o método de influência local em modelos de regressão mais gerais que o modelo de regressão Normal. Além disso, alguns autores investigaram a influência local em modelos de análise de sobrevivência. Por exemplo, Pettit e Bin Daud (1989) investigaram influência local em modelos de regressão de riscos proporcionais, Escobar e Meeker (1992) adaptaram métodos de influência local para análise de regressão sob censura, Silva et al. (2010) consideraram o problema de avaliar influência local em modelo de regressão log-Weibull estendida, Hashimoto et al. (2013) investigaram influência local e global em modelo de regressão log-Gama generalizada para dados com censura intervalar e Ortega, Cancho e Paula (2013) aplicaram influência local no modelo de regressão log-Beta Weibull com aplicação para prever a recorrência de câncer de próstata. Recentemente, Fachini,

Ortega e Cordeiro (2013) adaptaram métodos de influência local para um modelo de regressão bivariado com fração de cura. Aqui, foi considerado um método para detectar indivíduos influentes considerando a abordagem de influência local para os modelos de regressão LOLLW considerando a metodologia de cópulas.

O trabalho está organizado como segue. Na Seção 5.2, é discutido os modelos de resposta bivariada usando as cópulas de Frank e de Clayton e a distribuição LOLLW. Na Seção 5.3, são propostos modelos de regressão LOLLW-Frank e LOLLW-Clayton e inferências baseadas nas estimativas de máxima verossimilhança. Na Seção 5.4, são apresentados algumas medidas de diagnósticos considerando a deleção de casos e as curvaturas normais de influência local e foram obtidas a função de verossimilhança sob diferentes esquemas de perturbação para os modelos de regressão com resposta bivariada. Na Seção 5.5, os métodos propostos são aplicados a um conjunto de dados reais. Na última seção, Seção 5.6, são apresentadas algumas conclusões finais.

5.2 Modelos de resposta bivariada usando as cópulas de Frank-LOLLW e Clayton-LOLLW

Os modelos considerados neste trabalho foram apresentados por He e Lawless (2005), e tem variável resposta bivariada com distribuição conjunta dada por:

$$F(y_1, y_2) = C_\lambda \left(\frac{y_1 - \mu_1}{\sigma_1}, \frac{y_2 - \mu_2}{\sigma_2} \right), \quad (5.1)$$

em que C é uma função cópula sobre R^2 , λ é o parâmetro de associação e as distribuições marginais de Y_1, Y_2 têm a forma de locação-escala

$$Y_k = \mu_k + z_k, \quad (5.2)$$

em que $-\infty < \mu_1 < \infty$ e $-\infty < \mu_2 < \infty$ são parâmetros de locação, $\sigma_1 > 0$ e $\sigma_2 > 0$ são parâmetros de escala e z_1 e z_2 são os erros do modelo.

Cópulas são funções que fornecem uma maneira de criar funções de distribuições multivariadas com diferentes estruturas de dependência baseadas em funções marginais arbitrárias. Seja Y_k uma variável aleatória com distribuição marginal contínua $F_k, k = 1, 2$. A distribuição conjunta será

$$F(y_1, y_2) = C_\lambda \{F_1(y_1), F_2(y_2)\},$$

condicionada a $F_k(Y_k) \sim U(0, 1), \forall k = 1, 2$. A distribuição conjunta C_λ descreve a dependência das variáveis aleatórias Y_1 e Y_2 , por meio do parâmetro de associação λ .

De uma lista longa de diferentes tipos de cópulas, serão apresentados aqui apenas dois exemplos: Frank (1979) e Clayton (1978) for $k = 2$. Claramente, muitas outras cópulas existem e detalhes podem ser encontrados em outros textos como Nelsen (2006). As expressões para estas cópulas selecionadas são apresentadas juntamente com algumas propriedades de limite e foram usadas para criar modelos de regressão com resposta bivariada. Essas cópulas são amplamente usadas e têm muitas propriedades atrativas, mais detalhes, ver por exemplo, Genest (1987) e Nelsen (2006).

- Cópula de Frank

A função de distribuição conjunta para a cópula de Frank é dada por

$$C_\lambda(y_1, y_2) = -\frac{1}{\lambda} \log \left\{ 1 + \frac{\{\exp[-\lambda F_1(y_1)] - 1\}\{\exp[-\lambda F_2(y_2)] - 1\}}{\exp(-\lambda) - 1} \right\}, \quad (5.3)$$

em que $\lambda \in \mathbf{R} \setminus \{0\}$, para $\lambda \rightarrow 0$ obtém-se $C_0(y_1, y_2) = F_1(y_1)F_2(y_2)$ com funções marginais arbitrárias $F_1(y_1)$ e $F_2(y_2)$. Para $\lambda \rightarrow -\infty$ obtém-se $C_{-\infty}(y_1, y_2) = \max(F_1(y_1) + F_2(y_2) - 1, 0)$ e analogamente para $\lambda \rightarrow \infty$ obtém-se $C_\infty(y_1, y_2) = \min(F_1(y_1), F_2(y_2))$.

- Cópula de Clayton

A função de distribuição conjunta para a cópula de Clayton é dada por

$$C_\lambda(y_1, y_2) = \{[F_1(y_1)]^{-\lambda} + [F_2(y_2)]^{-\lambda} - 1\}^{-1/\lambda}, \quad (5.4)$$

em que $\lambda \in [-1, +\infty)$ e $\lambda \neq 0$. Para $\lambda \rightarrow 0$, obtém-se $C_0(y_1, y_2) = F_1(y_1)F_2(y_2)$, para $\lambda = 1$, obtém-se $C_1(y_1, y_2) = \max(F_1(y_1) + F_2(y_2) - 1, 0)$ e analogamente para $\lambda \rightarrow \infty$, obtém-se $C_\infty(y_1, y_2) = \min(F_1(y_1), F_2(y_2))$.

Como pode ser visto, nas equações (5.3) e (5.4), é necessário definir as distribuições marginais $F_1(y_1)$ e $F_2(y_2)$. Neste trabalho, será adotada a distribuição log-odd log-Logística Weibull (LOLLW), apresentada por Cruz, Ortega e Cordeiro (2015).

Em estatística, as distribuições de Weibull e Valor-Extremo (log-Weibull) são os modelos mais populares em aplicações à dados reais. Quando o número de observações é grande, elas podem servir como uma distribuição aproximada para outros modelos. A função densidade de probabilidade (fdp) e função de distribuição acumulada (fda) da Log-Weibull (para $y \in \mathbf{R}$) são dadas, respectivamente, por

$$g(y; \mu, \sigma) = \frac{1}{\sigma} \exp \left[\left(\frac{y - \mu}{\sigma} \right) - \exp \left(\frac{y - \mu}{\sigma} \right) \right] \text{ e } G(y; \mu, \sigma) = 1 - \exp \left[- \exp \left(\frac{y - \mu}{\sigma} \right) \right] \quad (5.5)$$

em que $\mu \in \mathbf{R}$ é um parâmetro de locação e $\sigma > 0$ é um parâmetro de escala.

A função densidade acumulada (fda) da distribuição log-odd log-Logística Weibull (LOLLW), com um parâmetro de forma adicional $\alpha > 0$, é definida por

$$F(y; \mu, \sigma, \alpha) = \int_0^{\frac{G(y; \mu, \sigma)}{\bar{G}(y; \mu, \sigma)}} \frac{\alpha y^{\alpha-1}}{(1+y^\alpha)^2} dy = \frac{G(y; \mu, \sigma)^\alpha}{G(y; \mu, \sigma)^\alpha + \bar{G}(y; \mu, \sigma)^\alpha}, \quad (5.6)$$

em que $\bar{G}(y; \mu, \sigma) = 1 - G(y; \mu, \sigma)$.

A distribuição log-Weibull $G(y; \mu, \sigma)$ é claramente um caso especial de (5.6) quando $\alpha = 1$. Note que não há uma função complicada na equação (5.6), em contraste com a família beta generalizada (EUGENE; LEE; FAMOYE, 2002), a qual inclui dois parâmetros extras e também a função beta incompleta. A função densidade de probabilidade (fdp) da distribuição LOLW é dada por

$$f(y; \mu, \sigma, \alpha) = \frac{\alpha g(y; \mu, \sigma) \{G(y; \mu, \sigma)[1 - G(y; \mu, \sigma)]\}^{\alpha-1}}{\{G(y; \mu, \sigma)^\alpha + [1 - G(y; \mu, \sigma)]^\alpha\}^2}. \quad (5.7)$$

Ainda, o parâmetro de forma adicional α pode ser escrito como

$$\alpha = \frac{\log \left[\frac{F(y)}{F(y)} \right]}{\log \left[\frac{G(y)}{G(y)} \right]} \quad \text{e} \quad \bar{G}(y) = 1 - G(y).$$

Portanto, o parâmetro α representa o quociente do logaritmo da razão de chances para as distribuições gerada e de base.

A fda e a fdp da distribuição LOLW podem ser expressas como

$$F(y; \mu, \sigma, \alpha) = \frac{\{1 - \exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha}{\{1 - \exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha + \{\exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha} \quad (5.8)$$

e

$$f(y; \mu, \sigma, \alpha) = \frac{\alpha \exp(\frac{y-\mu}{\sigma}) \{\exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha \{1 - \exp[-\exp(\frac{y-\mu}{\sigma})]\}^{\alpha-1}}{\sigma \{ \{1 - \exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha + \{\exp[-\exp(\frac{y-\mu}{\sigma})]\}^\alpha \}^2}, \quad (5.9)$$

respectivamente. Note que $\alpha > 0$ é um parâmetro de forma. Doravante, uma variável aleatória com função de densidade (5.9) será denotada por $Y \sim \text{LOLLW}(\alpha, \mu, \sigma)$. Para $\sigma = 1$, a distribuição log odd log-Logística Exponencial (LOLLE) é obtida. Ainda, a distribuição LOLLE com $\alpha = 1$ reduz-se a distribuição log-Weibull.

A função quantílica (fq) é muito utilizada em estatísticas gerais. A equação (5.6) tem propriedades tratáveis especialmente por simulações, uma vez que sua fq toma a

simples forma

$$y = Q_Z(v) = Q_{LW} \left(\frac{v^{1/\alpha}}{[1-v]^{1/\alpha} + v^{1/\alpha}} \right), \quad (5.10)$$

em que $z = Q_{LW}(v) = \mu + \sigma \log[-\log(1-v)]$ é a fq da distribuição log-Weibull, obtida invertendo (5.5), ou seja, $G(y; \mu, \sigma) = v$, com $v \sim U(0, 1)$.

Assumindo que as distribuições marginais dos erros do modelo vêm de (5.2) e substituindo (5.8) e (5.9) nas equações (5.3) e (5.4), obtém-se a função de distribuição conjunta, a saber

- Função de distribuição conjunta Frank-LOLLW

$$F_F(y_1, y_2) = -\frac{1}{\lambda} \log \left\{ 1 + \frac{\left\{ \exp \left[\frac{-\lambda(1-u_1)^{\alpha_1}}{(1-u_1)^{\alpha_1} + u_1^{\alpha_1}} \right] - 1 \right\} \left\{ \exp \left[\frac{-\lambda(1-u_2)^{\alpha_2}}{(1-u_2)^{\alpha_2} + u_2^{\alpha_2}} \right] - 1 \right\}}{\exp(-\lambda) - 1} \right\} \quad (5.11)$$

- Função de distribuição conjunta Clayton-LOLLW

$$F_C(y_1, y_2) = \left\{ \left[\frac{(1-u_1)^{\alpha_1}}{(1-u_1)^{\alpha_1} + u_1^{\alpha_1}} \right]^{-\lambda} + \left[\frac{(1-u_2)^{\alpha_2}}{(1-u_2)^{\alpha_2} + u_2^{\alpha_2}} \right]^{-\lambda} - 1 \right\}^{-1/\lambda}. \quad (5.12)$$

As funções densidade de probabilidade conjunta são dadas por

- função densidade de probabilidade conjunta Frank-LOLLW

$$f_F(y_1, y_2) = \frac{\alpha_1 \alpha_2 \exp(z_1 + z_2) \lambda u_1^{\alpha_1} u_2^{\alpha_2} [-1 + \exp(\lambda)] \exp\{\lambda[1 + q_1 + q_2]\}}{\{\exp(\lambda) - \exp[\lambda(1 + q_1)] + \exp[\lambda(q_1 + q_2)] - \exp[\lambda(1 + q_2)]\}^2} \times \frac{(1-u_1)^{\alpha_1-1} (1-u_2)^{\alpha_2-1}}{\sigma_1 \sigma_2 [(1-u_1)^{\alpha_1} + u_1^{\alpha_1}]^2 [(1-u_2)^{\alpha_2} + u_2^{\alpha_2}]^2}. \quad (5.13)$$

- função densidade de probabilidade conjunta Clayton-LOLLW

$$f_C(y_1, y_2) = \frac{\alpha_1 \alpha_2 \exp(z_1 + z_2) u_1^{\alpha_1} u_2^{\alpha_2} (1-u_1)^{\alpha_1} (1-u_2)^{\alpha_2} (1+\lambda) (q_1 q_2)^{-(1+\lambda)}}{\sigma_1 \sigma_2 [(1-u_1)^{\alpha_1} + u_1^{\alpha_1}]^2 [(1-u_2)^{\alpha_2} + u_2^{\alpha_2}]^2 [q_1^{-\lambda} + q_2^{-\lambda} - 1]^{(2+1/\lambda)}} \quad (5.14)$$

em que

$$q_k = \frac{(1-u_k)^{\alpha_k}}{(1-u_k)^{\alpha_k} + u_k^{\alpha_k}}, \quad u_k = \exp[-\exp(z_k)]e \quad z_k = \frac{y_k - \mu_k}{\sigma_k}, \quad \text{para } k = 1, 2.$$

Os modelos (5.13) e (5.14) são referidos como modelos bivariados de Frank-LOLLW e de Clayton-LOLLW, respectivamente. Os modelos de Frank-LOLLW e de Clayton-LOLLW contêm como casos especiais muitas distribuições conhecidas, por exemplo, reduzem-se aos modelos bivariados de Frank-log-Weibull (Frank-LW) e de Clayton-log-Weibull (Clayton-LW), quando $\alpha_1 = \alpha_2 = 1$. Se $\sigma_1 = \sigma_2 = 1$, reduzem aos modelos bivariados de Frank-log odd log-Logístico Exponencial (Frank-LOLLE) e de Clayton-log odd log-Logístico Exponencial (Clayton-LOLLE). Se $\alpha_1 = \alpha_2 = 1$ e $\sigma_1 = \sigma_2 = 1$, os modelos de Frank-LOLLW e de Clayton-LOLLW reduzem-se aos modelos bivariados de Frank-log-Exponencial (Frank-LE) e de Clayton-log-Exponencial (Clayton-LE).

Alguns gráficos da função densidade de probabilidade da cópula de Frank são apresentados na Figura 5.1.

Figura 5.1 - (a): $\lambda = 0,01, \alpha_1 = 2, \alpha_2 = 0,75, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 2, \sigma_2 = 2$; (b): $\lambda = 5, \alpha_1 = 1,5, \alpha_2 = 1,3, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 2, \sigma_2 = 2$

Alguns gráficos da função densidade de probabilidade da cópula de Clayton são apresentados na Figura 5.2.

Figura 5.2 - (c): $\lambda = 0,01, \alpha_1 = 2, \alpha_2 = 0,75, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 2, \sigma_2 = 2$; (d): $\lambda = 5, \alpha_1 = 0,25, \alpha_2 = 1,75, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 2, \sigma_2 = 2$

5.3 Modelo de regressão e métodos de inferência

Em muitas aplicações práticas, a variável resposta é afetada por variáveis explanatórias ou covariáveis, tais como nível de colesterol, pressão sanguínea e muitas outras. Deste modo, é importante explorar a relação entre a variável resposta e as variáveis explanatórias. Uma abordagem baseada em modelos de regressão pode ser usada. Neste trabalho, considerou-se a classe de modelos locação-escala.

Sejam Y_1 e Y_2 representando duas variáveis aleatórias relacionadas a variável resposta bivariada de um indivíduo e assumindo que há uma relação linear entre as variáveis $Y_k; k = 1, 2$, e um vetor de covariáveis $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{pk})^T$. Um modelo de regressão log-linear pode ser representado por

$$Y_k = \mathbf{x}_k^T \boldsymbol{\beta}_k + \sigma_k z_k, \quad (5.15)$$

em que $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são os coeficientes da regressão associados aos vetores de covariáveis $\mathbf{x}_1$ e $\mathbf{x}_2$ com p termos, $\sigma_1 > 0$ e $\sigma_2 > 0$ são os parâmetros de escala e z_1 e z_2 são os erros do modelo com distribuição conjunta dada nas equações (5.13) e (5.14) e independentes de $\mathbf{x}_1$ e $\mathbf{x}_2$. Se z_1 e z_2 são independentes, o modelo dado em (5.2) é reduzido ao modelo de regressão locação-escala tradicional.

Considere uma amostra $(y_{1k}, \mathbf{x}_{1k}), \dots, (y_{nk}, \mathbf{x}_{nk})$, para $k = 1, 2$, de n observações independentes, o modelo dado em (5.15) e as funções densidade de probabilidade conjuntas dadas nas equações (5.13) e (5.14). Assim, o logaritmo da função de verossimilhança para o vetor de parâmetros $\boldsymbol{\psi} = (\lambda, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, em que $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^T, \sigma_k, \alpha_k)^T$ e $\boldsymbol{\beta}_k^T = (\beta_{1k}, \dots, \beta_{pk})$ para $k = 1, 2$, é dado por

- Frank-LOLLW

$$\begin{aligned}
l(\boldsymbol{\psi}) = & n \log \left\{ \frac{\lambda \alpha_1 \alpha_2 [\exp(\lambda) - 1]}{\sigma_1 \sigma_2} \right\} + \sum_{i=1}^n (z_{i1} + z_{i2}) + \alpha_1 \sum_{i=1}^n \log(u_{i1}) + \alpha_2 \sum_{i=1}^n \log(u_{i2}) \\
& + \lambda \sum_{i=1}^n (1 + q_{i1} + q_{i2}) + (\alpha_1 - 1) \sum_{i=1}^n \log(1 - u_{i1}) + (\alpha_2 - 1) \sum_{i=1}^n \log(1 - u_{i2}) \\
& - 2 \sum_{i=1}^n \log \{ \exp(\lambda) - \exp[\lambda(1 + q_{i1})] + \exp[\lambda(q_{i1} + q_{i2})] - \exp[\lambda(1 + q_{i2})] \} \\
& - 2 \sum_{i=1}^n \log[(1 - u_{i1})^{\alpha_1} + u_{i1}^{\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2})^{\alpha_2} + u_{i2}^{\alpha_2}], \tag{5.16}
\end{aligned}$$

- Clayton-LOLLW

$$\begin{aligned}
l(\boldsymbol{\psi}) = & n \log \left[\frac{(1 + \lambda) \alpha_1 \alpha_2}{\sigma_1 \sigma_2} \right] + \sum_{i=1}^n (z_{i1} + z_{i2}) + \alpha_1 \sum_{i=1}^n \log(u_{i1}) + \alpha_2 \sum_{i=1}^n \log(u_{i2}) \\
& + \alpha_1 \sum_{i=1}^n \log(1 - u_{i1}) + \alpha_2 \sum_{i=1}^n \log(1 - u_{i2}) - (1 + \lambda) \sum_{i=1}^n \log(q_{i1} q_{i2}) \\
& - 2 \sum_{i=1}^n \log[(1 - u_{i1})^{\alpha_1} + u_{i1}^{\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2})^{\alpha_2} + u_{i2}^{\alpha_2}] \\
& - \left(2 + \frac{1}{\lambda} \right) \sum_{i=1}^n \log[q_{1i}^\lambda + q_{2i}^\lambda - 1], \tag{5.17}
\end{aligned}$$

em que

$$q_{ik} = \frac{(1 - u_{ik})^{\alpha_k}}{(1 - u_{ik})^{\alpha_k} + u_{ik}^{\alpha_k}}, \quad u_{ik} = \exp[-\exp(z_{ik})] \quad \text{e} \quad z_{ik} = \frac{(y_{ik} - \mathbf{x}_{ik}^T \boldsymbol{\beta}_k)}{\sigma_k},$$

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

O estimador de máxima verossimilhança (EMV) para o vetor de parâmetros $\boldsymbol{\psi}$ pode ser obtido maximizando a função (5.16) ou (5.17). Os EMVs dos coeficientes da regressão e parâmetros desconhecidos são as soluções das equações não-lineares dos vetores escore $U_\lambda = 0$, $U_{\boldsymbol{\theta}_1} = \mathbf{0}$ e $U_{\boldsymbol{\theta}_2} = \mathbf{0}$. Para determinar as raízes das equações não-lineares, utilizou-se métodos iterativos. A subrotina MaxBFGS do programa de linguagem matricial Ox (ver, DOORNIK, 2007) foi utilizada para maximizar o logaritmo da função de verossimilhança $\ell(\boldsymbol{\psi})$. Os valores iniciais para $\boldsymbol{\beta}_k$, σ_k e λ , foram tomados a partir do ajuste do modelo de regressão bivariado de Frank-LW ou Clayton-LW com $\alpha_1 = \alpha_2 = 1$.

Para estimação intervalar e testes de hipóteses sobre os parâmetros do modelo, exige-se as matrizes de informação observada ou esperada $(2p + 7) \times (2p + 7)$ ($\ddot{\mathbf{L}}(\boldsymbol{\psi})$) e ($\mathbf{I}(\boldsymbol{\psi})$), calculadas numericamente. Sob condições de regularidade gerais, os intervalos de

confiança aproximados podem ser construídos individualmente para cada parâmetro com base na distribuição Normal multivariada $N_{2p+7}(0, \mathbf{I}(\hat{\boldsymbol{\psi}})^{-1})$, em que $\hat{\boldsymbol{\psi}}$ é o EMV de $\boldsymbol{\psi}$.

Os valores máximos do logaritmo das funções de verossimilhança restrita e não-restrita podem ser calculados, em ordem, para construir o teste de razão de verossimilhança (TRV) e testar alguns submodelos dos modelos de Frank-LOLLW ou de Clayton-LOLLW. Por exemplo, o teste de $H_0 : \alpha_1 = \alpha_2 = 1$ versus $H : H_0$ não é verdadeira é equivalente a comparar os modelos de Frank-LOLLW ou Clayton-LOLLW e Frank-LW ou Clayton-LW, respectivamente, e a estatística do teste reduz-se a

$$w = 2[\ell(\hat{\lambda}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\alpha}_1, \hat{\alpha}_2) - \ell(\tilde{\lambda}, \tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \tilde{\sigma}_1, \tilde{\sigma}_2, 1, 1)],$$

em que $\hat{\lambda}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\alpha}_1$ e $\hat{\alpha}_2$ são os EMVs sob H e $\tilde{\lambda}, \tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \tilde{\sigma}_1$ e $\tilde{\sigma}_2$ são as estimativas sob H_0 .

5.4 Diagnósticos de influência

Uma vez que os modelos de regressão são sensíveis às pressuposições do modelo subjacente, geralmente realizar uma análise de sensibilidade é fortemente aconselhável. Cook (1986) usou esta ideia para motivar sua avaliação da análise de influência. O autor sugere que mais confiança deve ser depositada a um modelo, o qual é relativamente estável sob pequenas modificações. Os esquemas de perturbação mais conhecidos são baseados em influência local, no qual os efeitos são estudados removendo completamente os casos da análise. Este raciocínio irá formar a base para a metodologia de influência local utilizado neste trabalho e ao fazer isso, será possível determinar quais observações poderiam ser influentes na análise.

Outra abordagem sugerida por Cook (1986) que, em vez de remover as observações, dá-se pesos a elas. O cálculo da influência local pode ser realizado para os modelos 5.16 e 5.17. Se a verossimilhança deslocada $LD(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\psi}}) - l(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}})\}$ for usada, em que $\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$ denota EMVs sob o modelo perturbado, a curvatura normal para $\boldsymbol{\psi}$ na direção $\mathbf{d}$, $\|\mathbf{d}\| = 1$ é dada por $C_{\mathbf{d}}(\boldsymbol{\psi}) = 2|\mathbf{d}^T \boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\psi})]^{-1} \boldsymbol{\Delta} \mathbf{d}|$, em que $\boldsymbol{\Delta}$ é uma matriz $(2p+5) \times n$ que depende do esquema de perturbação, dos quais os elementos são dados por $\Delta_{vi} = \partial^2 l(\boldsymbol{\psi}|\boldsymbol{\omega}) / \partial \psi_v \partial \omega_i$, $i = 1, 2, \dots, n$ e $v = 1, 2, \dots, 2p+5$ avaliados em $\hat{\boldsymbol{\psi}}$ e $\boldsymbol{\omega}_0$, em que $\boldsymbol{\omega}_0$ é o vetor não-perturbado (ver COOK, 1986).

As curvaturas normais podem ser calculadas $C_{\mathbf{d}}(\lambda)$, $C_{\mathbf{d}}(\boldsymbol{\theta}_1)$ e $C_{\mathbf{d}}(\boldsymbol{\theta}_2)$ para fazer vários gráficos de índices, por exemplo, o gráfico de índices de $\mathbf{d}_{max}$, o autovetor correspondente a $C_{\mathbf{d}_{max}}$, o maior autovalor da matriz $\mathbf{B} = -\boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\psi})]^{-1} \boldsymbol{\Delta}$ e o gráfico de índices de $C_{\mathbf{d}_i}(\lambda)$, $C_{\mathbf{d}_i}(\boldsymbol{\theta}_1)$ e $C_{\mathbf{d}_i}(\boldsymbol{\theta}_2)$, chamado influência local total (ver, por exemplo, LESAFFRE; VERBEKE, 1998), em que $\mathbf{d}_i$ denota um vetor $n \times 1$ de zeros com 1 na i -ésima

posição. Portanto, a curvatura na direção $\mathbf{d}_i$ assume a forma $C_i = 2|\Delta_i^T [\ddot{\mathbf{L}}(\boldsymbol{\psi})]^{-1} \Delta_i|$, em que Δ_i^T denota a i -ésima linha de Δ . É usual assinalar os casos que $C_i \geq 2\bar{C}$, em que $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$.

A seguir é calculada, para três esquemas de perturbação, a matriz

$$\Delta = \left(\Delta_{vi} \right)_{[(2p+5) \times n]} = \left(\frac{\partial^2 l(\boldsymbol{\psi}|\boldsymbol{\omega})}{\partial \psi_v \partial \omega_i} \right)_{[(2p+5) \times n]},$$

em que $v = 1, 2, \dots, 2p + 5$ e $i = 1, 2, \dots, n$, considerando o modelo definido em (5.15) e seu logaritmo da função de verossimilhança apresentados em (5.16) e (5.17).

5.4.1 Perturbação de casos

Considerando o vetor de pesos $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^T$. Neste caso, o logaritmo da função de verossimilhança perturbada toma a forma

- Modelo de regressão bivariado de Frank-LOLLW

$$\begin{aligned} l(\boldsymbol{\psi}|\boldsymbol{\omega}) &= \sum_{i=1}^n \omega_i (z_{i1} + z_{i2}) + \alpha_1 \sum_{i=1}^n \omega_i \log(u_{i1}) + \alpha_2 \sum_{i=1}^n \omega_i \log(u_{i2}) \\ &+ \lambda \sum_{i=1}^n \omega_i (1 + q_{i1} + q_{i2}) + (\alpha_1 - 1) \sum_{i=1}^n \omega_i \log(1 - u_{i1}) \\ &+ (\alpha_2 - 1) \sum_{i=1}^n \omega_i \log(1 - u_{i2}) - 2 \sum_{i=1}^n \omega_i \log\{\exp(\lambda) - \exp[\lambda(1 + q_{i1})] \\ &+ \exp[\lambda(q_{i1} + q_{i2})] - \exp[\lambda(1 + q_{i2})]\} - 2 \sum_{i=1}^n \omega_i \log[(1 - u_{i1})^{\alpha_1} + u_{i1}^{\alpha_1}] \\ &- 2 \sum_{i=1}^n \omega_i \log[(1 - u_{i2})^{\alpha_2} + u_{i2}^{\alpha_2}]. \end{aligned}$$

- Modelo de regressão bivariado de Clayton-LOLLW

$$\begin{aligned} l(\boldsymbol{\psi}|\boldsymbol{\omega}) &= \sum_{i=1}^n \omega_i (z_{i1} + z_{i2}) + \alpha_1 \sum_{i=1}^n \omega_i \log(u_{i1}) + \alpha_2 \sum_{i=1}^n \omega_i \log(u_{i2}) \\ &+ \alpha_1 \sum_{i=1}^n \omega_i \log(1 - u_{i1}) + \alpha_2 \sum_{i=1}^n \omega_i \log(1 - u_{i2}) - (1 + \lambda) \sum_{i=1}^n \omega_i \log(q_{i1} q_{i2}) \\ &- 2 \sum_{i=1}^n \omega_i \log[(1 - u_{i1})^{\alpha_1} + u_{i1}^{\alpha_1}] - 2 \sum_{i=1}^n \omega_i \log[(1 - u_{i2})^{\alpha_2} + u_{i2}^{\alpha_2}] \\ &- \left(2 + \frac{1}{\lambda} \right) \sum_{i=1}^n \omega_i \log[q_{i1}^\lambda + q_{i2}^\lambda - 1], \end{aligned}$$

em que $z_{ik} = y_{ik} - \mathbf{x}_{ik}^T \boldsymbol{\beta}_k / \sigma_k$, $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$, para $i = 1, \dots, n$ e $k = 1, 2$. Note que q_{i1} , q_{i2} , u_{i1} e u_{i2} são definidos em (5.16) e (5.17). Os elementos da matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\lambda, \boldsymbol{\Delta}_{\boldsymbol{\theta}_1}, \boldsymbol{\Delta}_{\boldsymbol{\theta}_2})^T$ são obtidos numericamente.

5.4.2 Perturbação da variável resposta

Neste caso, considera-se que cada y_{i1} e y_{i2} é perturbada como $y_{i1w} = y_{i1} + \omega_i S_{y1}$ e $y_{i2w} = y_{i2} + \omega_i S_{y2}$, respectivamente, em que S_{y_k} é um fator escala que pode ser estimado pelo desvio padrão de Y_k e $\omega_i \in \mathbf{R}$.

O logaritmo da função de verossimilhança perturbada é expressa como

- Modelo de regressão bivariado de Frank-LOLLW

$$\begin{aligned}
 l(\boldsymbol{\psi}|\boldsymbol{\omega}) = & n \log \left\{ \frac{\lambda \alpha_1 \alpha_2 [\exp(\lambda) - 1]}{\sigma_1 \sigma_2} \right\} + \sum_{i=1}^n (z_{i1}^* + z_{i2}^*) + \alpha_1 \sum_{i=1}^n \log(u_{i1}^*) + \alpha_2 \sum_{i=1}^n \log(u_{i2}^*) \\
 & + \lambda \sum_{i=1}^n (1 + q_{i1}^* + q_{i2}^*) + (\alpha_1 - 1) \sum_{i=1}^n \log(1 - u_{i1}^*) + (\alpha_2 - 1) \sum_{i=1}^n \log(1 - u_{i2}^*) \\
 & - 2 \sum_{i=1}^n \log \{ \exp(\lambda) - \exp[\lambda(1 + q_{i1}^*)] + \exp[\lambda(q_{i1}^* + q_{i2}^*)] - \exp[\lambda(1 + q_{i2}^*)] \} \\
 & - 2 \sum_{i=1}^n \log[(1 - u_{i1}^*)^{\alpha_1} + u_{i1}^{*\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2}^*)^{\alpha_2} + u_{i2}^{*\alpha_2}], \tag{5.18}
 \end{aligned}$$

- Modelo de regressão bivariado de Clayton-LOLLW

$$\begin{aligned}
 l(\boldsymbol{\psi}|\boldsymbol{\omega}) = & n \log \left[\frac{(1 + \lambda) \alpha_1 \alpha_2}{\sigma_1 \sigma_2} \right] + \sum_{i=1}^n (z_{i1}^* + z_{i2}^*) + \alpha_1 \sum_{i=1}^n \log(u_{i1}^*) + \alpha_2 \sum_{i=1}^n \log(u_{i2}^*) \\
 & + \alpha_1 \sum_{i=1}^n \log(1 - u_{i1}^*) + \alpha_2 \sum_{i=1}^n \log(1 - u_{i2}^*) - (1 + \lambda) \sum_{i=1}^n \log(q_{i1}^* q_{i2}^*) \\
 & - 2 \sum_{i=1}^n \log[(1 - u_{i1}^*)^{\alpha_1} + u_{i1}^{*\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2}^*)^{\alpha_2} + u_{i2}^{*\alpha_2}] \\
 & - \left(2 + \frac{1}{\lambda} \right) \sum_{i=1}^n \log[q_{i1}^{*\lambda} + q_{i2}^{*\lambda} - 1], \tag{5.19}
 \end{aligned}$$

em que $q_{ik}^* = \frac{(1 - u_{ik}^*)^{\alpha_k}}{(1 - u_{ik}^*)^{\alpha_k} + u_{ik}^{*\alpha_k}}$, $u_{ik}^* = \exp[-\exp(z_{ik}^*)]$, $z_{i1}^* = (y_{i1}^* - \mathbf{x}_i^T \boldsymbol{\beta}_1) / \sigma_1$, $z_{i2}^* = (y_{i2}^* - \mathbf{x}_i^T \boldsymbol{\beta}_2) / \sigma_2$, $y_{i1}^* = y_{i1} + \omega_i S_{y1}$, $y_{i2}^* = y_{i2} + \omega_i S_{y2}$, para $i = 1, \dots, n$ e $k = 1, 2$. Mais uma vez os elementos da matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\lambda, \boldsymbol{\Delta}_{\boldsymbol{\theta}_1}, \boldsymbol{\Delta}_{\boldsymbol{\theta}_2})^T$ são obtidos numericamente.

5.4.3 Perturbação de uma variável explanatória

Considerando agora uma perturbação aditiva, sob uma particular variável explanatória contínua, normalmente X_t , fazendo $x_{it\omega} = x_{it} + \omega_i S_x$, em que S_x é um fator escala, $\omega_i \in \mathbf{R}$. Este esquema de perturbação leva as seguintes expressões para o logaritmo da função de verossimilhança

- Modelo de regressão bivariado de Frank-LOLLW

$$\begin{aligned}
 l(\psi|\omega) = & n \log \left\{ \frac{\lambda \alpha_1 \alpha_2 [\exp(\lambda) - 1]}{\sigma_1 \sigma_2} \right\} + \sum_{i=1}^n (z_{i1}^\dagger + z_{i2}^\dagger) + \alpha_1 \sum_{i=1}^n \log(u_{i1}^\dagger) + \alpha_2 \sum_{i=1}^n \log(u_{i2}^\dagger) \\
 & + \lambda \sum_{i=1}^n (1 + q_{i1}^\dagger + q_{i2}^\dagger) + (\alpha_1 - 1) \sum_{i=1}^n \log(1 - u_{i1}^\dagger) + (\alpha_2 - 1) \sum_{i=1}^n \log(1 - u_{i2}^\dagger) \\
 & - 2 \sum_{i=1}^n \log \{ \exp(\lambda) - \exp[\lambda(1 + q_{i1}^\dagger)] + \exp[\lambda(q_{i1}^\dagger + q_{i2}^\dagger)] - \exp[\lambda(1 + q_{i2}^\dagger)] \} \\
 & - 2 \sum_{i=1}^n \log[(1 - u_{i1}^\dagger)^{\alpha_1} + u_{i1}^{\dagger\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2}^\dagger)^{\alpha_2} + u_{i2}^{\dagger\alpha_2}],
 \end{aligned}$$

- Modelo de regressão bivariado de Clayton-LOLLW

$$\begin{aligned}
 l(\psi|\omega) = & n \log \left[\frac{(1 + \lambda) \alpha_1 \alpha_2}{\sigma_1 \sigma_2} \right] + \sum_{i=1}^n (z_{i1}^\dagger + z_{i2}^\dagger) + \alpha_1 \sum_{i=1}^n \log(u_{i1}^\dagger) + \alpha_2 \sum_{i=1}^n \log(u_{i2}^\dagger) \\
 & + \alpha_1 \sum_{i=1}^n \log(1 - u_{i1}^\dagger) + \alpha_2 \sum_{i=1}^n \log(1 - u_{i2}^\dagger) - (1 + \lambda) \sum_{i=1}^n \log(q_{i1}^\dagger q_{i2}^\dagger) \\
 & - 2 \sum_{i=1}^n \log[(1 - u_{i1}^\dagger)^{\alpha_1} + u_{i1}^{\dagger\alpha_1}] - 2 \sum_{i=1}^n \log[(1 - u_{i2}^\dagger)^{\alpha_2} + u_{i2}^{\dagger\alpha_2}] \\
 & - \left(2 + \frac{1}{\lambda} \right) \sum_{i=1}^n \log[q_{i1}^{\dagger\lambda} + q_{i2}^{\dagger\lambda} - 1],
 \end{aligned}$$

em que $q_{ik}^\dagger = \frac{(1 - u_{ik}^\dagger)^{\alpha_k}}{(1 - u_{ik}^\dagger)^{\alpha_k} + u_{ik}^{\dagger\alpha_k}}$, $u_{ik}^\dagger = \exp[-\exp(z_{ik}^\dagger)]$, $z_{ik}^\dagger = \{y_{ik} - \mathbf{x}_i^{\dagger\top} \beta_k\} / \sigma_k$, $\mathbf{x}_i^{\dagger\top} \beta_k = \beta_{0k} + \beta_{1k}x_{i1} + \beta_{2k}x_{i2} + \dots + \beta_{tk}[x_{it} + \omega_i S_x] + \dots + \beta_{pk}x_{ip}$ $i = 1, \dots, n$ e $k = 1, 2$. Os elementos da matriz $\Delta = (\Delta_\lambda, \Delta_{\theta_1}, \Delta_{\theta_2})^T$ são obtidos numericamente.

5.5 Aplicação: dados de qualidade de vida relacionada à saúde bucal

O presente estudo tem como objetivos avaliar as mudanças na qualidade de vida relacionada a saúde bucal (OHRQoL) de alunos em um exame de acompanhamento de três anos e avaliar o impacto da incidência de cárie na OHRQoL de adolescentes. Em 2009, uma amostra inicial de 515 adolescentes, representativa da população de 12 anos

de idade da cidade de Juiz de Fora em Minas Gerais/Brasil, foi avaliada (Paula et al., 2012). A amostral final, reavaliada 3 anos após o exame baseline, foi composta de 291 adolescentes e, representava uma taxa de 56,5% da amostra inicial. Destes, 150 (51,5%) eram do sexo feminino e 238 (81,8%) estudavam em escolas públicas.

A qualidade de vida relacionada a saúde bucal (OHRQoL), como o nome indica, envolve aspectos da qualidade de vida relacionada a boca. A variável resposta foi a avaliação da qualidade de vida relacionada a saúde bucal de crianças (OHRQoL). Os dados foram obtidos de um questionário chamado de CPQ (Children Perceptions Questionnaire), administrado a crianças entre as idades de 11 e 14 anos (*CPQ11 – 14*). O mesmo questionário foi aplicado duas vezes para a mesma criança: primeiro para estabelecer a baseline (início do estudo) e então 3 anos mais tarde. Portanto, a variável resposta associada a cada criança é bivariada. A (*CPQ11 – 14*) determina a qualidade de vida relacionada a saúde bucal e é composta de 36 itens agrupados em 4 domínios da saúde

- sintomas orais, composto de 6 questões;
- limitações funcionais, composto de 9 questões;
- bem-estar emocional, composto de 9 questões;
- bem-estar social, composto de 12 questões.

Em cada item, há questões a respeito da frequência de eventos envolvendo os dentes, boca e mandíbula nos três últimos anos. As respostas são dadas em uma escala Likert de 1 a 4: “nenhum” = 0; “uma ou duas vezes” = 1; “às vezes” = 2; “frequentemente” = 3; e “muito frequentemente” = 4. Isto permite obter um escore global adicionando o número de cada item. Escores altos na (*CPQ11 – 14*) indicam piora na qualidade de vida relacionada à saúde bucal, pois os indivíduos quando dizem “frequentemente” ou “muito frequentemente”, por exemplo, serão mais propensos a terem dores de dentes ou outros problemas dentários. Outro índice usado no estudo foi o DMFT, o qual reflete a incidência de cáries em dentes permanentes das crianças. A sigla DFMT significa Decayed, Missing or Filled Teeth.

Nessa análise, foram considerados o modelo de regressão bivariado por meio das cópulas de Clayton e Frank, a distribuição LOLLW e seus correspondentes submodelos como distribuições marginais. As variáveis utilizadas foram

- t_{i1} = OHRQoL(0): Escore geral da OHRQL na baseline (medida no início do estudo);
- t_{i2} =OHRQoL(3): Escore geral da OHRQL no tempo follow up (medida três anos depois);

- x_{i1} – incidência do componente D do DMFT: Se houve um aumento dos dentes cariados durante os três anos (0= aumentou 1= não mudou, 2= diminuiu);
- x_{i2} – incidência do componente M do DMFT: Se houve um aumento de perda dos dentes durante os três anos (0= aumentou, 1= não mudou);
- x_{i3} – incidência do componente F do DMFT: Se houve um aumento do número de restaurações durante os três anos (0= menor número de dentes obturados - provavelmente devido a extração, 1= não mudou, 2= maior número de dentes com obturações);
- x_{i4} – escore DMFT total: soma dos escores dos três elementos mencionados acima - formando o número completo de dentes permanentes cariados, perdidos e obturados. Portanto, neste caso, foi dicotomizado se houve ou não um aumento no índice DMFT durante os três anos, subtraindo o escore final - escore inicial (0= aumentou, 1=não mudou);
- x_{i5} – incidência de sangramento nas gengivas: a presença de sangramento nas gengivas foi medido no início e após três anos: 0 = sem sangramento gengival em 2009 e sangramento em 2012 (novos casos), 1 = sangramento gengival continuou, 2= sangramento gengival em 2009 e nenhum sangramento em 2012 (cura);
- x_{i6} – necessidade de tratamento odontológico: avalia se houve algum aumento da necessidade de tratamento odontológico da criança durante os três anos, medida pelo Índice de Estética Dentária (DAI - Dental Aesthetic Index). Este índice mede a aceitabilidade social da aparência dentária por meio da avaliação de 10 características oclusais (0= novos casos/necessidade de tratamento dentário em 2009, mas tal necessidade apareceu durante os três anos, 1= nenhuma alteração das características, 2= necessidade de tratamento dentário em 2009 mas não em 2012);
- x_{i7} – uso de aparelho ou fixadores: avalia se a criança recebeu tratamento odontológico (fixador e/ou aparelho) alguma vez durante o período de estudo (1= não, 2= sim);
- x_{i8} – aglomeração domiciliar: mede a mudança na proporção do número de pessoas na casa da criança com o número de quartos em 2009 e 2012 (0= aumentou, 1= permaneceu inalterado, 2= diminuiu);
- x_{i9} – questionário geral de qualidade de vida: refere a diferença na OHRQoL entre 2009 e 2012 (0= piorou, 1=permaneceu inalterado, 2= melhorou).
- x_{i10} – visita ao dentista: mede se a criança visitou um dentista nos últimos três anos (1=não, 2=sim);

- x_{i11} – gênero (1=feminino, 2=masculino).

em que $i = 1, \dots, 291$.

A seguir são apresentadas algumas análises descritivas da variável resposta bidimensional.

A Tabela 5.1 apresenta um resumo descritivo para estes dados mostrando diferentes graus de assimetria e curtose.

Tabela 5.1 - Estatísticas descritivas para o conjunto de dados DENTAL

Variável	Média	Med.	Moda	D.P.	Var.	Ass.	Curt.	Mín.	Max.
OHRQoL(0)	25,471	18,0	4,0	23,421	548,547	1,224	0,872	0,10	106,0
OHRQoL(3)	20,896	15,0	0,1	19,712	388,588	1,340	1,431	0,10	90,0

Na Figura 5.3 (a) são apresentados os gráficos da distribuição bivariada. O gráfico de contorno (com curvas de nível) fornece informações, ao mesmo tempo, quanto a normalidade e correlação das variáveis. A Figura 5.3 (b) mostra uma correlação positiva entre as variáveis OHRQoL(0) e OHRQoL(3), pois as linhas de contorno (curvas de nível) estão posicionadas em torno da diagonal principal.

Figura 5.3 - (a) Distribuição bivariada. (b) Correlação positiva entre as variáveis OHRQoL(0) e OHRQoL(3)

O coeficiente de correlação Tau de Kendall calculado para o baseline e follow-up foi $\tau = 0,49$ (valor-p $< 0,001$), indicando uma associação positiva entre as variáveis, como também pode ser observado no gráfico de dispersão (Figura (5.4) (b)).

Dessa forma, nesse trabalho foi usado o método de cópulas para avaliar as mudanças

Figura 5.4 - (a) QQplot. (b) Gráfico de dispersão OHRQoL(0) e OHRQoL(3)

dos escores no baseline e follow-up de acordo com as mudanças nos perfis ambientais e clínicos durante o período do estudo e a dependência entre esses escores.

Para explicar o tempo até a ocorrência ou evento distinto em um único indivíduo, regressões bidimensionais obtidas por meio de cópulas são usadas. As cópulas de Frank e de Clayton, apresentadas na Seção 5.3, são usadas na modelagem.

O conjunto de dados DENTAL foi analisado considerando as variáveis aleatórias $Y_{i1} = \log(T_{i1})$ e $Y_{i2} = \log(T_{i2})$, as quais estão relacionadas a variáveis explanatórias por meio de um modelo log-linear

$$\begin{aligned}
 y_{ik} = & \beta_{0k} + \beta_{11k}D_{i11} + \beta_{12k}D_{i12} + \beta_{2k}x_{i2} + \beta_{31k}D_{i31} + \beta_{32k}D_{i32} + \beta_{4k}x_{i4} + \beta_{51k}D_{i51} \\
 & + \beta_{52k}D_{i52} + \beta_{61k}D_{i61} + \beta_{62k}D_{i62} + \beta_{7k}x_{i7} + \beta_{81k}D_{i81} + \beta_{82k}D_{i82} + \beta_{91k}D_{i91} \\
 & + \beta_{92k}D_{i92} + \beta_{10k}x_{i10} + \beta_{1k}x_{i11} + \sigma_k z_{ik}, \quad i = 1, \dots, 291, \quad k = 1, 2, \quad (5.20)
 \end{aligned}$$

em que z_1 e z_2 são os erros aleatórios do modelo com distribuições conjuntas dadas nas equações (5.13) e (5.14).

Note que as variáveis (D_{11}, D_{12}) , (D_{31}, D_{32}) , (D_{51}, D_{52}) , (D_{61}, D_{62}) , (D_{81}, D_{82}) , (D_{91}, D_{92}) são as variáveis dummies correspondentes as variáveis x_1, x_3, x_5, x_6, x_8 e x_9 , respectivamente.

As EMVs dos parâmetros do modelo foram obtidas usando a subrotina MaxBFGS do programa de linguagem matricial Ox (ver DOORNIK, 2007). Os valores das estatísticas deviance global (DG), AIC (Akaike Information Criterion), CAIC (Consistent Akaike Information Criterion) e BIC (Bayesian Information Criterion) são listados nas Tabelas 5.2 e 5.3. Os menores valores dos critérios de informação correspondem ao modelo de

regressão bivariada de Frank-LOLLW e Clayton-LOLLW, o que proporciona um melhor ajuste para os dados atuais do que os outros modelos.

Tabela 5.2 - DG, critérios AIC, BIC e CAIC para o modelo de regressão bivariado de Clayton-LOLLW e submodelos aplicados ao conjunto de dados DENTAL

Modelos de regressão bivariado $Y_{i1} \times Y_{i2}$	DG	AIC	CAIC	BIC
LOLLW $\times$ LOLW	1.799,8207	1.881,8207	1.895,6520	2.032,4269
LOLLW $\times$ LOLLE	1.818,9485	1.898,9485	1.912,0685	2.045,8814
LOLLW $\times$ LW	1.825,1329	1.905,1329	1.918,2529	2.052,0658
LOLLW $\times$ LE	1.827,5811	1.905,5811	1.918,0114	2.048,8407
LOLLE $\times$ LOLW	1.811,3240	1.891,3240	1.904,4440	2.038,2569
LW $\times$ LOLW	1.809,4347	1.889,4347	1.902,5547	2.036,3677
LE $\times$ LOLW	1.812,0360	1.890,0360	1.902,4663	2.033,2956
LW $\times$ LW	1.831,4803	1.909,4803	1.921,9106	2.052,7399
LW $\times$ LE	1.833,1898	1.909,1898	1.920,9517	2.048,7761
LE $\times$ LW	1.837,7974	1.913,7974	1.925,5593	2.053,3837
LE $\times$ LE	1.842,0951	1.916,0951	1.927,2097	2.052,0081

Tabela 5.3 - DG, critérios AIC, BIC e CAIC para o modelo de regressão bivariado Frank-LOLLW e submodelos aplicados ao conjunto de dados DENTAL

Modelos de regressão bivariado $Y_{i1} \times Y_{i2}$	DG	AIC	CAIC	BIC
LOLLW $\times$ LOLW	1.685,4972	1.767,4972	1.781,3285	1.918,1035
LOLLW $\times$ LOLLE	1.710,6342	1.790,6342	1.803,7542	1.937,5671
LOLLW $\times$ LW	1.714,4440	1.794,4440	1.807,5640	1.941,3769
LOLLW $\times$ LE	1.714,9614	1.792,9614	1.805,3917	1.936,2210
LOLLE $\times$ LOLW	1.704,0638	1.784,0638	1.797,1838	1.930,9967
LW $\times$ LOLW	1.702,3472	1.782,3472	1.795,4672	1.929,2801
LE $\times$ LOLW	1.704,2458	1.782,2458	1.794,6761	1.925,5054
LW $\times$ LW	1.726,8332	1.804,8332	1.817,2635	1.948,0928
LW $\times$ LE	1.727,4266	1.803,4266	1.815,1885	1.943,0129
LE $\times$ LW	1.730,0146	1.806,0146	1.817,7765	1.945,6009
LE $\times$ LE	1.731,9501	1.805,9501	1.817,0647	1.941,8631

Pelas Tabelas 5.2 e 5.3, observa-se que, tanto para a cópula de Clayton como para a cópula de Frank, os modelos que ajustam melhor aos dados, analisando as menores estatísticas AIC, CAIC e BIC, são os que consideram em ambas as componentes a dis-

tribuição LOLLW. Assim, para decidir qual das cópulas será utilizada nas análises posteriores, realizou-se o Teste de Razão de Verossimilhança Generalizada (TRVG) (para discriminar entre os modelos não-encaixados, ver CAMERON; TRIVEDI, 1998). A estatística TRVG foi $TRVG = 10.4664$. Uma vez que $TRVG = 10.4664 > 1,96$, rejeita-se a hipótese nula (ao nível de 5% de significância) de equivalência entre os modelos (atribuindo a distribuição LOLLW para ambas as componentes da cópula) considerando a cópula de Frank e de Clayton e verifica-se, portanto, que o modelo que se ajusta melhor aos dados é o modelo considerando a cópula de Frank.

As EMVs para os parâmetros do modelo de regressão bivariado Frank-LOLLW são dadas na Tabela 5.4.

Tabela 5.4 - EMVs para o modelo de regressão bivariado Frank-LOLLW aplicado ao conjunto de dados de qualidade de vida relacionada à saúde bucal

Parâmetro	Estimativa	E.P.	valor- p	Parâmetro	Estimativa	E.P.	valor- p
λ	6,5401	0,5822		β_{122}	0,1608	0,2993	0,5911
β_{01}	2,8875	0,4801	$< 0,01^{**}$	β_{22}	-0,1228	0,3021	0,6843
β_{111}	0,0882	0,1589	0,5789	β_{312}	-0,2325	0,2155	0,2806
β_{121}	0,1117	0,279	0,6888	β_{322}	-0,2052	0,3214	0,5233
β_{21}	0,2284	0,2791	0,4133	β_{42}	-0,1673	0,1836	0,3622
β_{311}	-0,145	0,1904	0,4463	β_{512}	-0,127	0,1466	0,3863
β_{321}	0,0117	0,2783	0,9664	β_{522}	0,3009	0,2392	0,2084
β_{41}	0,1433	0,1652	0,3855	β_{612}	0,1642	0,3793	0,6651
β_{511}	-0,2423	0,1288	0,0599	β_{622}	0,4168	0,363	0,2509
β_{521}	0,4393	0,2053	0,0323 **	β_{72}	0,13	0,1716	0,4488
β_{611}	0,1012	0,3484	0,7714	β_{812}	0,0469	0,1457	0,7473
β_{621}	0,4567	0,3381	0,1768	β_{822}	-0,0366	0,1681	0,8277
β_{71}	0,1086	0,1485	0,4648	β_{912}	0,0019	0,1442	0,9897
β_{811}	0,1763	0,1249	0,1579	β_{922}	0,2005	0,1013	0,0477 **
β_{821}	0,4036	0,1468	0,0060 **	β_{102}	-0,3069	0,1289	0,0173 **
β_{911}	0,0873	0,1271	0,4921	β_{12}	-0,0504	0,0965	0,6016
β_{921}	0,2182	0,0917	0,0173 **	σ_1	0,5178	0,0644	
β_{101}	-0,3948	0,1078	0,0003 **	σ_2	0,5452	0,054	
β_{11}	-0,0575	0,0844	0,496	α_1	0,4779	0,0723	
β_{02}	3,4165	0,498	$< 0,01^{**}$	α_2	0,4377	0,0557	
β_{112}	-0,0502	0,1907	0,7924				

** significativo a um nível de 5%.

5.6 Análise de sensibilidade

A teoria de influência local, desenvolvida na Seção 5.4, é utilizada aqui. Obteve-se, considerando a perturbação de casos, o valor de máxima curvatura $C_{\mathbf{d}_{max}} = 2,8461$. A Figura 5.5(a), gráfico do autovetor correspondente a $\mathbf{d}_{max}$, mostra que as observações #61 e #86 são bem distintas com relação as outras.

A influência das perturbações sobre os tempos será analisado agora (esquema de perturbação da variável resposta). O valor da curvatura máxima é $C_{\mathbf{d}_{max}} = 9,3510$. A Figura 5.5(b), gráfico de $\mathbf{d}_{max}$ versus índices das observações, indica que não há nenhuma observação discordante das demais.

A influência local total C_i é apresentada nas Figuras 5.6(a) e 5.6(b), para esquemas de perturbação de casos e da variável resposta, respectivamente. As observações #55 e #86 são bem distintas com relação as demais.

Desta forma, a partir da análise de diagnóstico (influência local), #55, #61 e #86 foram detectadas como possíveis observações influentes. Primeiramente, cada uma dessas observações é eliminada individualmente. Em seguida, são removidas do conjunto “A” (conjunto dos dados originais) a totalidade das observações possivelmente influentes.

A Tabela 5.6 fornece as mudanças relativas (em porcentagens), de cada estimativa, definida por $\mathbf{RC}_{\theta_j} = \left[(\hat{\theta}_j - \hat{\theta}_{j(i)}) / \hat{\theta}_j \right] \times 100$, e os correspondentes valores- p , em que $\hat{\theta}_j(I)$ é o EMV de θ_j após o conjunto “I” de observações ter sido removido. A Tabela 5.6 fornece os seguintes conjuntos: $C_1 = \{\#55\}$, $C_2 = \{\#61\}$, $C_3 = \{\#86\}$, $C_4 = \{\#55, \#61\}$, $C_5 = \{\#55, \#86\}$, $C_6 = \{\#61, \#86\}$ e $C_7 = \{\#55, \#61, \#86\}$.

Figura 5.5 - Gráficos de influência local $|\mathbf{d}_{max}|$ do modelo de regressão considerando a Cópula de Frank aplicado ao conjunto de dados de saúde bucal relacionada a qualidade de vida: (a) perturbação de casos e (b) perturbação da resposta

Figura 5.6 - Gráficos de influência local total $|C_i|$ do modelo de regressão considerando a Cópula de Frank aplicado ao conjunto de dados de saúde bucal relacionada a qualidade de vida: (a) perturbação de casos e (b) perturbação da resposta

A Tabela 5.6 (ver Apêndice C) mostra que as observações #55, #61 e #86 são diagnosticadas como observações influentes, uma vez que causam mudanças significativas em algumas das variáveis explanatórias (D_{51} , D_{52} , X_{10}). A observação 55 refere-se ao escolar com tempos OHRQoL(0) e OHRQoL(3) iguais a, respectivamente, 92 e 87, que: mantiveram os índices C e P do CPOD, restaurou e não aumentou o índice de cáries, aumentou o índice CPOD, teve sangramento gengival tratado, manteve o tratamento ortodôntico, não usou aparelho, manteve a aglomeração por domicílio, piorou a qualidade de vida, foi ao dentista nos últimos 3 anos e do gênero feminino. A observação 61 refere-se ao escolar com tempos OHRQoL(0) e OHRQoL(3) iguais a, respectivamente, 66 e 69, que: manteve os índices C, P e O do CPOD, manteve o índice CPOD, manteve a presença de sangramento gengival, manteve o tratamento ortodôntico, não usou aparelho, manteve a aglomeração por domicílio, piorou a qualidade de vida, não foi ao dentista nos últimos 3 anos e do gênero masculino. Já a observação 86 refere-se ao escolar com tempos OHRQoL(0) e OHRQoL(3) iguais a, respectivamente, 73 e 67, que: manteve os índices C, P e O do CPOD, manteve o índice CPOD, manteve a presença de sangramento gengival, manteve o tratamento ortodôntico, não usou aparelho, manteve a aglomeração por domicílio, manteve a qualidade de vida, não foi ao dentista nos últimos 3 anos e do gênero masculino.

Portanto, o modelo ajustado após a remoção de observações 55, 61 e 86 é apresentado na Tabela 5.5. Esta remoção também foi sugerida pelo pesquisador responsável pelo estudo.

Tabela 5.5 - EMVs para o modelo de regressão bivariado Frank-LOLLW aplicado ao conjunto de dados DENTAL após a análise de influência

Parâmetro	Estimativa	E.P.	Valor- <i>p</i>	Parâmetro	Estimativa	E.P.	Valor- <i>p</i>
λ	6,2470	0,5690		D_{12}	0,2176	0,2930	0,4578
β_{01}	2,8938	0,5000	$< 0,01^{**}$	β_{22}	0,0184	0,2831	0,948
D_{11}	0,0447	0,1636	0,7846	D_{31}	-0,2316	0,2165	0,2847
D_{12}	0,1323	0,2865	0,6442	D_{32}	-0,2537	0,3299	0,4418
β_{21}	0,3324	0,2844	0,2424	β_{42}	-0,1568	0,1835	0,3928
D_{31}	-0,1455	0,1976	0,4615	D_{51}	-0,1673	0,1409	0,2351
D_{32}	-0,0140	0,2893	0,9614	D_{52}	0,1424	0,2510	0,5706
β_{41}	0,1558	0,1689	0,3563	D_{61}	0,0192	0,3662	0,9583
D_{51}	-0,2733	0,1314	0,0375 ^{**}	D_{62}	0,3214	0,3516	0,3607
D_{52}	0,3427	0,2268	0,1308	β_{72}	0,1092	0,1680	0,5159
D_{61}	-0,0031	0,3617	0,9931	D_{81}	0,0204	0,1432	0,8868
D_{62}	0,4022	0,3546	0,2567	D_{82}	-0,0204	0,1619	0,8998
β_{71}	0,0886	0,1482	0,5499	D_{91}	-0,0732	0,1433	0,6095
D_{81}	0,1540	0,1284	0,2304	D_{92}	0,2876	0,1005	0,0042 ^{**}
D_{82}	0,4016	0,1481	0,0067 ^{**}	β_{102}	-0,2534	0,1321	0,0551
D_{91}	0,0255	0,1321	0,8472	β_{112}	-0,0935	0,1021	0,3596
D_{92}	0,2838	0,0959	0,0031 ^{**}	σ_1	0,5484	0,0692	
β_{101}	-0,347	0,1136	0,0023 ^{**}	σ_2	0,5293	0,0522	
β_{111}	-0,0988	0,0925	0,2852	α_1	0,5163	0,0803	
β_{02}	3,4341	0,4814	$< 0,01^{**}$	α_2	0,4248	0,0539	
D_{11}	-0,1077	0,1886	0,5681				

^{**} significativo a um nível de 5%.

Por exemplo, a covariável D_{82} (aglomeração domiciliar) é significativa no primeiro tempo, isso demonstra que existe uma diferença significativa entre os níveis (0 = aumentada) e (2 = Diminuição). Todas estas variáveis significativas são mostradas na Figura 5.7. Mais informação é fornecida por uma comparação visual do histograma dos dados com as funções de densidade estimadas, que são apresentadas na Figura 5.8. Concluímos também que a distribuição LOLLW fornece uma adaptação adequada a esses dados e, em seguida, o modelo de regressão bivariada Frank-LOLLW é mais apropriado para a modelagem deste conjunto de dados.

Figura 5.7 - Gráfico das variáveis significativas

Figura 5.8 - Ajuste da densidade LOLLW para o conjunto de dados de saúde bucal relacionada a qualidade de vida: (a) para a variável OHRQoL(0) e (b) para a variável OHRQoL(3)

5.7 Considerações Finais

Neste trabalho foi considerada uma abordagem paramétrica para dados bivariados usando modelos marginais de locação-escala com distribuições log-odd log-logística Weibull. A estrutura de dependência entre os eventos foi modelada usando a cópula de Frank, pertencente a família Arquimediana. Os modelos, técnicas de estimação e métodos de influência local foram apresentados. Essas técnicas demonstraram serem ferramentas importantes para avaliar a robustez dos modelos de regressão com resposta bivariada. A metodologia de influência local foi precisa para detectar possíveis observações discrepantes. Para ilustrar o modelo proposto e algumas metodologias de análise de sensibilidade, um conjunto de dados reais que diz respeito a saúde bucal relacionada a qualidade de vida de pacientes infantis foi utilizado.

Referências

- CHATTERJEE, N.; SHIH, J.A bivariate cure-mixture approach for modeling familial association in diseases. **Biometrics**, Washington, v.57, p.779-786, 2001.
- CRUZ, J.N.; ORTEGA, E.M.M; CORDEIRO, G.M. The log-odd log-Logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. **Journal of Statistical Computation and Simulation**, London, p.1-23, 2015.
- CLAYTON, D.G. A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence. **Biometrika**, Cambridge, v.65, p.141-151, 1978.
- COOK, R.D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society B**, London, v.48, p.133-169, 1986
- ESCOBAR, L.A.; MEEKER, W.Q. Assessing influence in regression analysis with censored data. **Biometrics**, Washington, v.48, p.507-528, 1992.
- EUGENE, N.; LEE, C.; FAMOYE, F. Beta-Normal distribution and its applications. **Communications in Statistics - Theory and Methods**, London, v.31, p.497-512, 2002.
- FACHINI, J.B.; ORTEGA, E.M.M.; CORDEIRO, G.M. A bivariate regression model with cure fraction. **Journal of Statistical Computation and Simulation**, New York, v.1, p.1-16, 2013.
- FRANK, M.J. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x, y)$. **Aequationes Mathematicae**, Basel, v.19, p.194-226, 1979.

GENEST, C. Frank's family of bivariate distributions. **Biometrika**, Cambridge, v.74, p.549-555, 1987.

HASHIMOTO, E.M.; ORTEGA, E.M.M; CANCHO, V.G.; CORDEIRO, G.M. On estimation and diagnostics analysis in log-generalized Gamma regression model for interval-censored data. **Statistics**, Berlin, v. 47, p.379-398, 2013.

HE, W.; LAWLESS, J.F. Bivariate location-scale models for regression analysis, with applications to lifetime data. **Journal of the Royal Statistical Society**, London, v.67, 63-78, 2005.

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

NELSEN, R.B. **An introduction to copulas**. 2.ed. New York: Springer-Verlag, 2006. 270p.

Núñez, J.S.R. **Modelagem Bayesiana para dados de sobrevivência bivariados através de cópulas**. 2005. 101p. Tese (Doutorado em Estatística)- Instituto de Matemática e Estatística, IME/USP, São Paulo, 2005.

ORTEGA, E.M.M.; CORDEIRO, G.M.; KATTAN, M.W. The log-Beta Weibull regression model with application to predict recurrence of prostate cancer. **Statistical Papers**, New York, v.54, p.113-132, 2013.

PAULA, J.S.; OLIVEIRA, M.; SOARES, M.S.P.; CHAVES, M.G.A.M.; MIALHE, F.L. Perfil epidemiológico dos pacientes atendidos no pronto atendimento da Faculdade de Odontologia da Universidade Federal de Juiz de Fora. **Arquivos em Odontologia (UFMG)**, Belo Horizonte, v. 48, p.257-262, 2012.

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

SILVA, G.O.; ORTEGA, E.M.M.; CORDEIRO, G.M. The Beta modified Weibull distribution. **Lifetime Data Analysis**, London, v. 16, p.409-430, 2010.

APÊNDICES

APÊNDICE A - Séries de potência úteis (Capítulo 2)

Aqui são apresentadas três expansões em séries de potência necessárias para a prova do resultado geral na Seção 2.5. Primeiro, para $a > 0$ real, tem-se a expansão binomial

$$(1 - u)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} u^j, \quad (5.21)$$

em que o coeficiente binomial é definido para todo real.

Segundo, expandindo z^λ em série de Taylor, pode-se escrever

$$z^\lambda = \sum_{k=0}^{\infty} (\lambda)_k (z - 1)^k / k! = \sum_{i=0}^{\infty} f_i z^i, \quad (5.22)$$

$$f_i = f_i(\lambda) = \sum_{k=i}^{\infty} \frac{(-1)^{k-i}}{k!} \binom{k}{i} (\lambda)_k,$$

e $(\lambda)_k = \lambda(\lambda - 1) \dots (\lambda - k + 1)$ é o fatorial decrescente definido na Seção 2.6.

Terceiro, usando as equações (2.12) e (2.13), obtém-se uma expressão para $[G(x)^a + \bar{G}(x)^a]^c$. Pode-se escrever a partir de (2.7) e (5.21)

$$[G(x)^a + \bar{G}(x)^a] = \sum_{j=0}^{\infty} t_j G(x)^j,$$

em que $t_j = t_j(a) = a_j(a) + (-1)^j \binom{a}{j}$ e $a_j(a)$ é como definido em (2.7). Então, usando (5.22), temos

$$[G(x)^a + \bar{G}(x)^a]^c = \sum_{i=0}^{\infty} f_i \left(\sum_{j=0}^{\infty} t_j G(x)^j \right)^i,$$

em que $f_i = f_i(c)$. Finalmente, usando novamente as equações (2.12) e (2.13), obtém-se

$$[G(x)^a + \bar{G}(x)^a]^c = \sum_{j=0}^{\infty} h_j(a, c) G(x)^j, \quad (5.23)$$

em que $h_j(a, c) = \sum_{i=0}^{\infty} f_i m_{i,j}$ e (para $i \geq 0$) $m_{i,j} = (j t_0)^{-1} \sum_{m=1}^j [m(j+1) - j] t_m m_{i,j-m}$ (para $j \geq 1$) e $m_{i,0} = t_0^i$.

APÊNDICE B - Componentes do vetor escore (Capítulo 4)

As componentes do vetor escore $U(\boldsymbol{\theta})$ são dadas por

$$\begin{aligned}
U_{\alpha^{(T^*)}}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{\delta_i \dot{h}_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})}{h_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})} - \sum_{i=1}^n \frac{(\phi+1) \dot{H}_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})}{\phi + H_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}, \\
U_{\sigma^{(T^*)}}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{\delta_i \dot{h}_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})}{h_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})} - \sum_{i=1}^n \frac{(\phi+1) \dot{H}_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa})}{\phi + H_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}, \\
U_{\alpha^{(C^*)}}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{(1-\delta_i) \dot{h}_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{h_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})} - \sum_{i=1}^n \frac{(\phi+1) \dot{H}_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{\phi + H_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}, \\
U_{\sigma^{(C^*)}}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{(1-\delta_i) \dot{h}_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{h_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})} - \sum_{i=1}^n \frac{(\phi+1) \dot{H}_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{\phi + H_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}, \\
U_{\phi}(\boldsymbol{\theta}) &= r \left[\log(\phi) + 1 + \frac{\Gamma'(\phi+1)}{\Gamma(\phi)} - \frac{\Gamma'(\phi)}{\Gamma(\phi)} \right] - \sum_{i=1}^n \frac{1}{\left[\phi + H_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) \right]}, \\
U_{\beta_j}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{\delta_i \dot{h}_{(\boldsymbol{\beta})}^{(T^*)}(y_i|\boldsymbol{\kappa})}{h_{(\boldsymbol{\beta})}^{(T^*)}(y_i|\boldsymbol{\kappa})} - \sum_{i=1}^n \frac{(\phi+1) \left[\dot{H}_{0(\boldsymbol{\beta})}^{(T^*)}(y_i|\boldsymbol{\kappa}) \right]}{\phi + H_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}, \\
U_{\mu^{(C^*)}}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{(1-\delta_i) \dot{h}_{0(\mu^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{h_{0(\mu^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})} - \sum_{i=1}^n \frac{(\phi+1) \dot{H}_{0(\mu^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})}{\phi + H_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) + H_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu})},
\end{aligned}$$

em que

$$\begin{aligned}
\dot{h}_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) &= \frac{\exp\left(z_i^{(T^*)}\right) \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1} \left\{ \left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \left[1 + \alpha^{(T^*)} \log\left(\frac{1-u_i^{(T^*)}}{u_i^{(T^*)}}\right) \right] \right.}{\sigma^{(T^*)} \left[\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \right]^2} \\
&\quad \left. \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \right\} \\
&\quad + \frac{\left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}}}{\sigma^{(T^*)} \left[\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \right]^2},
\end{aligned}$$

$$\begin{aligned} \dot{h}_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) &= \frac{\exp\left(z_i^{(C^*)}\right)\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}\left\{\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\left[1+\alpha^{(C^*)}\log\left(\frac{1-u_i^{(C^*)}}{u_i^{(C^*)}}\right)\right]\right.}{\sigma^{(C^*)}\left[\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}+\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\right]^2} \\ &\quad \left.+\frac{\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}}{\sigma^{(C^*)}\left[\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}+\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\right]^2}, \end{aligned}$$

$$\dot{H}_{0(\alpha^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) = \frac{\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}}\log\left(\frac{1-u_i^{(T^*)}}{u_i^{(T^*)}}\right)}{\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}}+\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}}},$$

$$\dot{H}_{0(\alpha^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) = \frac{\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\log\left(\frac{1-u_i^{(C^*)}}{u_i^{(C^*)}}\right)}{\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}+\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}},$$

$$\begin{aligned} \dot{h}_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) &= \frac{\alpha^{(T^*)}\exp\left(z_i^{(T^*)}\right)\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}}{(\sigma^{(T^*)})^2\left[\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}}+\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}}\right]^2}\left\{\left[\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}}\right.\right. \\ &\quad \left.+\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}}\right]\left\{\sigma^{(T^*)}\left(\dot{z}_{i(\sigma^{(T^*)})}^{(T^*)}+\frac{(1-\alpha^{(T^*)})\dot{u}_{i(\sigma)}^{(T^*)}}{1-u_i^{(T^*)}}\right)-1\right\} \\ &\quad \left.-\alpha^{(T^*)}\sigma^{(T^*)}\dot{u}_{i(\sigma^{(T^*)})}^{(T^*)}\left[\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}-\left(1-u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}\right]\right\}, \end{aligned}$$

$$\begin{aligned} \dot{h}_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) &= \frac{\alpha^{(C^*)}\exp\left(z_i^{(C^*)}\right)\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}}{(\sigma^{(C^*)})^2\left[\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}+\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\right]^2}\left\{\left[\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\right.\right. \\ &\quad \left.+\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}\right]\left\{\sigma^{(C^*)}\left(\dot{z}_{i(\sigma^{(C^*)})}^{(C^*)}+\frac{(1-\alpha^{(C^*)})\dot{u}_{i(\sigma^{(C^*)})}^{(C^*)}}{1-u_i^{(C^*)}}\right)-1\right\} \\ &\quad \left.-\alpha^{(C^*)}\sigma^{(C^*)}\dot{u}_{i(\sigma^{(C^*)})}^{(C^*)}\left[\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}-\left(1-u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}\right]\right\}, \end{aligned}$$

$$\dot{H}_{0(\sigma^{(T^*)})}^{(T^*)}(y_i|\boldsymbol{\kappa}) = \alpha^{(T^*)}\dot{u}_{i(\sigma^{(T^*)})}^{(T^*)} \left\{ \left[\frac{\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1} - \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}}{\left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}}} \right] - \frac{1}{u_i^{(T^*)}} \right\},$$

$$\dot{H}_{0(\sigma^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) = \alpha^{(C^*)}\dot{u}_{i(\sigma^{(C^*)})}^{(C^*)} \left\{ \left[\frac{\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1} - \left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}}{\left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}} + \left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}} \right] - \frac{1}{u_i^{(C^*)}} \right\},$$

$$\begin{aligned} \dot{h}_{0(\boldsymbol{\beta})}^{(T^*)}(y_i|\boldsymbol{\kappa}) &= \frac{\alpha^{(T^*)} \exp\left(z_i^{(T^*)}\right) \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}}{\sigma^{(T^*)} \left[\left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \right]^2} \left\{ \left[\left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}} \right] \right. \\ &\quad \left. \left[\dot{z}_{i(\boldsymbol{\beta})}^{(T^*)} + \frac{(1 - \alpha^{(T^*)})\dot{u}_{i(\boldsymbol{\beta})}^{(T^*)}}{1 - u_i^{(T^*)}} \right] - \alpha^{(T^*)}\dot{u}_{i(\boldsymbol{\beta})}^{(T^*)} \left[\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1} - \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1} \right] \right\}, \end{aligned}$$

$$\dot{H}_{0(\boldsymbol{\beta})}^{(T^*)}(y_i|\boldsymbol{\kappa}) = \alpha^{(T^*)}\dot{u}_{i(\boldsymbol{\beta})}^{(T^*)} \left\{ \left[\frac{\left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1} - \left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}-1}}{\left(1 - u_i^{(T^*)}\right)^{\alpha^{(T^*)}} + \left(u_i^{(T^*)}\right)^{\alpha^{(T^*)}}} \right] - \frac{1}{u_i^{(T^*)}} \right\},$$

$$\begin{aligned} \dot{h}_{0(\mu^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\kappa}) &= \frac{\alpha^{(C^*)} \exp\left(z_i^{(C^*)}\right) \left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}}{\sigma^{(C^*)} \left[\left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}} + \left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}} \right]^2} \left\{ \left[\left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}} + \left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}} \right] \right. \\ &\quad \left. \left[\dot{z}_{i(\mu^{(C^*)})}^{(C^*)} + \frac{(1 - \alpha^{(C^*)})\dot{u}_{i(\mu^{(C^*)})}^{(C^*)}}{1 - u_i^{(C^*)}} \right] - \alpha^{(C^*)}\dot{u}_{i(\mu^{(C^*)})}^{(C^*)} \left[\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1} - \left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1} \right] \right\}, \end{aligned}$$

$$\dot{H}_{0(\mu^{(C^*)})}^{(C^*)}(y_i|\boldsymbol{\nu}) = \alpha^{(C^*)}\dot{u}_{i(\mu^{(C^*)})}^{(C^*)} \left\{ \left[\frac{\left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1} - \left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}-1}}{\left(1 - u_i^{(C^*)}\right)^{\alpha^{(C^*)}} + \left(u_i^{(C^*)}\right)^{\alpha^{(C^*)}}} \right] - \frac{1}{u_i^{(C^*)}} \right\},$$

$$\dot{u}_{i(\boldsymbol{\beta})}^{(T^*)} = -\dot{z}_{i(\boldsymbol{\beta})}^{(T^*)} \exp\left\{z_i^{(T^*)} - \exp\left(z_i^{(T^*)}\right)\right\}, \quad \dot{u}_{i(\sigma^{(T^*)})}^{(T^*)} = -\dot{z}_{i(\sigma^{(T^*)})}^{(T^*)} \exp\left\{z_i^{(T^*)} - \exp\left(z_i^{(T^*)}\right)\right\},$$

$$\dot{u}_{i(\sigma)}^{(C^*)} = -\dot{z}_{i(\sigma^{(C^*)})}^{(C^*)} \exp\left\{z_i^{(C^*)} - \exp\left(z_i^{(C^*)}\right)\right\}, \quad \dot{u}_{i(\mu^{(C^*)})}^{(C^*)} = -\dot{z}_{i(\mu^{(C^*)})}^{(C^*)} \exp\left\{z_i^{(C^*)} - \exp\left(z_i^{(C^*)}\right)\right\}$$

e

$$\dot{z}_{i(\boldsymbol{\beta})}^{(T^*)} = \frac{\partial z_i^{(T^*)}}{\partial \beta_j} = \frac{x_j}{\sigma^{(T^*)}}, \quad \dot{z}_{i(\sigma^{(T^*)})}^{(T^*)} = \frac{\mathbf{x}_i^\top \boldsymbol{\beta}^{(T^*)} - y_i}{(\sigma^{(T^*)})^2}, \quad \dot{z}_{i(\mu^{(C^*)})}^{(C^*)} = \frac{1}{\sigma^{(C^*)}}.$$

APÊNDICE C - Tabela de análise de influência local (Capítulo 5)

Tabela 5.6 - Estimativas, [mudanças relativas] e (valores- p) dos parâmetros para os tempos baseline e follow-up considerando a cópula de Frank

(continua)								
Exclusão	Nenhuma	C_1	C_2	C_3	C_4	C_5	C_6	C_7
λ	[-]	[0,6896]	[2,2125]	[1,4052]	[2,922]	[2,1636]	[3,6039]	[4,4816]
	6,5401	6,495	6,3954	6,4482	6,349	6,3986	6,3044	6,247
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
β_{01}	[-]	[-2,8814]	[0,8035]	[0,9697]	[-2,019]	[-1,5723]	[2,2684]	[-0,2182]
	2,8875	2,9707	2,8643	2,8595	2,9458	2,9329	2,822	2,8938
	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)
β_{111}	[-]	[56,7481]	[48,0554]	[-39,3183]	[90,5407]	[18,3417]	[8,5782]	[49,2792]
	0,0882	0,0381	0,0458	0,1228	0,0083	0,072	0,0806	0,0447
	(0,5789)	(0,8054)	(0,7806)	(0,4396)	(0,9589)	(0,6468)	(0,6223)	(0,7846)
β_{121}	[-]	[22,2073]	[41,0150]	[-76,9871]	[56,0598]	[-45,5961]	[-41,0043]	[-18,4032]
	0,1117	0,0869	0,0659	0,1977	0,0491	0,1627	0,1575	0,1323
	(0,6888)	(0,7482)	(0,8199)	(0,4810)	(0,8621)	(0,555)	(0,5859)	(0,6442)
β_{21}	[-]	[11,7796]	[-15,8828]	[-40,2128]	[0,5080]	[-23,5549]	[-65,3486]	[-45,5596]
	0,2284	0,2015	0,2646	0,3202	0,2272	0,2822	0,3776	0,3324
	(0,4133)	(0,4698)	(0,3513)	(0,2488)	(0,4254)	(0,3128)	(0,1773)	(0,2424)
β_{311}	[-]	[35,3161]	[-6,3366]	[-20,7819]	[28,2217]	[15,5278]	[-32,1313]	[-0,3517]
	-0,145	-0,0938	-0,1542	-0,1752	-0,1041	-0,1225	-0,1916	-0,1455
	(0,4463)	(0,6125)	(0,4291)	(0,364)	(0,5879)	(0,5201)	(0,3312)	(0,4615)
β_{321}	[-]	[-584,6088]	[0,6058]	[677,5275]	[-572,6303]	[38,123]	[773,8077]	[219,3328]
	0,0117	0,0802	0,0117	-0,0677	0,0788	0,0073	-0,0790	-0,0140
	(0,9664)	(0,7653)	(0,9678)	(0,8091)	(0,7800)	(0,9791)	(0,7850)	(0,9614)
β_{41}	[-]	[-43,6475]	[-4,7513]	[24,1401]	[-42,4545]	[-17,212]	[23,8471]	[-8,6862]
	0,1433	0,2059	0,1501	0,1087	0,2042	0,168	0,1092	0,1558
	(0,3855)	(0,2016)	(0,3724)	(0,5109)	(0,2195)	(0,3066)	(0,5158)	(0,3563)
β_{511}	[-]	[0,1527]	[-5,7376]	[-7,9006]	[-4,2475]	[-5,8202]	[-15,8384]	[-12,8209]
	-0,2423	-0,2419	-0,2562	-0,2614	-0,2526	-0,2564	-0,2806	-0,2733
	(0,0599)	(0,0539)	(0,0499)	(0,0460)	(0,0489)	(0,0459)	(0,0342)	(0,0375)
β_{521}	[-]	[48,2997]	[-9,8648]	[2,8861]	[31,9274]	[42,2839]	[-8,4832]	[22,0012]
	0,4393	0,2271	0,4827	0,4267	0,2991	0,2536	0,4766	0,3427
	(0,0323)	(0,2804)	(0,0201)	(0,0386)	(0,1773)	(0,2403)	(0,0227)	(0,1308)
β_{611}	[-]	[85,5329]	[18,4501]	[27,7082]	[92,3940]	[100,4204]	[43,6462]	[103,0815]
	0,1012	0,0146	0,0826	0,0732	0,0077	-0,0004	0,057	-0,0031
	(0,7714)	(0,9655)	(0,8168)	(0,8370)	(0,9825)	(0,9990)	(0,8753)	(0,9931)
β_{621}	[-]	[14,6667]	[0,7533]	[1,9729]	[13,1646]	[14,6645]	[2,0605]	[11,9252]
	0,4567	0,3897	0,4532	0,4477	0,3966	0,3897	0,4473	0,4022
	(0,1768)	(0,2365)	(0,1916)	(0,1971)	(0,2470)	(0,2522)	(0,2087)	(0,2567)
β_{71}	[-]	[-12,7027]	[-8,4377]	[36,3965]	[-17,0321]	[19,1710]	[28,2682]	[18,3447]
	0,1086	0,1224	0,1177	0,069	0,1271	0,0877	0,0779	0,0886
	(0,4648)	(0,3886)	(0,4360)	(0,6426)	(0,3836)	(0,5422)	(0,6070)	(0,5499)
β_{811}	[-]	[17,5383]	[1,4407]	[-0,5275]	[17,5213]	[15,0539]	[-1,0550]	[12,6773]
	0,1763	0,1454	0,1738	0,1772	0,1454	0,1498	0,1782	0,154
	(0,1579)	(0,2336)	(0,1731)	(0,1599)	(0,2481)	(0,2288)	(0,1656)	(0,2304)
β_{821}	[-]	[1,0655]	[1,0209]	[-1,7617]	[2,7033]	[-0,4237]	[-1,6527]	[0,4931]
	0,4036	0,3993	0,3995	0,4107	0,3927	0,4053	0,4103	0,4016

Tabela 5.6 - Estimativas, [mudanças relativas] e (valores- p) dos parâmetros para os tempos baseline e follow-up considerando a cópula de Frank

(continuação)								
Exclusão	Nenhuma	C_1	C_2	C_3	C_4	C_5	C_6	C_7
	(0,006)	(0,0056)	(0,0073)	(0,0050)	(0,0077)	(0,0052)	(0,0057)	(0,0067)
β_{911}	[-]	[12,3705]	[-71,9775]	[127,1981]	[-46,5702]	[117,4943]	[68,9857]	[70,8474]
	0,0873	0,0765	0,1502	-0,0238	0,128	-0,0153	0,0271	0,0255
	(0,4921)	(0,5372)	(0,2607)	(0,8547)	(0,3297)	(0,9055)	(0,8374)	(0,8472)
β_{921}	[-]	[-4,3624]	[-25,0195]	[-3,3451]	[-23,0903]	[-6,1816]	[-33,0752]	[-30,0600]
	0,2182	0,2278	0,2728	0,2255	0,2686	0,2317	0,2904	0,2838
	(0,0173)	(0,0103)	(0,0047)	(0,0149)	(0,0043)	(0,0103)	(0,0027)	(0,0031)
β_{101}	[-]	[-13,2866]	[9,7224]	[12,3011]	[-4,2000]	[-1,2793]	[24,6732]	[12,0985]
	-0,3948	-0,4472	-0,3564	-0,3462	-0,4113	-0,3998	-0,2974	-0,347
	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)
β_{11}	[-]	[91,6717]	[-67,7171]	[-58,2826]	[17,6614]	[24,5902]	[-138,7507]	[-71,9523]
	-0,0575	-0,0048	-0,0964	-0,091	-0,0473	-0,0433	-0,1372	-0,0988
	(0,4960)	(0,9552)	(0,2675)	(0,2883)	(0,5994)	(0,6245)	(0,1166)	(0,2852)
β_{02}	[-]	[-1,7591]	[-0,1083]	[0,7552]	[-1,7679]	[-0,6439]	[0,8400]	[-0,5151]
	3,4165	3,4766	3,4202	3,3907	3,4769	3,4385	3,3878	3,4341
	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)	(< 0, 01)
β_{112}	[-]	[-165,2968]	[-116,7258]	[125,3916]	[-247,8495]	[-45,3920]	[12,9270]	[-114,5733]
	-0,0502	-0,1331	-0,1087	0,0127	-0,1745	-0,0729	-0,0437	-0,1077
	(0,7924)	(0,4748)	(0,5709)	(0,9458)	(0,3579)	(0,6963)	(0,8146)	(0,5681)
β_{122}	[-]	[17,9331]	[30,9663]	[-81,7249]	[44,5081]	[-53,2645]	[-57,1757]	[-35,2755]
	0,1608	0,132	0,111	0,2923	0,0892	0,2465	0,2528	0,2176
	(0,5911)	(0,6464)	(0,7162)	(0,3179)	(0,7638)	(0,3905)	(0,3909)	(0,4578)
β_{22}	[-]	[-13,5716]	[31,0160]	[100,3102]	[6,4886]	[74,0291]	[149,3235]	[115,0191]
	-0,1228	-0,1395	-0,0847	0,0004	-0,1149	-0,0319	0,0606	0,0184
	(0,6843)	(0,6469)	(0,7761)	(0,9989)	(0,7040)	(0,9125)	(0,8249)	(0,9481)
β_{312}	[-]	[39,2112]	[-4,8815]	[-22,6829]	[30,9879]	[16,1972]	[-31,8051]	[0,3957]
	-0,2325	-0,1413	-0,2439	-0,2853	-0,1605	-0,1949	-0,3065	-0,2316
	(0,2806)	(0,5029)	(0,2570)	(0,1804)	(0,4553)	(0,3623)	(0,1471)	(0,2847)
β_{322}	[-]	[63,3405]	[-4,4498]	[-64,9917]	[54,1403]	[0,4386]	[-78,0729]	[-23,6573]
	-0,2052	-0,0752	-0,2143	-0,3385	-0,0941	-0,2043	-0,3654	-0,2537
	(0,5233)	(0,8104)	(0,5133)	(0,2872)	(0,7730)	(0,5239)	(0,2571)	(0,4418)
β_{42}	[-]	[59,6874]	[1,7870]	[-34,0306]	[52,7582]	[21,7487]	[-36,9651]	[6,2874]
	-0,1673	-0,0675	-0,1643	-0,2243	-0,079	-0,1309	-0,2292	-0,1568
	(0,3622)	(0,7098)	(0,3721)	(0,2111)	(0,6689)	(0,4727)	(0,1980)	(0,3928)
β_{512}	[-]	[1,4721]	[-8,0926]	[-29,5914]	[-1,9366]	[-20,1763]	[-45,0051]	[-31,7090]
	-0,127	-0,1252	-0,1373	-0,1646	-0,1295	-0,1527	-0,1842	-0,1673
	(0,3863)	(0,3857)	(0,3374)	(0,2570)	(0,3611)	(0,2886)	(0,1907)	(0,2351)
β_{522}	[-]	[100,9776]	[-25,4129]	[10,5148]	[65,7938]	[95,2973]	[-17,8592]	[52,6902]
	0,3009	-0,0029	0,3774	0,2693	0,1029	0,0142	0,3547	0,1424
	(0,2084)	(0,9904)	(0,1105)	(0,2515)	(0,6821)	(0,9538)	(0,1248)	(0,5706)
β_{612}	[-]	[73,3693]	[22,6689]	[20,0987]	[81,5366]	[79,7052]	[44,7341]	[88,3361]
	0,1642	0,0437	0,127	0,1312	0,0303	0,0333	0,0907	0,0192
	(0,6651)	(0,9061)	(0,7336)	(0,7251)	(0,9348)	(0,9283)	(0,8026)	(0,9583)
β_{622}	[-]	[22,3382]	[5,0190]	[4,3064]	[22,5685]	[22,5925]	[10,1339]	[22,8972]
	0,4168	0,3237	0,3959	0,3989	0,3228	0,3227	0,3746	0,3214
	(0,2509)	(0,3606)	(0,2684)	(0,2650)	(0,3629)	(0,3627)	(0,2828)	(0,3607)
β_{72}	[-]	[-29,7453]	[-10,2254]	[50,1416]	[-35,2851]	[15,1650]	[43,0284]	[16,0191]
	0,1300	0,1686	0,1433	0,0648	0,1758	0,1103	0,0740	0,1092
	(0,4488)	(0,3087)	(0,4027)	(0,7041)	(0,2922)	(0,5092)	(0,6622)	(0,5159)
β_{812}	[-]	[100,3907]	[22,2708]	[-32,6317]	[113,4617]	[60,5910]	[-25,7052]	[56,5706]
	0,0469	-0,0002	0,0365	0,0623	-0,0063	0,0185	0,0590	0,0204
	(0,7473)	(0,9990)	(0,8025)	(0,6613)	(0,9653)	(0,8962)	(0,6771)	(0,8868)
β_{822}	[-]	[10,5188]	[-26,1467]	[69,5096]	[-27,8880]	[67,4813]	[62,8643]	[44,2431]
	-0,0366	-0,0327	-0,0461	-0,0112	-0,0468	-0,0119	-0,0136	-0,0204