

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

Análise de sensibilidade e resíduos em modelos de regressão com respostas  
bivariadas por meio de cópulas

Eduardo Monteiro de Castro Gomes

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

Piracicaba  
2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Eduardo Monteiro de Castro Gomes  
Bacharel em Estatística

**Análise de sensibilidade e resíduos em modelos de regressão com respostas  
bivariadas por meio de cópulas**

Orientador:  
Prof. Dr. **Edwin Moises Marcos Ortega**

Dissertação apresentada para obtenção do título de  
Mestre em Agronomia. Área de concentração: Es-  
tatística e Experimentação Agrônômica

**Piracicaba**  
**2007**

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

Gomes, Eduardo Monteiro de Castro

Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas / Eduardo Monteiro de Castro Gomes. - - Piracicaba, 2007. 103 p. : il.

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

1. Análise de regressão e correlação 2. Análise de sobrevivência 3. Dados censurados 4. Métodos Monte Carlo 5. Simulação (estatística) I. Título

CDD 519.536

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

## Dedicatória

*A meu irmão Bernardo (in memoriam),  
que vive comigo em tudo que faço.*

## AGRADECIMENTOS

Aos meus pais, Cíntia Monteiro de Castro e Vicente Gomes Neto pelo amor e apoio incondicional.

A minha avó Gina por ser minha maior incentivadora.

A Ana Maria e meus irmãos Mariana e Rodrigo por todo o carinho e apoio.

Ao professor Dr. Edwin Moises Marcos Ortega, pela orientação e formação, por ser um exemplo e principalmente um bom amigo.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, Profa. Dra. Clarice Garcia Borges Demétrio, Prof. Dr. Décio Barbin, Profa. Dra. Roseli Aparecida Leandro e Prof. Dr. Silvio Sandoval Zocchi pelos ensinamentos e amizade.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, em especial a Solange de Assis Paes Sabadin pelos auxílios permanentes.

Aos funcionários do restaurante universitário (RUCAS) pelos ótimos serviços prestados, essenciais para meu bem-estar em Piracicaba.

Às amigas Juliana, Fernanda, Giovana, Camila e Sarah pela amizade e confiança.

Aos amigos Bruno e Jalmar, companheiros de Estatística e de tantas outras alegrias.

A Ana Clara pelo companherismo, paciência e carinho.

Ao CNPq pela concessão de bolsa de estudos.

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

## SUMÁRIO

RESUMO . . . . .	8
ABSTRACT . . . . .	9
1 INTRODUÇÃO . . . . .	10
2 REVISÃO DE LITERATURA . . . . .	13
2.1 Análise de Sobrevivência . . . . .	13
2.1.1 Modelos de Locação e Escala . . . . .	16
2.1.2 Modelo de regressão Weibull . . . . .	16
2.2 Cópulas . . . . .	17
2.2.1 Cópula de Frank . . . . .	18
2.2.2 Cópula de Clayton . . . . .	19
2.3 Técnicas de Diagnóstico . . . . .	19
2.3.1 Influência Local . . . . .	19
2.3.2 Influência Global . . . . .	21
2.3.3 Análise de Resíduos . . . . .	22
3 MATERIAL E MÉTODOS . . . . .	24
3.1 Material . . . . .	24
3.1.1 Dados de retinopatia . . . . .	24
3.1.2 Dados de pacientes com insuficiência renal . . . . .	25
3.2 Métodos . . . . .	26
3.3 Modelos de Regressão Bivariados . . . . .	26
3.4 Estratégias de Inferência . . . . .	27
3.4.1 Estimadores de máxima verossimilhança . . . . .	29
3.4.2 Estimadores de Jackknife . . . . .	30

	6
3.4.3	Influência local . . . . . 31
3.4.4	Influência global . . . . . 33
3.4.5	Análise de resíduos . . . . . 34
3.4.6	Impacto das observações influentes . . . . . 34
3.5	Metodologia de simulação . . . . . 35
3.5.1	Envelopes simulados . . . . . 36
4	RESULTADOS . . . . . 38
4.1	Resultados das simulações . . . . . 38
4.2	Aplicação com os dados de pacientes diabéticos . . . . . 50
4.2.1	Análise exploratória . . . . . 50
4.2.2	Modelo de regressão bivariado através da cópula de Frank . . . . . 52
4.2.2.1	Análise de influência local . . . . . 54
4.2.2.2	Análise de influência global . . . . . 55
4.2.2.3	Análise de resíduos . . . . . 55
4.2.2.4	Impacto das observações influentes . . . . . 57
4.2.2.5	Modelo final proposto utilizando a cópula de Frank . . . . . 58
4.2.3	Modelo de regressão bivariado através da cópula de Clayton . . . . . 61
4.2.3.1	Análise de influência local . . . . . 63
4.2.3.2	Análise de influência global . . . . . 64
4.2.3.3	Análise de resíduos . . . . . 65
4.2.3.4	Impacto das observações influentes . . . . . 65
4.2.3.5	Modelo final proposto utilizando a cópula de Clayton . . . . . 66
4.3	Aplicação com os dados de infecção em pacientes com insuficiência renal . . . . . 69
4.3.1	Análise exploratória . . . . . 70
4.3.2	Modelo de regressão bivariado através da cópula de Frank . . . . . 71



	7
4.3.2.1	Análise de influência local . . . . . 71
4.3.2.2	Análise de influência global . . . . . 73
4.3.2.3	Análise de resíduos . . . . . 75
4.3.2.4	Impacto das observações influentes . . . . . 77
4.3.2.5	Modelo final proposto utilizando a cópula de Frank . . . . . 80
4.3.3	Modelo de regressão bivariado através da cópula de Clayton . . . . . 82
4.3.3.1	Análise de influência local . . . . . 83
4.3.3.2	Análise de influência global . . . . . 85
4.3.3.3	Análise de resíduos . . . . . 85
4.3.3.4	Impacto das observações influentes . . . . . 86
4.3.3.5	Modelo final proposto utilizando a cópula de Clayton . . . . . 86
5	CONSIDERAÇÕES FINAIS . . . . . 92
5.1	Trabalhos futuros . . . . . 93
	REFERÊNCIAS . . . . . 94
	ANEXOS . . . . . 97

## RESUMO

### **Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas**

Neste trabalho são apresentados modelos de regressão com respostas bivariadas obtidos através de funções de cópulas. O objetivo de utilizar estes modelos bivariados é modelar a correlação entre eventos e captar nos parâmetros de regressão a influência da associação entre as variáveis resposta na presença de censuras nos dados. Os parâmetros dos modelos, são estimados por meio dos métodos de máxima verossimilhança e jackknife. Alguns métodos da análise de sensibilidade como influência global, local e local total de um indivíduo, são introduzidos e calculados considerando diferentes esquemas de perturbação. Uma análise de resíduos foi proposta para verificar a qualidade de ajuste dos modelos utilizados e também foi proposta novas medidas de resíduos para respostas bivariadas. Métodos de simulação de Monte Carlo foram conduzidos para estudar a distribuição empírica dos resíduos marginais e bivariados propostos. Finalmente, os resultados serão aplicados à dois conjuntos de dados disponíveis na literatura.

*Palavras-chaves:* Modelos de regressão bivariados; Dados censurados; Cópulas arquimedianas; Análise de sensibilidade; Análise de resíduo; Simulação de Monte Carlo

## ABSTRACT

### **Bivariate response regression models with copulas: Sensitivity and residual analysis**

In this work bivariate response regression models are presented with the use of copulas. The objective of this approach is to model the correlation between events and capture the influence of this correlation in the regression parameters. The models are used in the context of survival analysis and are fitted to two data sets available in the literature. Inferences are obtained using maximum likelihood and Jackknife methods. Sensitivity techniques such as local and global influence are proposed and calculated. A residual analysis is proposed to check the adequacy of the models and simulation methods are used to assess the empirical distribution of the marginal univariate and bivariate residual measures proposed.

*Keywords:* Regression models; Censored data; Archimedean copulas; Sensitivity analysis; Residual analysis; Monte Carlo simulation

## 1 INTRODUÇÃO

Em diversos tipos de estudos em áreas como Agronomia ou Medicina tem-se interesse em conhecer ou explicar o tempo até a ocorrência de um determinado evento de interesse e determinar efeitos de outros fatores sobre esse tempo. Técnicas de análise de Sobrevivência permitem a utilização de modelos de regressão para estudos deste tipo que se particularizem pela ocorrência de informações parciais do tempo de um ou mais elementos da amostra. Essas informações parciais são chamadas de censuras e acontecem por motivos que segundo sua natureza caracterizam o tipo de censura e requerem técnicas específicas de modelagem e inferência.

A aplicação de técnicas de análise de sobrevivência é comum em estudos referentes ao tempo até a germinação de sementes, tempo de vida de árvores, tempo de vida de pacientes com doença crônica diagnosticada, dentre outros possíveis exemplos.

Existem na literatura diversas formas de abordagens para dados de sobrevivência. O estimador não paramétrico proposto por Kaplan e Meier (1958) e os modelos de regressão de localização e escala e os de riscos proporcionais de Cox são algumas das ferramentas mais utilizadas. Essas abordagens, entretanto, são geralmente mais difundidas em casos com apenas um evento de interesse ou variável resposta observada por elemento da amostra.

A necessidade de relacionar tempos até a ocorrência de eventos distintos e não independentes em um mesmo indivíduo demanda por modelos de respostas multivariadas em análise de sobrevivência. Existem na literatura algumas formas propostas de realizar este tipo de modelagem. Dentre elas a utilização de modelos paramétricos multivariados clássicos como a normal multivariada e a utilização de cópulas para obtenção de modelos multivariados mais flexíveis. Neste contexto da utilização de cópulas, Lawless (2005) sugere a utilização de modelos marginais de localização e escala como alternativa aos modelos marginais de Cox utilizados por exemplo em Nuñez (2005) e Tibaldi (2004). Esta alternativa proposta por Lawless é importante nos casos em que a suposição de riscos proporcionais, necessária para a utilização dos modelos de Cox, não pode ser sustentada. Neste trabalho são utilizados marginalmente os modelos de localização e escala com distribuições conjuntas especificadas através de funções de cópulas.

Segundo Nelsen (2006) cópulas são funções que fornecem um meio de relacionar funções de distribuição multivariadas com funções de distribuição marginais. Ou alternativa-

mente, cópulas são funções de distribuição multivariadas, cujas marginais unidimensionais são uniformes no intervalo  $[0, 1]$ . Nelsen (2006) traz uma abordagem ampla sobre a teoria de cópulas, enquanto tópicos utilizados neste trabalho podem ser consultados em Genest e Rivest (1993), He e Lawless (1995) e Tibaldi (2004). Mediante a utilização das cópulas é possível reunir múltiplas respostas em modelo de regressão, podendo cada resposta ser de diferentes famílias e formas de distribuição de probabilidade.

Quando os modelos propostos são ajustados a conjuntos de dados é importante estudar a robustez dos resultados obtidos com relação à presença de pontos extremos ou a perturbações na forma do modelo adotado. Detectar observações aberrantes ou influentes constitui um passo importante na validação da análise estatística de um conjunto de dados.

Existem na literatura numerosas técnicas para se determinar a ocorrência de observações influentes nas estimativas dos parâmetros. Diversos resultados, comentários e sugestões a respeito podem ser encontrados, por exemplo, Cook e Weisberg (1982) e Cook (1986) dentre outros. Métodos de influência global e local são abordados neste trabalho. Três esquemas de perturbações nos modelos são propostos para os modelos apresentados.

Além da avaliação da robustez dos resultados, a validação do ajuste de um modelo deve ser realizada com auxílio de uma análise de resíduos. Neste tipo de análise procura-se detectar pontos mal ajustados, assim como verificar a ocorrência de afastamentos sérios das suposições feitas para o erro do modelo.

Nas aplicações contidas neste trabalho é realizada uma análise marginal de algumas medidas de resíduos comumente utilizados em modelos para análise de sobrevivência. Sendo estas o resíduo Martingale proposto por Therneau, Grambsch e Fleming (1990) e o resíduo Deviance, proposto por Fahrmeir e Tutz (1994) e Pierce e Schafer (1986).

Uma medida de resíduo é proposta neste trabalho para a aplicação com os modelos de sobrevivência com resposta bivariada. Esta medida é obtida mediante adaptações nos resíduos martingale e deviance.

Por meio de metodologia de simulação de Monte Carlo investiga-se o comportamento probabilístico das medidas de resíduo quando são utilizados os modelos e a metodologia de inferência propostos. O ambiente de programação R é utilizado nos procedimentos de simu-

lação. Nas aplicações dos modelos, os gráficos de envelopes simulados propostos por Atkinson (1985) são utilizados na validação do ajuste dos modelos nas aplicações neste trabalho.

Os modelos propostos são ajustados a dois conjunto de dados disponíveis na literatura. O primeiro conjunto de dados é referente a tempos até a ocorrência de retinopatia leve e grave em pacientes com diabetes melitus diagnosticados no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP), mais detalhes ver Achcar, et al (2006). O segundo conjunto é referente a tempos até dois eventos distintos de infecção em pacientes com deficiência renal mais detalhes ver McGilchrist, Aisbet (1991).

Este trabalho está organizado de forma em que na seção 2 são apresentados alguns conceitos de Análise de Sobrevivência, modelos de locação e escala, cópulas e métodos de diagnóstico. Na seção 3 são apresentados os conjuntos de dados e as metodologias de ajuste, diagnóstico e simulação utilizados. Na seção 4.1 estão os resultados das simulações realizadas e nas seções 4.2 e 4.3 os resultados dos ajustes aos conjuntos de dados.

## 2 REVISÃO DE LITERATURA

Nesta seção são apresentados alguns tópicos de análise de sobrevivência necessários para obtenção dos modelos de regressão paramétricos, conceitos de cópulas, metodologias de inferência e medidas de diagnóstico e influência.

### 2.1 Análise de Sobrevivência

Análise de Sobrevivência é um conjunto de metodologias e técnicas estatísticas utilizadas em estudos das mais diversas áreas que têm interesse em conhecer ou explicar o tempo até a ocorrência de um evento de interesse e determinar a influência de outros fatores ou covariáveis sobre este tempo. Essas técnicas são particulares por permitirem a utilização de informações parciais na modelagem. A estas informações parciais dá-se o nome de censuras e elas ocorrem conforme as seguintes classificações:

#### 1. Censuras à esquerda

Este mecanismo de censura é menos frequente e acontece quando o tempo inicial de observação do indivíduo até a ocorrência do evento não corresponde ao tempo inicial de exposição do indivíduo a ocorrência do evento. Como exemplo pode ser citado o paciente que entra no estudo quando do diagnóstico de uma doença e a informação do tempo até o óbito do paciente em consequência desta doença é incompleta, pois não se sabe certamente quando a doença foi contraída, mas sabe-se que foi antes do diagnóstico.

#### 2. Censuras à direita

Este mecanismo de censura é subdividido em três categorias:

Censura tipo I: Experimentos com tempo de duração pré-determinados podem terminar antes da observação do evento em um ou mais elementos da amostra. Para esses elementos sabe-se apenas que o evento não ocorreu até o final do experimento, e essa informação não deve ser desconsiderada.

Censura tipo II: Experimentos com número pré-determinado  $k$  de eventos a serem observados. No momento em que o  $k$ -ésimo evento for observado termina-se o experimento

e registra-se que os tempos até a ocorrência do evento para os  $(n - k)$  elementos que não vivenciaram o evento seriam maiores ou iguais ao tempo do evento para o  $k$ -ésimo elemento.

Censura Aleatória: Esse mecanismo contempla todos os casos em que elementos não tenham o evento observado por motivos não controláveis pelo estudo, tais como ventos fortes que podem derrubar árvores de experimentos, o abandono de um tratamento por parte de um paciente, ou ocorrência de outro evento que não permita a observação do evento de interesse.

### 3. Censura Intervalar

Este tipo de censura é mais geral e acontece em estudos em que os elementos da amostra têm acompanhamento em ocasiões periódicas. Para a ocorrência do evento de interesse pode-se conhecer apenas o intervalo entre acompanhamentos no qual o evento ocorreu. Como exemplo pode ser citado o caso em que não se sabe ao certo o tempo até a germinação de uma semente, mas sabe-se que foi no período entre a visita em que foi constatada a germinação e a visita anterior.

Neste trabalho serão considerados casos com censuras aleatórias à direita. Na modelagem de dados com este mecanismo de censura assume-se que existem duas variáveis independentes  $T$  e  $C$  representando respectivamente os tempos de falha e censura de um indivíduo. Para cada elemento da amostra observa-se

$$t = \min(T, C).$$

A esta observação associa-se a variável indicadora de censura definida por:

$$\delta = \begin{cases} 0 & \text{se } T > C \\ 1 & \text{se } T \leq C \end{cases}$$

No contexto da utilização dessas informações parciais da variável resposta, a metodologia de análise de sobrevivência aplicada neste trabalho depende da utilização de funções de probabilidade que são relacionadas diretamente às funções de densidade e distribuição de probabilidades na formulação dos modelos. Essas funções utilizadas são definidas e seguindo a notação utilizada na literatura, o tempo até ocorrência do evento de interesse será muitas vezes referido por tempo de falha.



Seja  $T$  uma variável aleatória positiva representando o tempo de falha e  $f(t)$  a respectiva função de densidade. A função de sobrevivência de uma variável aleatória representa a probabilidade do elemento não falhar até um tempo  $t$  e pode ser expressa por:

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

A função de risco determina a taxa instantânea de falha no tempo  $t$  condicionada à sobrevivência até o instante  $t$ . O comportamento dessa função é de fundamental importância na discriminação de modelos e ela pode ser definida por:

$$h(t) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

No contexto da utilização de modelos paramétricos de probabilidade para descrever a variável  $T$  é necessário realizar estimativas referentes às funções de probabilidade utilizadas. Usualmente uma amostra da população que apresenta a variável em estudo é utilizada para estimar os parâmetros. Na presença de censura nos dados pode-se definir a amostra como um conjunto de observações de pares  $(t_i, \delta_i)$ ,  $i = 1, 2, \dots, n$  em que  $t_i$  é o valor observado para o tempo de falha e  $\delta_i$  a variável indicadora de censura para o  $i$ -ésimo indivíduo e  $n$  é o tamanho da amostra.

Uma importante ferramenta na área de análise de sobrevivência é o estimador não-paramétrico de Kaplan-Meier, proposto por Kaplan e Meier (1958) para estimar a função de sobrevivência. A expressão geral deste estimador tem a forma:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right)$$

em que  $t_1 < t_2 \dots < t_r$  são os  $r$  tempos distintos e ordenados de falha e  $d_j$  e  $n_j$  são respectivamente o número de falhas e o número de indivíduos em risco em  $t_j$ ,  $j = 1, \dots, r$ .

As estimativas referentes aos modelos utilizados também podem ser obtidas de formas paramétricas. Uma opção é através de metodologia de máxima verossimilhança. No contexto da utilização de dados com mecanismo de censuras à direita a função de verossimilhança é adaptada na forma

$$L(\theta) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \quad (1)$$

Pode-se notar que nessa função de verossimilhança os indivíduos que apresentaram o evento têm sua função de densidade incluída no produtório enquanto os indivíduos que tiveram seus tempos censurados têm sua função de sobrevivência incluída. A interpretação dessa adaptação é bastante intuitiva uma vez que a função de sobrevivência representa a probabilidade de o indivíduo não apresentar falha até um instante  $t$ , que é exatamente a informação que se tem sobre o evento no caso de censura aleatória à direita.

Pode-se, então, estimar os parâmetros para os modelos utilizados para descrever o tempo até a ocorrência do evento. É de interesse prático relacionar o tempo a covariáveis que supostamente influenciam esse tempo. Nesse contexto, são utilizados os modelos de regressão.

### 2.1.1 Modelos de Locação e Escala

Seja  $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$  um vetor associado a  $p$  variáveis que supostamente influenciam a variável aleatória  $T$ . A classe de modelos de locação e escala emprega famílias paramétricas de distribuições para modelar a variável transformada  $Y = \log(T)$ , como visto em Lawless (2003). Esta classe se caracteriza pelo fato da distribuição de  $Y$  ter parâmetros de locação  $\mu(x)$  dependendo das covariáveis e um parâmetro constante de escala  $\sigma$ . Se utilizar  $\mu(x) = \mathbf{x}^T \boldsymbol{\beta}$ , sendo  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ , pode-se escrever:

$$Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma W,$$

em que  $W$  é resíduo que tem distribuição que não depende de  $\mathbf{x}$ .

### 2.1.2 Modelo de regressão Weibull

Os modelos com distribuição Weibull são bastante versáteis por apresentarem funções de risco crescentes ou decrescentes em função do tempo. Considerando que a variável aleatória  $T$  tem distribuição Weibull com função de densidade:

$$f(t) = \frac{\varphi}{\lambda \varphi} t^{\varphi-1} \exp \left\{ - \left( \frac{t}{\lambda} \right)^\varphi \right\}, \quad t \geq 0$$

em que  $\lambda > 0$  é o parâmetro de escala e  $\varphi > 0$  é o parâmetro de forma.

Considerando a transformação  $Y = \log(T)$ , pode-se mostrar que  $Y$  terá distribuição valor extremo  $(\mu, \sigma)$ , sendo  $\lambda = \exp(\mu)$  e  $\varphi = \frac{1}{\sigma}$ .

Nessas condições a variável  $Y$  pode ser escrita como um modelo de locação e escala da forma  $Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma W$ . As funções de distribuição e sobrevivência para  $Y$  são dadas por:

$$f(y) = \frac{1}{\sigma} \exp \left\{ \left( \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right\} \exp \left\{ - \exp \left( \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right\}.$$

$$S_Y(y) = \exp \left\{ - \exp \left( \frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma} \right) \right\}. \quad (2)$$

No contexto da utilização de amostras com presença de covariáveis tem-se observações do tipo  $(t_i, \delta_i, \mathbf{x}_i)$  para cada elemento da amostra, em que  $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$  é o vetor de covariáveis observadas para o  $i$ -ésimo indivíduo. Então o modelo Weibull de regressão de locação e escala é dado por:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma W_i, i = 1, 2, \dots, n$$

Os modelos com este tipo de formulação são utilizados nos casos em que se deseja estudar o tempo até a ocorrência de apenas um evento de interesse. Caso existam mais de um evento de interesse e estes sejam correlacionados, modelos para respostas multivariadas são necessários para captar a associação entre as respostas ou tempos. Nesse contexto as funções de cópulas foram utilizadas neste trabalho para reunir modelos marginais de locação e escala em modelos bivariados.

## 2.2 Cópulas

Cópulas são funções que fornecem um meio de relacionar funções de distribuições multivariadas com funções de distribuição marginais e apresentam enorme flexibilidade na escolha das marginais. Seja  $Y_j$  uma variável aleatória com função de distribuição marginal contínua  $F_j, j = 1, 2, \dots, k$ . A função de distribuição conjunta é da forma:

$$F(y_1, y_2, \dots, y_k) = C_\alpha(F_1(y_1), F_2(y_2), \dots, F_k(y_k))$$

em que  $C_\alpha(F_1(y_1), F_2(y_2), \dots, F_k(y_k))$  é a cópula do vetor aleatório  $(Y_1, Y_2, \dots, Y_k)$  tendo como condição que  $F_j(Y_j) \sim U(0, 1) \forall j$ . A distribuição conjunta  $C_\alpha$  descreve a dependência entre as variáveis aleatórias  $Y_1, Y_2, \dots, Y_k$ , que são independentes se e somente se

$C_\alpha(F_1(Y_1), F_2(Y_2), \dots, F_k(Y_k)) = \prod_{j=1}^k F_j(Y_j)$ . Existem na literatura cópulas com um ou mais parâmetros. Neste trabalho são abordadas apenas cópulas uniparamétricas, em que único parâmetro denominado  $\alpha$  descreve completamente a dependência entre as variáveis aleatórias.

Para trabalhos em análise de sobrevivência vale notar que se  $S_j$  é função de sobrevivência marginal de  $Y_j$ , então  $S_j(Y_j) \sim U(0, 1)$  e portanto funções de sobrevivência multivariadas também podem ser representadas por cópulas, isto é:  $S(y_1, y_2, \dots, y_k) = C_\alpha(S_1(y_1), S_2(y_2), \dots, S_p(y_k))$ .

Neste trabalho são utilizadas as cópulas de Frank e Clayton, que pertencem à classe de cópulas Archimedianas uniparamétricas, para criar modelos de regressão com duas variáveis resposta, definindo  $k = 2$ . Esta classe de cópulas possui diversas propriedades atrativas, sendo a simplicidade na obtenção da medida de associação  $\tau$  de Kendall uma propriedade bastante útil neste trabalho, como cita Nelsen (2006). As cópulas de Frank e Clayton são apresentadas a seguir como em Núñez (2005).

### 2.2.1 Cópula de Frank

Introduzida por Frank (1979), a função de sobrevivência da família de Frank para  $\{\alpha \in (-\infty, \infty) \text{ tal que } \alpha \neq 0\}$  é dada por:

$$H_\alpha(t_1, t_2) = -\frac{1}{\alpha} \log \left[ 1 + \frac{(\exp(-\alpha S_1(t_1)) - 1)(\exp(-\alpha S_2(t_2)) - 1)}{\exp(-\alpha) - 1} \right],$$

em que se  $\alpha \rightarrow 0$ ,  $S_{12}(t_1, t_2) = S_1(t_1)S_2(t_2)$  com funções de sobrevivência marginais arbitrárias  $S_1(t_1)$  e  $S_2(t_2)$ , e admitindo possíveis correlações fortes positivas ou negativas.

Interpretações sobre a correlação entre as variáveis aleatórias mediante a observação do parâmetro de associação  $\alpha$  são difíceis, devido a diferentes possíveis parametrizações e uma proposta é a utilização do  $\tau$  de Kendall que é relacionado ao parâmetro  $\alpha$  na cópula de Frank por:

$$\tau_\alpha(T_1, T_2) = 1 - \frac{4}{\alpha} \left[ 1 - \frac{1}{\alpha} \int_0^\alpha \frac{t}{e^t - 1} dt + 1 \right], \tau \in [-1, 1]. \quad (3)$$

### 2.2.2 Cópula de Clayton

Obtida por Clayton (1978) a função de sobrevivência conjunta para essa família de cópulas é dada por:

$$H_\alpha(t_1, t_2) = \left[ S_1(t_1)^{-\alpha} + S_2(t_2)^{-\alpha} - 1 \right]^{-\frac{1}{\alpha}}.$$

Note que se  $\alpha \rightarrow 0$ , obtem-se  $S_{12}(t_1, t_2) = S_1(t_1)S_2(t_2)$ .

O coeficiente  $\tau$  de Kendall na família de Clayton é relacionado ao parâmetro de associação  $\alpha$  por:

$$\tau_\alpha(T_1, T_2) = \frac{\alpha}{\alpha + 2}. \quad (4)$$

É importante ressaltar que cada cópula modela formas distintas de associação e, portanto, é necessário conhecer a forma de associação dos dados para escolher a cópula mais adequada. Metodologias para seleção de cópulas vêm sendo desenvolvidas principalmente com enfoque não-paramétrico.

### 2.3 Técnicas de Diagnóstico

Modelos estatísticos servem como descrições aproximadas de processos complexos e não são exatos. Metodologias de diagnóstico devem, portanto, ser utilizadas para detectar problemas com o modelo ajustado assim como analisar a sensibilidade das inferências a pequenas perturbações nos dados ou no modelo proposto.

Cook (1986) propôs avaliar a influência conjunta das observações sob pequenas perturbações no modelo, ao invés da avaliação pela retirada individual ou conjunta de pontos. Uma utilização de técnicas de influência local em casos com dados censurados pode ser consultada em Ortega, Cancho e Bolfarine (2006).

Uma observação é dita influente se ela produzir alterações desproporcionais nos resultados da análise, quando a mesma for excluída, ou submetida a uma pequena perturbação.

#### 2.3.1 Influência Local

Existem formas alternativas de se perturbar o modelo proposto. O objetivo é pensar em esquemas de perturbação que sejam interpretáveis. Na literatura, consideram-se,

frequentemente, os seguintes tipos de perturbação:

1. em lugar de supor que as respostas possuem a mesma variância, admitir a possibilidade de se ter um modelo heterocedástico, isto é, supor:  $Var(Y_i) = \frac{\sigma^2}{w_i}$ ,  $i = 1, 2, \dots, n$ , em que  $0 < w_i < 1$ ;
2. pode-se introduzir pequenas perturbações aditivas nas respostas, substituindo  $y_i$  por  $y_i + \sigma w_i$  em que  $\sigma$  é o desvio padrão de  $Y_i$  e  $w_i \in R$ ;
3. perturbar simultaneamente todas as componentes do  $i$ -ésimo caso, substituindo o vetor  $(x_i, y_i)$  por  $(x_{i1} + s_1 w_1, \dots, x_{ip} + s_p w_p, y_i + \sigma w_{p+1})$ , em que  $s_j$  é um fator de escala, que pode ser a norma da  $j$ -ésima coluna da matriz  $X$ , para  $j = 1, 2, \dots, p$  e  $w_i \in R$ .
4. perturbação de casos, quando para cada elemento da amostra é atribuído um peso.

Seguindo com a teoria de influência local o próximo passo é expressar o logaritmo da função de verossimilhança em função das perturbações. No caso da ponderação de casos, por exemplo, o logaritmo da função de verossimilhança tem a forma:

$$l(\Psi|\omega) = \sum_{i=1}^n \omega_i \log f(y_i; \Psi).$$

Assume-se que o vetor  $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$  pertence a um subconjunto aberto  $\Omega$  de  $R^n$  e  $\Psi$  é o vetor de parâmetros. Esse vetor  $\omega$  expressa um esquema de pesos individuais e o logaritmo da função de verossimilhança clássico  $L(\Psi)$  equivale a  $L^*(\Psi|\omega)$ , neste caso, quando o vetor n-dimensional  $\omega = \omega_0 = (1, 1, \dots, 1)^T$ . Isto é,

$$l(\Psi) = l(\Psi|\omega_0).$$

Nesse contexto, o vetor  $\hat{\Psi}$  é definido como a estimativa de máxima verossimilhança obtida através de  $L(\Psi)$  e  $\hat{\Psi}_\omega$  a estimativa de máxima verossimilhança sob  $L^*(\Psi|\omega)$ . O objetivo é comparar  $\hat{\Psi}$  e  $\hat{\Psi}_\omega$  quando  $\omega$  varia em  $\Omega$ . A comparação entre  $\hat{\Psi}$  e  $\hat{\Psi}_\omega$  pode ser feita por meio do afastamento da verossimilhança  $LD(\omega)$ , expresso da seguinte forma:

$$LD(\omega) = 2 \left[ L^*(\hat{\Psi}) - L^*(\hat{\Psi}_\omega) \right]. \quad (5)$$

A idéia de Cook (1986) é estudar o comportamento da função  $LD(\boldsymbol{\omega})$  numa vizinhança  $\boldsymbol{\omega}_0$ , que é o ponto em que as duas funções de verossimilhança são iguais. Para isso considerou a seguinte superfície geométrica denominada gráfico de influência:

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ LD(\boldsymbol{\omega}) \end{pmatrix}.$$

Cook sugere, ainda, analisar como a superfície  $\alpha(\boldsymbol{\omega})$  desvia-se de seu plano tangente em  $\boldsymbol{\omega}_0$ . Com a equação (5) pode-se avaliar a influência que pequenas perturbações podem exercer sobre os componentes do modelo tais como estimativas dos parâmetros e outros resultados da análise. Interesse particular está na direção que produz a maior influência local. Essa direção  $\mathbf{d}_{max}$  é o autovetor normalizado correspondente ao maior autovalor  $C_{\mathbf{d}_{max}}$  da matriz  $F = \Delta^T \ddot{\mathbf{L}}^{-1} \Delta$ , em que  $\ddot{\mathbf{L}} = \frac{\partial^2 l(\boldsymbol{\Psi})}{\partial(\boldsymbol{\Psi})\partial(\boldsymbol{\Psi})^T}$ . e a  $i$ -ésima linha da matriz  $\Delta$  pode ser escrita na forma

$$\Delta_i^T = \left[ \frac{\partial^2 l(\boldsymbol{\Psi}|\boldsymbol{\omega})}{\partial\omega_i\partial\sigma}, \frac{\partial^2 l(\boldsymbol{\Psi}|\boldsymbol{\omega})}{\partial\omega_i\partial\beta_1}, \dots, \frac{\partial^2 l(\boldsymbol{\Psi}|\boldsymbol{\omega})}{\partial\omega_i\partial\beta_p} \right],$$

Com esse autovetor  $\mathbf{d}_{max}$ , identificamos as observações mais influentes para o esquema de perturbação em análise.

### 2.3.2 Influência Global

Outra forma de analisar a sensibilidade no modelo é a partir da metodologia de deleção de casos, que se baseia na diferença entre as estimativas de máxima verossimilhança relativas à amostra completa e às estimativas obtidas da amostra reduzida por exclusão de um indivíduo, como em Cook (1977). Caso-deleção é um enfoque para estudar o efeito de retirar o  $i$ -ésimo indivíduo de um conjunto de dados. Esta metodologia foi originalmente desenvolvida para modelos Normais lineares mas atualmente estendida para diversas classes de modelos.

Para variáveis aleatórias contínuas  $T_1, T_2$  com distribuição dada pelos modelos propostos, o logaritmo da função de verossimilhança com vetor de parâmetros  $\boldsymbol{\Psi}$  é denotado por  $l(\boldsymbol{\Psi})$ . Seja  $\hat{\boldsymbol{\Psi}}_{(i)}$  o estimador de máxima verossimilhança de  $\boldsymbol{\Psi}$  usando  $l_{(i)}(\boldsymbol{\Psi})$ , em que  $i$  representa o  $i$ -ésimo indivíduo deletado. Para avaliar a influência do  $i$ -ésimo indivíduo no estimador de máxima verossimilhança de  $\hat{\boldsymbol{\Psi}}$ , deve-se comparar a diferença entre  $\hat{\boldsymbol{\Psi}}_{(i)}$  e  $\hat{\boldsymbol{\Psi}}$ .

Se a deleção de um indivíduo provoca variações nas estimativas, maior atenção deve ser dada a esse indivíduo. Se  $\hat{\Psi}_{(i)}$  está distante de  $\hat{\Psi}$ , então o  $i$ -ésimo indivíduo é considerado como uma observação influente.

Uma primeira medida de influência global é definida como a norma padronizada de  $\hat{\Psi}_{(i)} - \hat{\Psi}$ , e é conhecida também como a distância de Cook generalizada.

$$CD_i(\Psi) = (\hat{\Psi}_{(i)} - \hat{\Psi})^T [-\ddot{\mathbf{L}}(\Psi)]^{-1} (\hat{\Psi}_{(i)} - \hat{\Psi}).$$

Considerando uma partição do vetor de parâmetros  $\Psi = (\Psi_1^T, \Psi_2^T)$  uma alternativa é usar os valores de  $CD_i(\Psi_1)$  e  $CD_i(\Psi_2)$  que revelam o impacto do  $i$ -ésimo caso nas estimativas de  $\Psi_1$  e  $\Psi_2$  respectivamente. Ver, por exemplo, Zhu e Zhang (2004).

A seguir, apresenta-se outra medida popular para verificar a existência de pontos influentes no modelo, conhecida como a distância da verossimilhança, dada por:

$$LD_i(\Psi) = 2 \left\{ l(\hat{\Psi}) - l_{(i)}(\hat{\Psi}) \right\}.$$

### 2.3.3 Análise de Resíduos

Uma etapa imprescindível da análise de regressão é a validação do modelo, no sentido de avaliar a qualidade dessa aproximação. A análise de resíduo tem como objetivo detectar pontos mal ajustados ou aberrantes e avaliar indícios de afastamentos das suposições adotadas para o modelo, entre estas, a adequação das distribuições propostas para as variáveis resposta.

As técnicas gráficas utilizando resíduos são freqüentemente adotadas para a análise de diagnóstico. O uso de envelopes simulados, por exemplo, proposto por Atkinson (1981) inicialmente para o modelo de regressão com erros normais, permite avaliar a qualidade do ajuste do modelo postulado.

Comumente utilizado em análise de sobrevivência, o resíduo de Martingale tem a forma:

$$M_{ik} = \delta_{ik} + \log(\hat{S}_k(y_{ik})), \quad (6)$$

em que  $\hat{S}_k$  é uma estimativa da função de sobrevivência marginal para o  $k$ -ésimo tempo.

Para tornar este resíduo simétrico com média zero e variância 1 uma transformação



é realizada sendo o resíduo transformado chamado de resíduo Deviance e é definido por:

$$Dev_{ik} = \text{sinal}(M_{ik})[-2\{M_{ik} + \delta_{ik} \log(\delta_{ik} - M_{ik})\}]^{1/2} \quad (7)$$

No contexto da utilização destas medidas de resíduo para os modelos bivariados em estudo, estas medidas podem ser calculadas marginalmente para cada variável resposta.

### 3 MATERIAL E MÉTODOS

Nesta seção são descritos os conjuntos de dados que são utilizados para testar a aplicabilidade dos modelos propostos e também a metodologia para escolha, aplicação e validação destes modelos.

#### 3.1 Material

##### 3.1.1 Dados de retinopatia

O primeiro conjunto de dados é referente a observação de tempos até a ocorrência de retinopatia leve e grave em pacientes com diabetes melitus diagnosticados no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto(HCFMRP-USP), mais detalhes ver Achcar et al (2005). A Retinopatia Diabética é caracterizada por alterações vasculares causadas por altos índices de glicemia. São lesões que aparecem na retina, podendo causar pequenos sangramentos e, como consequência perda da acuidade visual. Exames de rotina podem detectar anormalidades em estágios primários, possibilitando o tratamento na fase inicial do problema. Tem-se como variáveis resposta de interesse os dois tempos até o diagnóstico em cada indivíduo de retinopatia leve e grave. As covariáveis sexo, cor e outras referentes à presença de hipertensão arterial, neuropatia e nefropatia dos 71 pacientes observados são consideradas na modelagem. A presença de informações parciais, ou censuras nas observações justifica a utilização de técnicas de análise de sobrevivência. Seguem as descrições das variáveis disponíveis para análise:

$t_{i1}$  = tempo em anos até retinopatia leve

$t_{i2}$  = tempo em anos até retinopatia grave

$\delta_{i1}$  = indicador de censura do evento 1

$\delta_{i2}$  = indicador de censura do evento 2

$x_{i1}$  = sexo do paciente (0:masculino, 1:feminino)

$x_{i2}$  = cor da pele(0:branca, 1:não branca)

$x_{i3}$  = presença de hipertensão (0:não ,1:sim)

$x_{i4}$  = nefropatia (0:não ,1:sim)

$x_{i5}$  = neuropatia (0:não ,1:sim)

### 3.1.2 Dados de pacientes com insuficiência renal

O segundo conjunto de dados é referente a observações de tempos até a ocorrência de dois eventos distintos de infecção em pacientes com insuficiência renal. Estes pacientes utilizam máquinas de diálise portáteis. As infecções ocorrem no ponto onde um cateter é inserido e quando uma infecção ocorre o cateter deve ser removido até que a infecção seja curada para que então o cateter seja reinserido. São registrados os tempos entre a inserção do cateter e a ocorrência de uma infecção. Podem ocorrer diversas infecções nos pacientes, tendo registradas no conjunto de dados informações referentes a apenas duas infecções. Muitas vezes o cateter precisa ser removido por motivos diferentes de infecções e portanto ocorrem censuras a direita. São consideradas variáveis resposta de interesse os tempos entre inserções do cateter e ocorrência de infecção. As covariáveis idade, sexo e referentes a presença de doenças dos tipos GN, AN e PKD dos 38 pacientes observados são consideradas na modelagem. A presença de informações censuradas demanda e justifica a utilização de técnicas de análise de sobrevivência. Seguem as descrições das variáveis disponíveis para análise:

$t_{i1}$  = tempo em semanas até infecção 1

$t_{i2}$  = tempo em semanas até infecção 2

$\delta_{i1}$  = indicador de censura do evento 1

$\delta_{i2}$  = indicador de censura do evento 2

$x_{i1}$  = idade do paciente em anos

$x_{i2}$  = sexo do paciente (0:masculino, 1:feminino)

$x_{i3}$  = presença de doença tipo GN (0:não ,1:sim)

$x_{i4}$  = presença de doença tipo AN (0:não ,1:sim)

$x_{i5}$  = presença de doença tipo PKD (0:não ,1:sim)

### 3.2 Métodos

Com o objetivo de explicar os tempos até a ocorrência de eventos distintos em um mesmo indivíduo, os modelos de regressão bivariados obtidos através de cópulas são utilizados. As cópulas de Frank e Clayton apresentadas são utilizadas para reunir as funções de sobrevivência marginais escolhidas.

### 3.3 Modelos de Regressão Bivariados

Uma vez definidos os modelos de locação e escala que são utilizados como distribuições marginais e as funções de cópulas que relacionarão estas marginais, os modelos de regressão bivariados podem ser escritos.

Supondo  $T_1$  e  $T_2$  variáveis aleatórias com distribuições Weibull e  $Y_k = \log(T_k)$ . A função de sobrevivência para o modelo de regressão log-linear bivariado com  $p$  covariáveis terá sua representação através da Família Frank na forma:

$$S(y_1, y_2) = -\frac{1}{\alpha} \log \left\{ 1 + \frac{(\exp(-\alpha \exp\{-\exp(z_1)\}) - 1)(\exp(-\alpha \exp(-\exp(z_2))) - 1)}{\exp(-\alpha) - 1} \right\} \quad (8)$$

em que  $z_k = \frac{y_k - \mathbf{x}^T \boldsymbol{\beta}_k}{\sigma_k}$ ,  $\boldsymbol{\Psi} = (\alpha, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ ,  $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^T, \sigma_k)^T$  e  $\boldsymbol{\beta}_k^T = (\beta_{0k}, \beta_{1k}, \dots, \beta_{pk})$  para  $k = 1, 2$ .

Deve-se observar que as funções marginais podem ser quaisquer funções de sobrevivência e não necessariamente iguais, podendo dessa forma serem criados modelos bivariados com combinação de marginais distintas, sendo esta uma das características mais atrativas desse tipo de modelagem por permitir grande flexibilidade.

Para a família de Clayton, a função de sobrevivência para o modelo de regressão log-linear bivariado sob as suposições marginais citadas terá sua representação dada por:

$$S(y_1, y_2) = \left( \exp \left\{ -\exp(z_1) \right\}^{-\alpha} + \exp \left\{ -\exp(z_2) \right\}^{-\alpha} - 1 \right)^{-\frac{1}{\alpha}}.$$

Uma vez definida a forma dos modelos, segue a etapa inferencial do trabalho.

### 3.4 Estratégias de Inferência

Duas estratégias de estimação de parâmetros são utilizados: máxima verossimilhança e jackknife.

Como apresentado na equação (1) a função de verossimilhança é adaptada nos casos de presença de censuras nos dados. A função de verossimilhança tem uma parte com a contribuição dos elementos que tiveram a falha observada e uma parte relativa aos elementos com informação incompleta. No caso de respostas bivariadas, a função de verossimilhança é mais uma vez adaptada para contemplar todos os possíveis casos de censuras nos dados.

Os possíveis casos de censura aleatória em respostas bivariadas são quatro e representam a observação dos dois eventos de interesse, a observação do evento 1 e observação censurada do evento 2, o oposto que é a censura do evento 1 e observação do evento 2 e o quarto caso que representa observações censuradas para ambos eventos. Neste contexto, considera-se uma amostra observada  $(y_{i1}, y_{i2}, \delta_{i1}, \delta_{i2}, \mathbf{x}_i)$  em que  $y_{ik}$  é o logaritmo do tempo até a observação do  $k$ -ésimo evento,  $\delta_{ik}$  é a respectiva variável indicadora de censura e  $\mathbf{x}_i$  o vetor de covariáveis associados ao  $i$ -ésimo indivíduo,  $k = 1, 2$  e  $i = 1, \dots, n$ . Na utilização de modelo bivariado considera-se que o logaritmo da função de verossimilhança pode ser descrito de uma forma geral como em Lawless (2002):

$$l(\Psi) = \sum_{i=1}^n \delta_{i1}\delta_{i2} \log \left\{ f(y_{i1}, y_{i2}) \right\} + \delta_{i1}(1 - \delta_{i2}) \log \left\{ \frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right\} + \\ (1 - \delta_{i1})\delta_{i2} \log \left\{ \frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right\} + (1 - \delta_{i1})(1 - \delta_{i2}) \log \left\{ S(y_{i1}, y_{i2}) \right\}$$

em que  $S(y_{i1}, y_{i2})$  é a função de sobrevivência bivariada obtida através de uma cópula,  $f(y_{i1}, y_{i2})$  é a função de densidade conjunta obtida por  $\frac{\partial^2 S(y_{i1}, y_{i2})}{\partial y_{i1} \partial y_{i2}}$ ,  $\Psi = (\alpha, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$  e  $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^T, \sigma_k)$ ,  $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \dots, \beta_{pk})$  para  $k = 1, 2$ .

Para a cópula de Frank com suposições marginais de distribuições valor extremo para os logaritmos dos tempos até os eventos 1 e 2 tem-se  $S_1(y_1)$  e  $S_2(y_2)$  como definido em (2) e:

$$S(y_{i1}, y_{i2}) = -\frac{1}{\alpha} \log \left[ 1 + \frac{(\exp(-\alpha S_1(y_{i1})) - 1)(\exp(-\alpha S_2(y_{i2})) - 1)}{\exp(-\alpha) - 1} \right].$$

$$\begin{aligned}
\frac{-\partial S(y_{i1}, y_{i1})}{\partial y_{i1}} &= \frac{\exp(z_{i1})S_1(y_{i1}) \exp\{-\alpha S_1(y_{i1})\} \left( \exp\{-\alpha S_2(y_{i2})\} - 1 \right)}{\sigma_1 \left\{ \exp(-\alpha) - 1 \right\} \exp\{-\alpha S(y_{i1}, y_{i2})\}}. \\
\frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} &= \frac{\exp(z_{i2})S_2(y_{i2}) \exp\{-\alpha S_2(y_{i2})\} \left( \exp\{-\alpha S_1(y_{i1})\} - 1 \right)}{\sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp\{-\alpha S(y_{i1}, y_{i2})\}}. \\
f(y_{i1}, y_{i2}) &= \frac{\exp(z_{i1})S_1(y_{i1}) \exp\{-\alpha S_1(y_{i1})\} \alpha \exp(z_{i2})S_2(y_{i2}) \exp\{-\alpha S_2(y_{i2})\}}{\sigma_1 \sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp\{-\alpha S(y_{i1}, y_{i2})\}} \times \\
&\quad \left( 1 + \frac{\left( \exp\{-\alpha S_2(y_{i2})\} - 1 \right) \left( \exp\{-\alpha S_1(y_{i1})\} - 1 \right)}{\left\{ \exp(-\alpha) - 1 \right\} \exp\{-\alpha S(y_{i1}, y_{i2})\}} \right)
\end{aligned}$$

Para a cópula de Clayton com suposições marginais de distribuições valor extremo para os logaritmos dos tempos até os eventos 1 e 2 tem-se  $S_1(y_1)$  e  $S_2(y_2)$  definidos em (2) e:

$$\begin{aligned}
S(y_{i1}, y_{i2}) &= \left[ S_1(y_{i1})^{-\alpha} + S_2(y_{i2})^{-\alpha} - 1 \right]^{-\frac{1}{\alpha}}. \\
\frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} &= \frac{S(y_{i1}, y_{i2}) S_1(y_{i1})^{-\alpha} \exp(z_{i1})}{\sigma_1 S(y_{i1}, y_{i2})^{-\alpha}}. \\
\frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} &= \frac{S(y_{i1}, y_{i2}) S_2(y_{i2})^{-\alpha} \exp(z_{i2})}{\sigma_2 S(y_{i1}, y_{i2})^{-\alpha}}. \\
f(y_{i1}, y_{i2}) &= \frac{S(y_{i1}, y_{i2}) S_1(y_{i1})^{-\alpha} \exp(z_{i1}) S_2(y_{i2})^{-\alpha} \exp(z_{i2})}{\sigma_1 \sigma_2 S(y_{i1}, y_{i2})^{-2\alpha}} (1 + \alpha).
\end{aligned}$$

Quando se dá a utilização de cópulas para criação de modelos bivariados, a estimação dos parâmetros é normalmente realizada em duas etapas. Na primeira etapa, os tempos são considerados independentes e os parâmetros de suas distribuições são estimados em cada marginal. As estimativas obtidas são utilizadas na segunda etapa, quando o parâmetro de associação da cópula é, então, estimado. Esse tipo de abordagem é coerente quando o objetivo principal do estudo é a estimação do parâmetro de associação. Em casos em que os coeficientes dos modelos de regressão são o foco, essa abordagem marginal não acrescenta mais informações às estimativas do que a utilização de modelos individuais para cada tempo de falha. Será utilizada, portanto uma estimação conjunta dos coeficientes de regressão e do parâmetro  $\alpha$ .

### 3.4.1 Estimadores de máxima verossimilhança

Estimadores de máxima verossimilhança para o vetor de parâmetros  $\Psi = (\alpha, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$  podem ser obtidos maximizando a função de verossimilhança apresentada em função dos parâmetros. Os softwares Ox (subrotina MAXBFGS) e R (função OPTIM) podem ser utilizados para obter estimativas para os parâmetros. Para fazer inferências, intervalos de confiança e testes de hipóteses a distribuição assintótica dos EMV's é utilizada.

As propriedades assintóticas dos estimadores de máxima verossimilhança são necessárias para construir intervalos de confiança, e testar hipóteses sobre os coeficientes do modelo. Utiliza-se do fato que  $\hat{\Psi}$  tem distribuição normal assintótica multivariada sob certas condições de regularidade com média  $\Psi$  e matriz de variâncias e covariâncias  $\mathbf{I}^{-1}(\Psi)$ , dado que  $\mathbf{I}(\Psi) = -E[\ddot{L}(\Psi)]$ , tal que,  $\ddot{L}(\Psi) = \left\{ \frac{\partial^2 l(\Psi)}{\partial \Psi \partial \Psi^T} \right\}$ , isto é,

$$\hat{\Psi} \sim N_{2p+5}(\Psi, \mathbf{I}^{-1}(\Psi))$$

Visto que o cálculo de  $\mathbf{I}(\Psi)$  (matriz de informação de Fischer) não é possível devido à presença de censura, pode-se utilizar alternativamente a matriz  $-\ddot{L}(\Psi)$  avaliada em  $\Psi = \hat{\Psi}$  denominada matriz de informação observada, que é uma estimativa consistente de  $\mathbf{I}(\Psi)$ .

Assim, o intervalo de confiança assintótico para  $\beta_{jk}$  em que  $j = 0, 1, 2, \dots, p$  e  $k = 1, 2$ , considerando  $(1 - \alpha)100\%$  de confiança é dado por

$$\hat{\beta}_{jk} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_{jk})}.$$

Para a construção de um teste de hipótese para os parâmetros também utiliza-se a diagonal principal da matriz  $-\ddot{L}^{-1}(\Psi)$  como uma estimativa para a variância dos estimadores, logo, a estatística para testar as hipóteses

$$\begin{cases} H_0 : \beta_{jk} = \beta_{jk0} \\ H_a : \beta_j \neq \beta_{jk0} \end{cases}$$

é dada por:  $z = \frac{\hat{\beta}_{jk} - \beta_{jk0}}{\sqrt{\widehat{Var}(\hat{\beta}_{jk})}} \sim N(0, 1)$ .

### 3.4.2 Estimadores de Jackknife

A metodologia para a obtenção destes estimadores consiste na subdivisão da amostra e estimações sucessivas em cada subamostra. Neste trabalho é utilizada a subdivisão de uma amostra de tamanho  $n$  em  $n$  subamostras com  $(n-1)$  elementos. Cada uma das subamostras representa os dados após a eliminação de uma linha na matriz de dados.

Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória e deseja-se estimar um parâmetro populacional  $\theta$  com um estimador  $\hat{\theta}$ . Para cada uma das subamostras obtem-se as estimativas  $\hat{\theta}_{-i}$ ,  $i = 1, 2, \dots, n$ , denominadas estimativas parciais. A partir dessas estimativas são calculados os "pseudo-valores" que têm a forma:

$$\hat{\theta}_{*i} = n\hat{\theta} - (n-1)\hat{\theta}_{-i}, i = 1, 2, \dots, n$$

O estimador de Jackknife é a média dos pseudo-valores.

$$\hat{\theta}_* = \frac{\sum_{i=1}^n \hat{\theta}_{*i}}{n}$$

Tukey propôs que os  $\hat{\theta}_{*j}$  são independentes e identicamente distribuídos. Dessa forma, pode-se dizer que  $Var(\hat{\theta}_*)$  pode ser estimada por:

$$S_T^2 = \frac{\sum_{i=1}^n (\hat{\theta}_{*i} - \hat{\theta}_*)^2}{n(n-1)}.$$

Como consequência deste resultado tem-se que:

$$t_T = \frac{(\hat{\theta}_* - \theta)}{S_T} \sim t_{(n-1)}.$$

Uma propriedade dos estimadores de Jackknife é a possibilidade de eliminar o viés de ordem  $n^{-1}$  de um estimador. O atrativo principal, entretanto, para a utilização do método nesse trabalho é a simplicidade na estimação da variância dos estimadores assim como o comportamento aproximadamente normal dos mesmos.

Comparações entre as estimativas pelos diferentes métodos abordados serão utilizadas e visam detectar diferenças significativas entre as diferentes metodologias de estimação.

Verificações quanto a robustez dos resultados obtidos com os modelos propostos são feitas mediante uma análise de sensibilidade baseada nas técnicas de influência global e local.



### 3.4.3 Influência local

São considerados três tipos de perturbações nos modelos: Ponderação de casos, perturbação da variável resposta e a perturbação de variável explicativa.

#### 1. Ponderação de casos

Considere o vetor de ponderação  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^T$ . Neste caso o logaritmo da função de verossimilhança perturbado é expresso por

$$l(\boldsymbol{\Psi}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \left[ \delta_{i1} \delta_{i2} \log \left\{ f(y_{i1}, y_{i2}) \right\} + \delta_{i1} (1 - \delta_{i2}) \log \left\{ \frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right\} + \right. \\ \left. (1 - \delta_{i1}) \delta_{i2} \log \left\{ \frac{-\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right\} + (1 - \delta_{i1})(1 - \delta_{i2}) \log \left\{ S(y_{i1}, y_{i2}) \right\} \right]$$

em que  $0 \leq \omega_i \leq 1$  e  $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ . Os elementos da matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\alpha, \boldsymbol{\Delta}_{\boldsymbol{\theta}_1}, \boldsymbol{\Delta}_{\boldsymbol{\theta}_2})^T$  definidos na Seção 2.3.1 são obtidos numericamente.

#### 2. Perturbação das variáveis resposta

Neste tipo de perturbação as variáveis resposta  $y_{i1}$  e  $y_{i2}$  são modificadas como  $y_{i1w} = y_{i1} + \omega_i u_{y1}$  e  $y_{i2w} = y_{i2} + \omega_i u_{y2}$ , sendo  $u_{y_k}$  são fatores de escala que podem ser, por exemplo, a estimativa do desvio padrão da variável  $Y_k$  e  $\omega_i \in \mathbf{R}$ . Neste caso, o logaritmo da função de verossimilhança perturbada fica dado por:

$$l(\boldsymbol{\Psi}|\boldsymbol{\omega}) = \sum_{i=1}^n \delta_{i1} \delta_{i2} \log \left\{ f^*(y_{i1}, y_{i2}) \right\} + \delta_{i1} (1 - \delta_{i2}) \log \left\{ - \left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right]^* \right\} + \\ (1 - \delta_{i1}) \delta_{i2} \log \left\{ - \left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^* \right\} + (1 - \delta_{i1})(1 - \delta_{i2}) \log \left\{ S^*(y_{i1}, y_{i2}) \right\},$$

em que para a cópula de Frank define-se:

$$S^*(y_{i1}, y_{i2}) = -\frac{1}{\alpha} \log \left[ 1 + \frac{(\exp(-\alpha S_1(y_{i1}^*)) - 1)(\exp(-\alpha S_2(y_{i2}^*)) - 1)}{\exp(-\alpha) - 1} \right].$$

$$\left[ \frac{\partial S(y_{i1}, y_{i1})}{\partial y_{i1}} \right]^* = \frac{\exp(z_{i1}^*) S_1(y_{i1}^*) \exp \left\{ -\alpha S_1(y_{i1}^*) \right\} \left( \exp \left\{ -\alpha S_2(y_{i2}^*) \right\} - 1 \right)}{\sigma_1 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S(y_{i1}^*, y_{i2}^*) \right\}}.$$

$$\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^* = \frac{\exp(z_{i2}^*) S_2(y_{i2}^*) \exp \left\{ -\alpha S_2(y_{i2}^*) \right\} \left( \exp \left\{ -\alpha S_1(y_{i1}^*) \right\} - 1 \right)}{\sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S(y_{i1}^*, y_{i2}^*) \right\}}.$$

$$f^*(y_{i1}, y_{i2}) = \frac{\exp(z_{i1}^*) S_1(y_{i1}^*) \exp \left\{ -\alpha S_1(y_{i1}^*) \right\} \alpha \exp(z_{i2}^*) S_2(y_{i2}^*) \exp \left\{ -\alpha S_2(y_{i2}^*) \right\}}{\sigma_1 \sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S(y_{i1}^*, y_{i2}^*) \right\}} \\ \times \left\{ 1 + \frac{[\exp \{-\alpha S_2(y_{i2}^*)\} - 1][\exp \{-\alpha S_1(y_{i1}^*)\} - 1]}{\left\{ \exp(-\alpha) - 1 \right\} \exp \{-\alpha S(y_{i1}^*, y_{i2}^*)\}} \right\}.$$

Para a cópula de Clayton tem-se

$$S^*(y_{i1}, y_{i2}) = \left[ S_1(y_{i1}^*)^{-\alpha} + S_2(y_{i2}^*)^{-\alpha} - 1 \right]^{-\frac{1}{\alpha}}.$$

$$-\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right]^* = \frac{S^*(y_{i1}, y_{i2}) S_1(y_{i1}^*)^{-\alpha} \exp(z_{i1}^*)}{\sigma_1 S^*(y_{i1}, y_{i2})^{-\alpha}}.$$

$$-\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^* = \frac{S^*(y_{i1}, y_{i2}) S_2(y_{i2}^*)^{-\alpha} \exp(z_{i2}^*)}{\sigma_2 S^*(y_{i1}, y_{i2})^{-\alpha}}.$$

$$f^*(y_{i1}, y_{i2}) = \frac{S^*(y_{i1}, y_{i2}) S_1(y_{i1}^*)^{-\alpha} \exp(z_{i1}^*) S_2(y_{i2}^*)^{-\alpha} \exp(z_{i2}^*)}{\sigma_1 \sigma_2 S^*(y_{i1}, y_{i2})^{-2\alpha}} (1 + \alpha),$$

em que  $y_{i1}^* = y_{i1} + \omega_i u_{y1}$ ,  $y_{i2}^* = y_{i2} + \omega_i u_{y2}$ ,  $z_{i1}^* = (y_{i1}^* - \mathbf{x}^T \boldsymbol{\beta}_1) / \sigma_1$  e  $z_{i2}^* = (y_{i2}^* - \mathbf{x}^T \boldsymbol{\beta}_2) / \sigma_2$ . Os elementos da matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\alpha, \boldsymbol{\Delta}_{\boldsymbol{\theta}_1}, \boldsymbol{\Delta}_{\boldsymbol{\theta}_2})^T$  são obtidos numericamente.

### 3. Perturbação de uma variável explicativa

Considere agora uma perturbação aditiva em uma covariável contínua particular  $X_t$  que pode ser dada por  $x_{it\omega} = x_{it} + \omega_i u_x$  em que  $u_x$  é um fator de escala e  $\omega_i \in \mathbf{R}$ . Neste caso o logaritmo da função de verossimilhança perturbada pode ser expresso como

$$l(\boldsymbol{\Psi} | \boldsymbol{\omega}) = \sum_{i=1}^n \delta_{i1} \delta_{i2} \log \left\{ f^\ddagger(y_{i1}, y_{i2}) \right\} + \delta_{i1} (1 - \delta_{i2}) \log \left\{ - \left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right]^\ddagger \right\} + \\ (1 - \delta_{i1}) \delta_{i2} \log \left\{ - \left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^\ddagger \right\} + (1 - \delta_{i1}) (1 - \delta_{i2}) \log \left\{ S^\ddagger(y_{i1}, y_{i2}) \right\}$$

em que para a cópula de Frank define-se:

$$S^\ddagger(y_{i1}, y_{i2}) = -\frac{1}{\alpha} \log \left[ 1 + \frac{(\exp(-\alpha S_1^\ddagger(y_{i1})) - 1)(\exp(-\alpha S_2^\ddagger(y_{i2})) - 1)}{\exp(-\alpha) - 1} \right].$$

$$\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right]^\ddagger = \frac{\exp(z_{i1}^\ddagger) S_1^\ddagger(y_{i1}) \exp \left\{ -\alpha S_1^\ddagger(y_{i1}) \right\} \left( \exp \left\{ -\alpha S_2^\ddagger(y_{i2}) \right\} - 1 \right)}{\sigma_1 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S^\ddagger(y_{i1}, y_{i2}) \right\}}$$

$$\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^\ddagger = \frac{\exp(z_{i2}^\ddagger) S_2^\ddagger(y_{i2}) \exp \left\{ -\alpha S_2^\ddagger(y_{i2}) \right\} \left( \exp \left\{ -\alpha S_1^\ddagger(y_{i1}) \right\} - 1 \right)}{\sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S^\ddagger(y_{i1}, y_{i2}) \right\}}$$

$$f^\ddagger(y_{i1}, y_{i2}) = \frac{\exp(z_{i1}^\ddagger) S_1^\ddagger(y_{i1}) \exp \left\{ -\alpha S_1^\ddagger(y_{i1}) \right\} \alpha \exp(z_{i2}^\ddagger) S_2^\ddagger(y_{i2}) \exp \left\{ -\alpha S_2^\ddagger(y_{i2}) \right\}}{\sigma_1 \sigma_2 \left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S^\ddagger(y_{i1}, y_{i2}) \right\}} \\ \times \left\{ 1 + \frac{\left[ \exp \left\{ -\alpha S_2^\ddagger(y_{i2}) \right\} - 1 \right] \left[ \exp \left\{ -\alpha S_1^\ddagger(y_{i1}) \right\} - 1 \right]}{\left\{ \exp(-\alpha) - 1 \right\} \exp \left\{ -\alpha S^\ddagger(y_{i1}, y_{i2}) \right\}} \right\}$$

Para a cópula de Clayton tem-se:

$$S^\ddagger(y_{i1}, y_{i2}) = \left[ S_1^\ddagger(y_{i1})^{-\alpha} + S_2^\ddagger(y_{i2})^{-\alpha} - 1 \right]^{-\frac{1}{\alpha}}$$

$$-\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i1}} \right]^\ddagger = \frac{S^\ddagger(y_{i1}, y_{i2}) S_1^\ddagger(y_{i1})^{-\alpha} \exp(z_{i1}^\ddagger)}{\sigma_1 S^\ddagger(y_{i1}, y_{i2})^{-\alpha}}$$

$$-\left[ \frac{\partial S(y_{i1}, y_{i2})}{\partial y_{i2}} \right]^\ddagger = \frac{S^\ddagger(y_{i1}, y_{i2}) S_2^\ddagger(y_{i2})^{-\alpha} \exp(z_{i2}^\ddagger)}{\sigma_2 S^\ddagger(y_{i1}, y_{i2})^{-\alpha}}$$

$$f^\ddagger(y_{i1}, y_{i2}) = \frac{S^\ddagger(y_{i1}, y_{i2}) S_1^\ddagger(y_{i1})^{-\alpha} \exp(z_{i1}^\ddagger) S_2^\ddagger(y_{i2})^{-\alpha} \exp(z_{i2}^\ddagger)}{\sigma_1 \sigma_2 S^\ddagger(y_{i1}, y_{i2})^{-2\alpha}} (1 + \alpha).$$

em que  $z_{ik}^\ddagger = \{y_{ik} - \mathbf{x}_i^{\ddagger T} \boldsymbol{\beta}_k\} / \sigma_k$  e  $\mathbf{x}_i^{\ddagger T} = \beta_{0k} + \beta_{1k} x_{i1} + \beta_{2k} x_{i2} + \dots + \beta_{tk} [x_{it} + \omega_i u_x] + \dots + \beta_{pk} x_{ip}$  e  $S_k^\ddagger(y_{ik}) = \exp \left\{ -\exp(z_{ik}^\ddagger) \right\}$ .

Os elementos da matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\alpha, \boldsymbol{\Delta}_{\boldsymbol{\theta}_1}, \boldsymbol{\Delta}_{\boldsymbol{\theta}_2})^T$  são obtidos numericamente.

### 3.4.4 Influência global

Para realização da análise de influência global são calculadas, no esquema de deleção de casos, as medidas de afastamento da verossimilhança e distância generalizada de Cook. A medida de afastamento da verossimilhança é calculada seguindo a formulação dada em (5).

No contexto da utilização dos modelos de respostas bivariadas a distância generalizada de Cook é calculada de forma marginal para cada resposta, podendo desta forma detectar pontos que influenciam as estimativas marginal ou conjuntamente. Tem-se então que

$$CD_{ik}(\boldsymbol{\theta}_k) = (\hat{\boldsymbol{\theta}}_{k(i)} - \hat{\boldsymbol{\theta}}_k)^T [-\ddot{\mathbf{L}}(\boldsymbol{\theta}_k)]^{-1} (\hat{\boldsymbol{\theta}}_{k(i)} - \hat{\boldsymbol{\theta}}_k), \quad k = 1, 2$$

é a distância generalizada de Cook referente à  $k$ -ésima resposta para o  $i$ -ésimo indivíduo.

### 3.4.5 Análise de resíduos

Para verificar a adequação do modelo aos dados e possíveis indícios de afastamentos sérios das suposições feitas para o erro do modelo proposto, utiliza-se a análise de resíduos. O resíduo de Deviance é calculado marginalmente para cada variável resposta na forma:

$$Dev_{ik} = \text{sinal}(M_{ik}) [-2\{M_{ik} + \delta_{ik} \log(\delta_{ik} - M_{ik})\}]^{1/2}, \quad i = 1, 2, \dots, n \quad \text{e} \quad k = 1, 2.$$

em que  $M_{ik}$  é o resíduo martingal definido em (6).

As seguintes medidas de resíduos propostas são adaptações dos resíduos martingal e deviance para utilização em modelos de respostas bivariadas em análise de sobrevivência.

$$M_i^* = \frac{(\delta_{i1} + \delta_{i2})}{2} + \log(\hat{S}(y_{i1}, y_{i2}))$$

em que  $\hat{S}(y_{i1}, y_{i2})$ , é a sobrevivência estimada utilizando os modelos propostos.

$$Dev_i^* = \text{sinal}(M_i^*) \left[ -2 \left\{ M_i^* + \frac{(\delta_{i1} + \delta_{i2})}{2} \log \left( \frac{(\delta_{i1} + \delta_{i2})}{2} - M_i^* \right) \right\} \right]^{1/2}, \quad i = 1, 2, \dots, n. \quad (9)$$

### 3.4.6 Impacto das observações influentes

Combinando os resultados das análises de resíduos, influência local e global, algumas observações são selecionadas por destacarem-se das demais nestas análises e serem, portanto, possíveis pontos influentes na modelagem. O impacto dessas observações sobre o modelo é avaliado com o auxílio de tabelas formadas pelas estimativas obtidas a partir de sub-amostras com a retirada destas observações individualmente e em grupo. As tabelas são formadas pelas estimativas para os parâmetros, medidas de mudança relativa e os p-valores. A medida de mudança

relativa foi definida por  $RC_{\alpha_j} = [(\hat{\alpha}_j - \hat{\alpha}_{j(I)})/\hat{\alpha}_j]$ , em que o sub-índice I refere-se ao conjunto de observações a serem retiradas na sub-amostra. Neste contexto, são consideradas influentes aquelas observações que quando retiradas causam alterações significativas tanto nas estimativas pontuais quanto nos p-valores e conseqüentemente nas significâncias das covariáveis.

Com a retirada das observações influentes os modelos são reajustados aos conjuntos de dados e modelos finais são propostos para descrever os fenômenos em questão. A validação destes modelos finais dá-se pela avaliação do comportamento das medidas de resíduo, incluindo a abordagem de envelopes simulados proposta por Atkinson (1985).

### 3.5 Metodologia de simulação

Neste trabalho técnicas de simulação de Monte Carlo são utilizadas para investigar a distribuição empírica das medidas de resíduo propostas quando os modelos apresentados são aplicados. Quando simulações de números aleatórios são realizadas, tem-se o controle sob a distribuição dos dados. Desta forma sabe-se se as suposições utilizadas na modelagem proposta são corretas ou não. Conhecendo o comportamento das medidas de resíduo quando as suposições são corretas pode-se utilizar este conhecimento para validar ou não a utilização dos modelos propostos e as suposições adotadas quando estes são ajustados à conjuntos de dados reais.

Para a realização da simulação, o pacote chamado copula, disponível para R, é utilizado. Com a função *rmvdc* (*random multivariate distribution via copula*) pode-se escolher a função de cópulas e o parâmetro de associação assim como a forma das distribuições marginais e as configurações paramétricas das mesmas para gerar números aleatórios.

Neste estudo foram utilizadas as cópulas de Frank e Clayton com distribuições marginais Weibull para simular respostas bivariadas. Para reproduzir o esquema de dados censurados, além das amostras geradas através das cópulas, são geradas pares independentes de amostras provenientes de distribuição Uniforme  $(0, \lambda)$  para representar os tempos de censurados na amostra. A variável considerada observada é então  $t_{ik} = \min(T_{ik}, C_{ik})$  e associa-se a variável indicadora de censura conforme definido na Seção(2.1). O parâmetro  $\lambda$ , utilizado para gerar os tempos censurados, é escolhido através de tentativa e erro para obtenção do percentual de censuras médio desejado para a amostra.

Foram definidas as seguintes configurações para este estudo de simulação:

1. Com a cópula de Frank são utilizadas combinações do parâmetro de associação assumindo valores (0.3, 6), tamanhos amostrais (30, 50, 100) e percentuais médios de censuras (10%, 20%, 30%).
2. Com a cópula de Clayton são utilizadas combinações do parâmetro de associação assumindo valores (2, 7), tamanhos amostrais (30, 50, 100) e percentuais médios de censuras (10%, 20%, 30%).

Os parâmetros das distribuições marginais Weibull foram mantidos constantes em (3, 5; 6) e (4; 20) para todas as configurações mencionadas.

Além dos tempos de falha e censura, uma covariável com distribuição Uniforme (0, 1) é também simulada. Para cada configuração de simulação são geradas 100 amostras. Uma vez geradas estas amostras, os modelos propostos são ajustados e as medidas de resíduo  $Dev_{ik}$  e  $Dev_i^*$  calculadas. Para cada amostra as medidas de resíduo calculadas são ordenadas e para cada posição são calculadas as médias entre as amostras. Com estas médias são construídos gráficos Normais de probabilidade, pois supõe-se que as medidas de resíduo utilizadas têm distribuição Normal (0, 1) quando as suposições adotadas na modelagem são corretas.

### 3.5.1 Envelopes simulados

Atkinson (1985) sugere que mediante a construção de envelopes simulados pode-se interpretar melhor os gráficos Normais de probabilidade para os resíduos. Os envelopes simulados são bandas de confiança obtidas por método de Monte Carlo a partir do modelo ajustado para avaliar a existência de afastamentos sérios das suposições adotadas. Esses gráficos podem ser utilizados mesmo que os resíduos não tenham distribuição de probabilidade Normal. Neste caso os pontos no gráfico estarão próximos da reta que faz um ângulo de  $45^\circ$  com o eixo das abscissas.

A construção dos envelopes segue os seguintes passos:

1. Ajustar o modelo proposto ao conjunto de dados e calcular os resíduos definidos em (7) e (9).
2. Simular, utilizando o pacote copula do R,  $m$  amostras das variáveis resposta utilizando os parâmetros estimados no passo anterior.

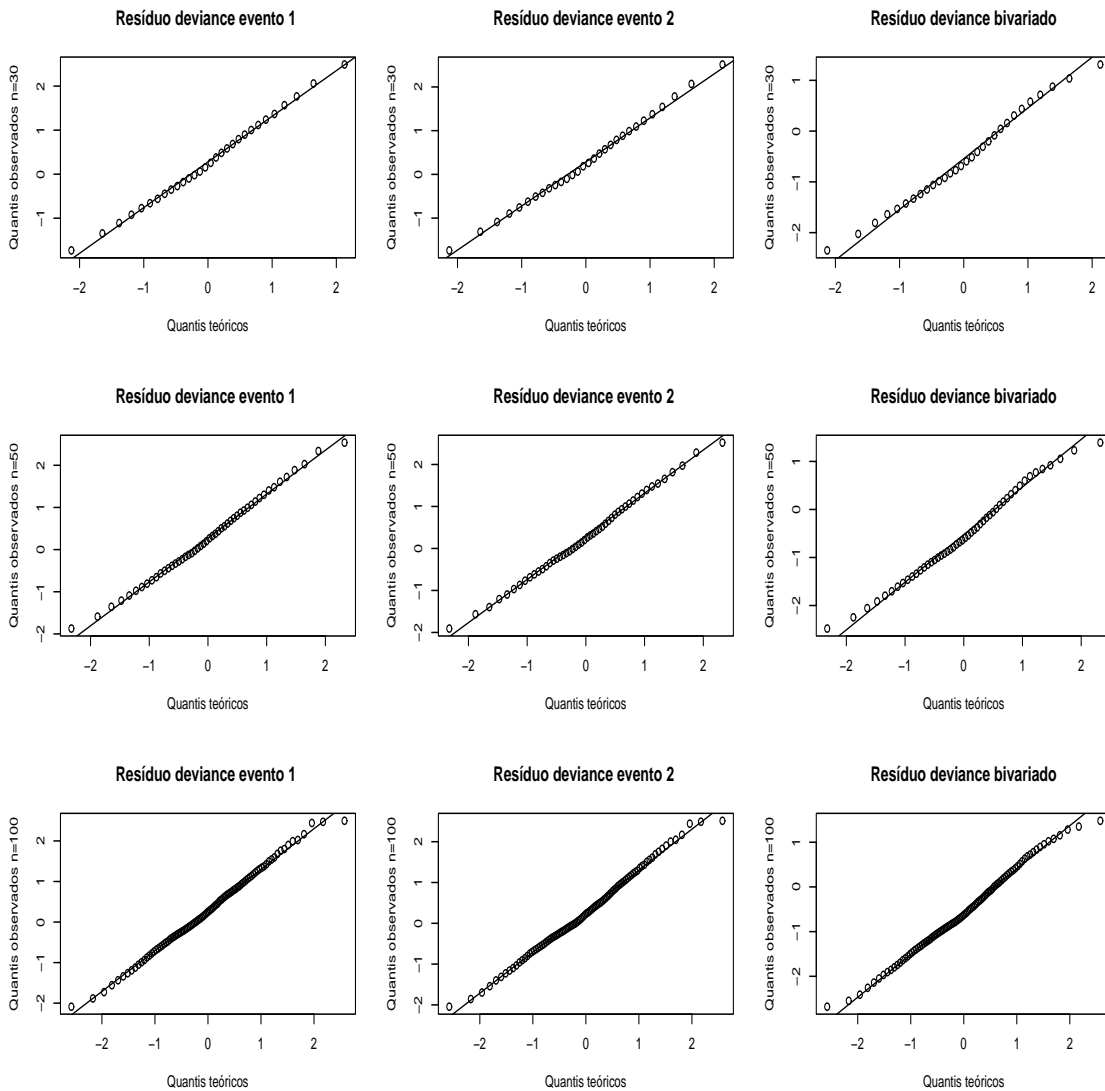
3. Para cada amostra simulada ajustar o modelo proposto e calcular os resíduos  $Dev_{ik}$  e  $Dev_i^*$ .
4. Ordenar os resíduos obtidos para cada amostra.
5. Para cada  $i$ -ésima posição do ordenamento obter o mínimo, a média e o máximo entre as amostras.
6. Incluir no gráfico Normal de probabilidades obtido para os resíduos do passo 1 os mínimos, médias e máximos.

Os mínimos e máximos calculados formam os envelopes. Se o modelo estimado é correto, espera-se que os valores observados estejam dentro das bandas e de forma aleatória. Segundo Atkinson (1985), utilizando-se  $m = 19$  amostras, existe a chance de 5% de que o maior valor absoluto dos resíduos do conjunto de dados esteja fora dos envelopes simulados quando o modelo ajustado for correto. Este resultado também pode ser utilizado para detectar a presença de pontos aberrantes.

## 4 RESULTADOS

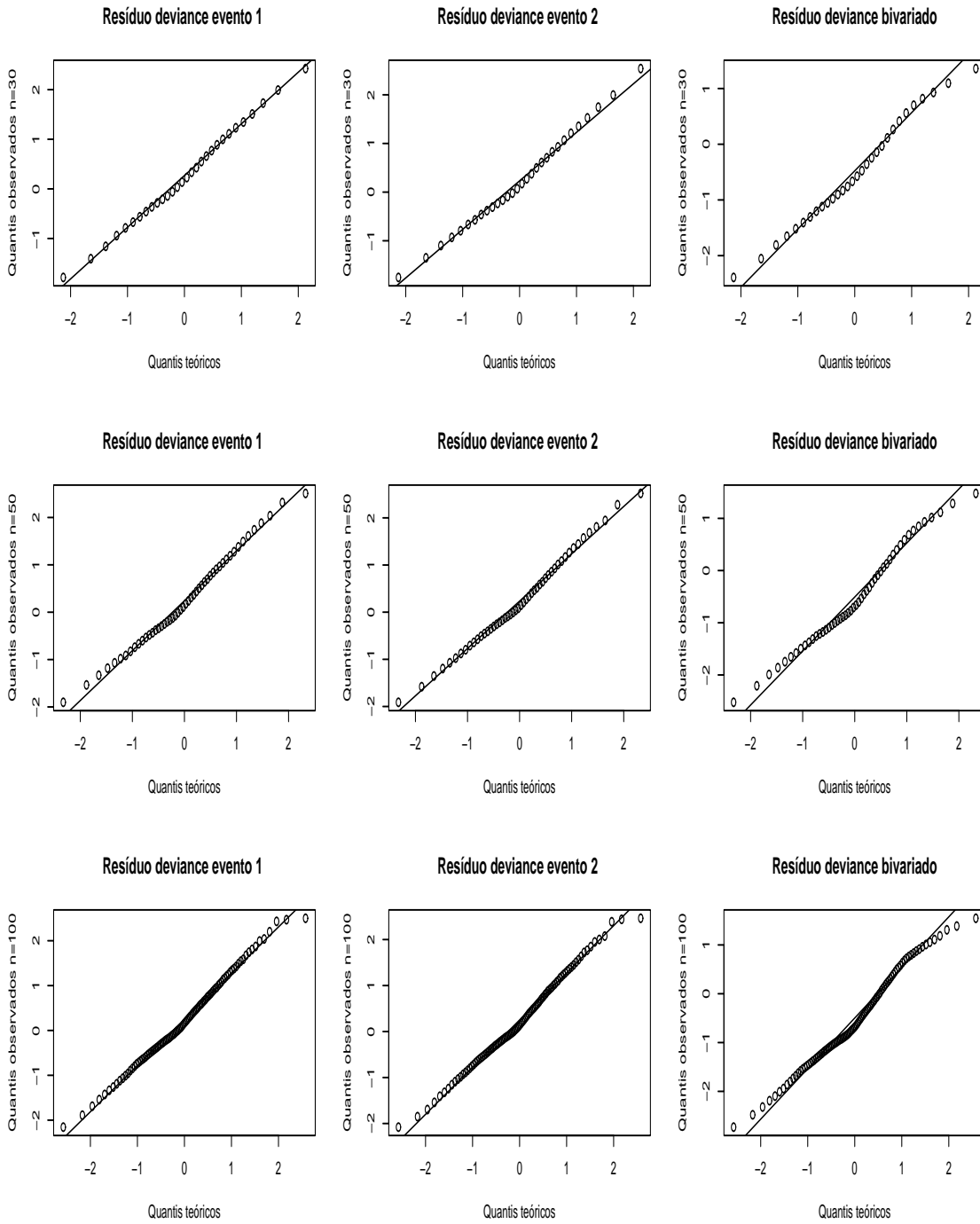
Nesta seção apresenta-se os resultados do estudo de simulação descrito na seção 3.5 e os resultados referentes a aplicação dos modelos propostos aos dois conjuntos de dados citados na seção 3.1.

### 4.1 Resultados das simulações

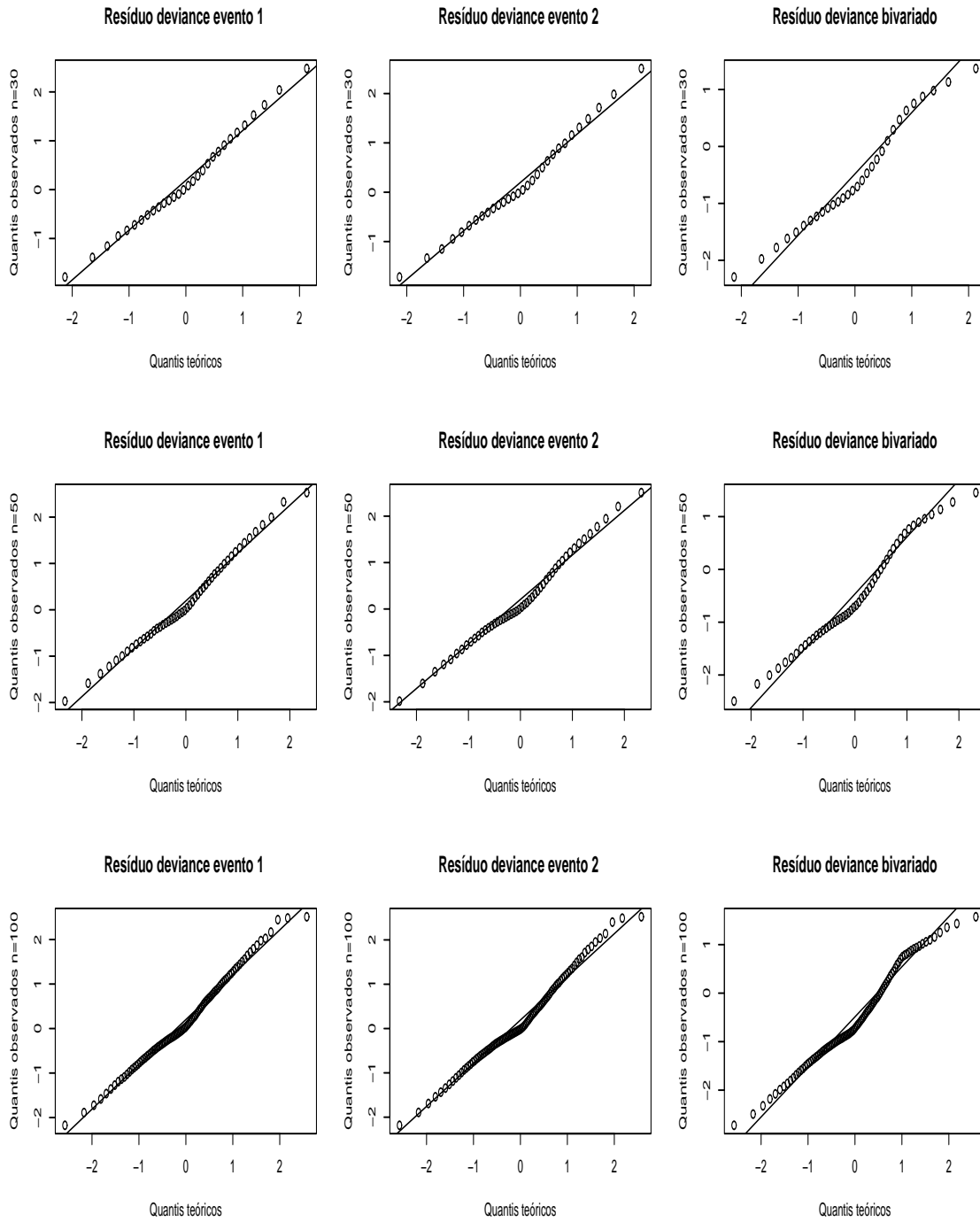


**Figura 1** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 0.3$  e 10% de dados censurados

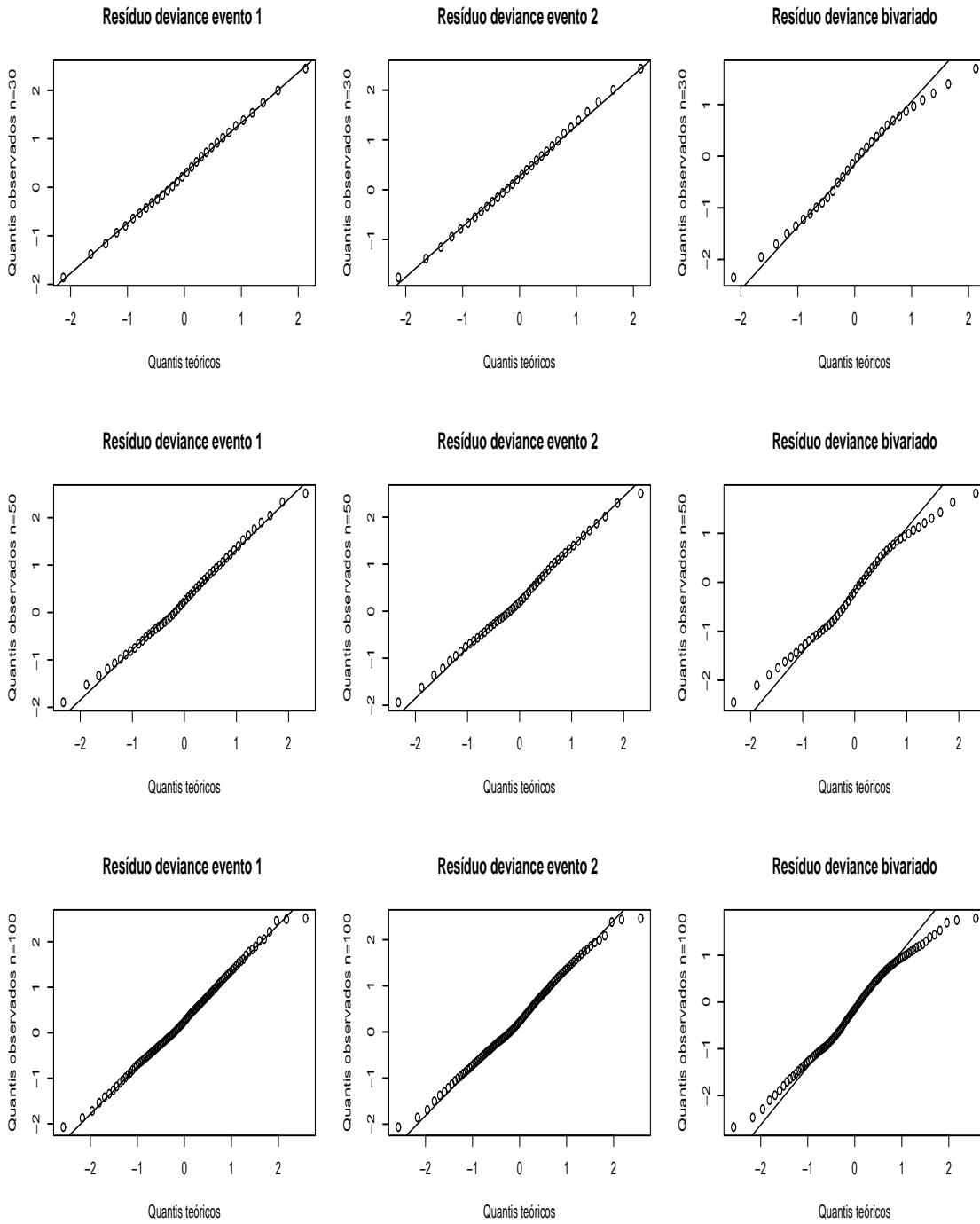




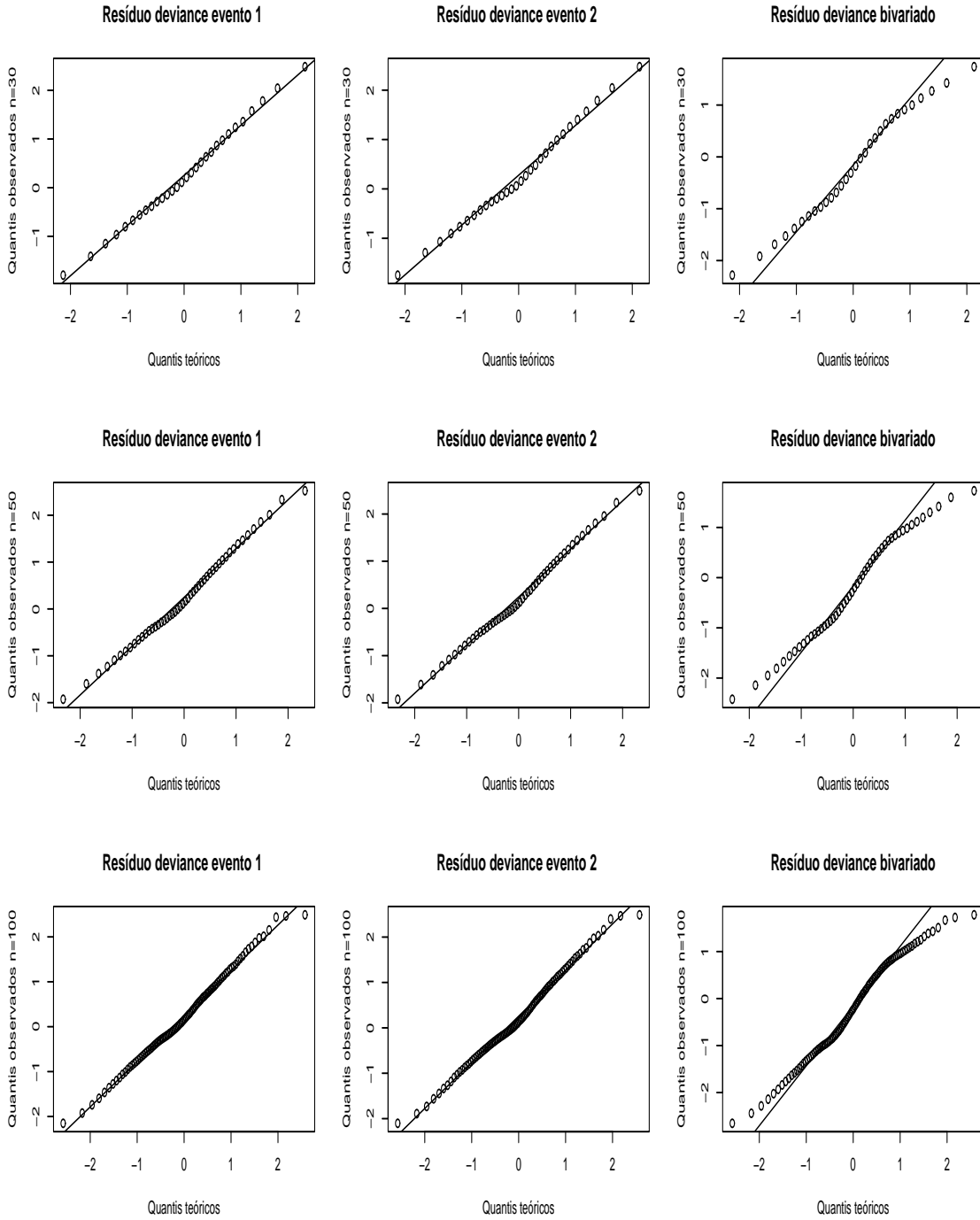
**Figura 2** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 0.3$  e 20% de dados censurados



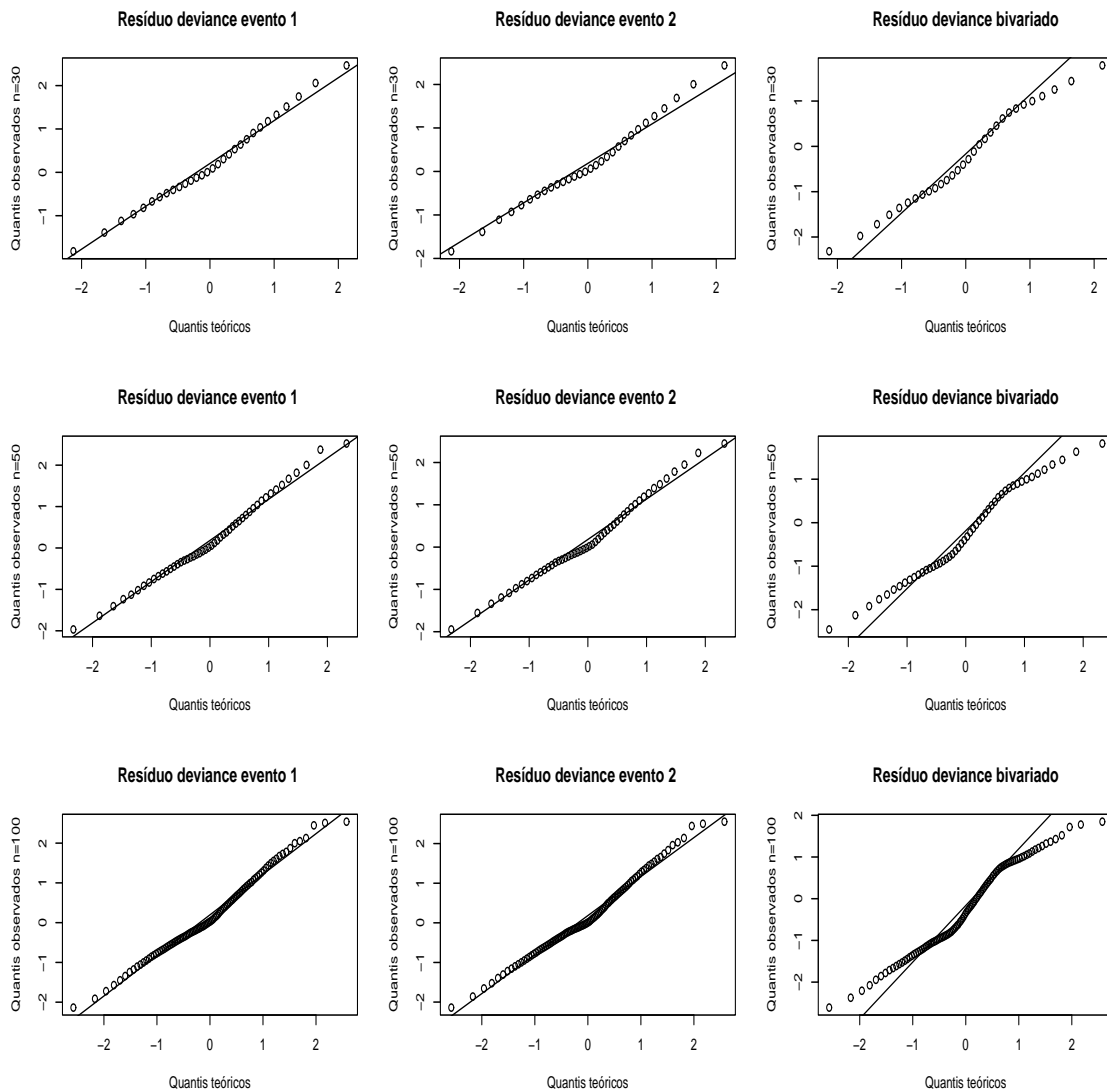
**Figura 3** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 0.3$  e 30% de dados censurados



**Figura 4** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 6$  e 10% de dados censurados

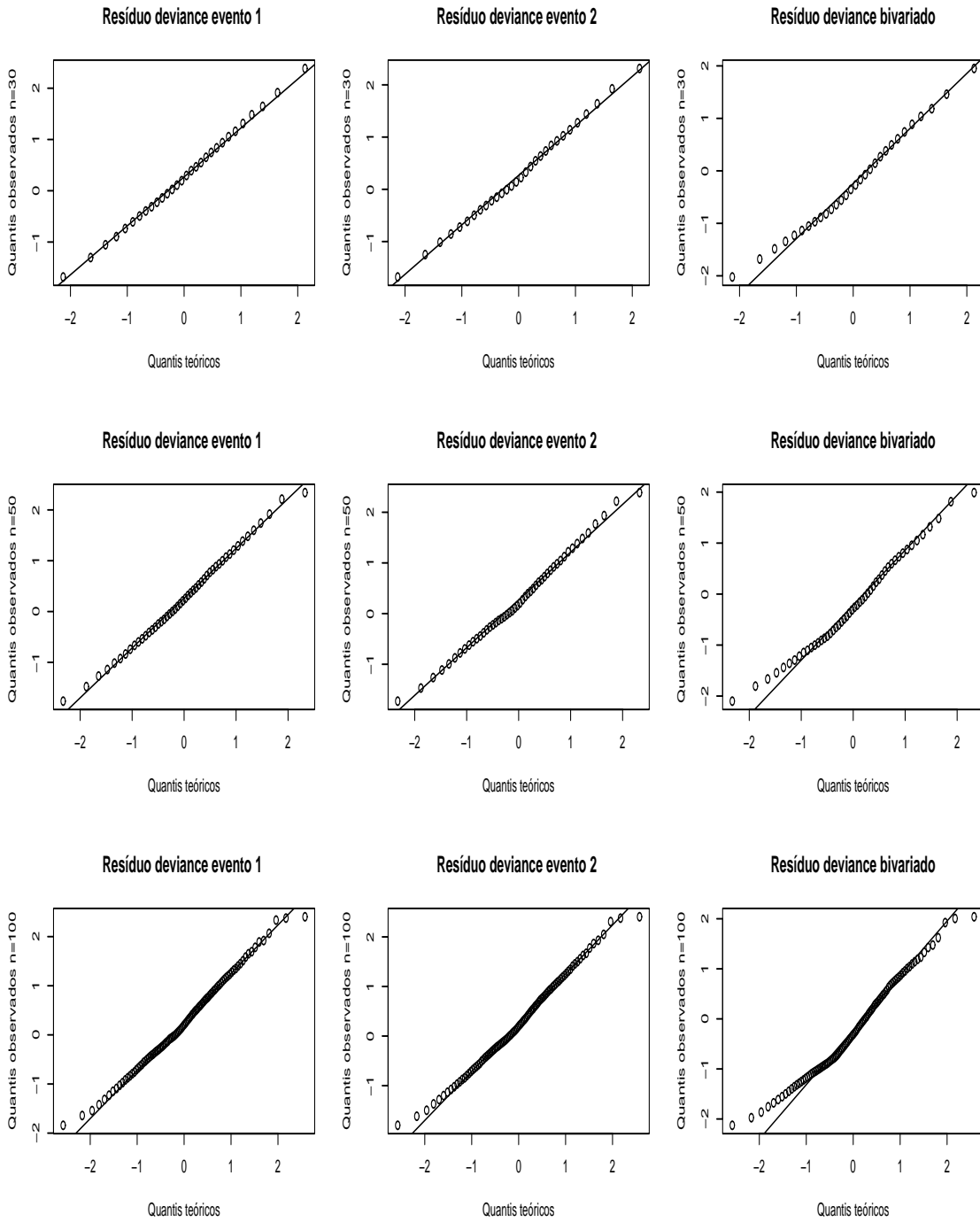


**Figura 5** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 6$  e 20% de dados censurados

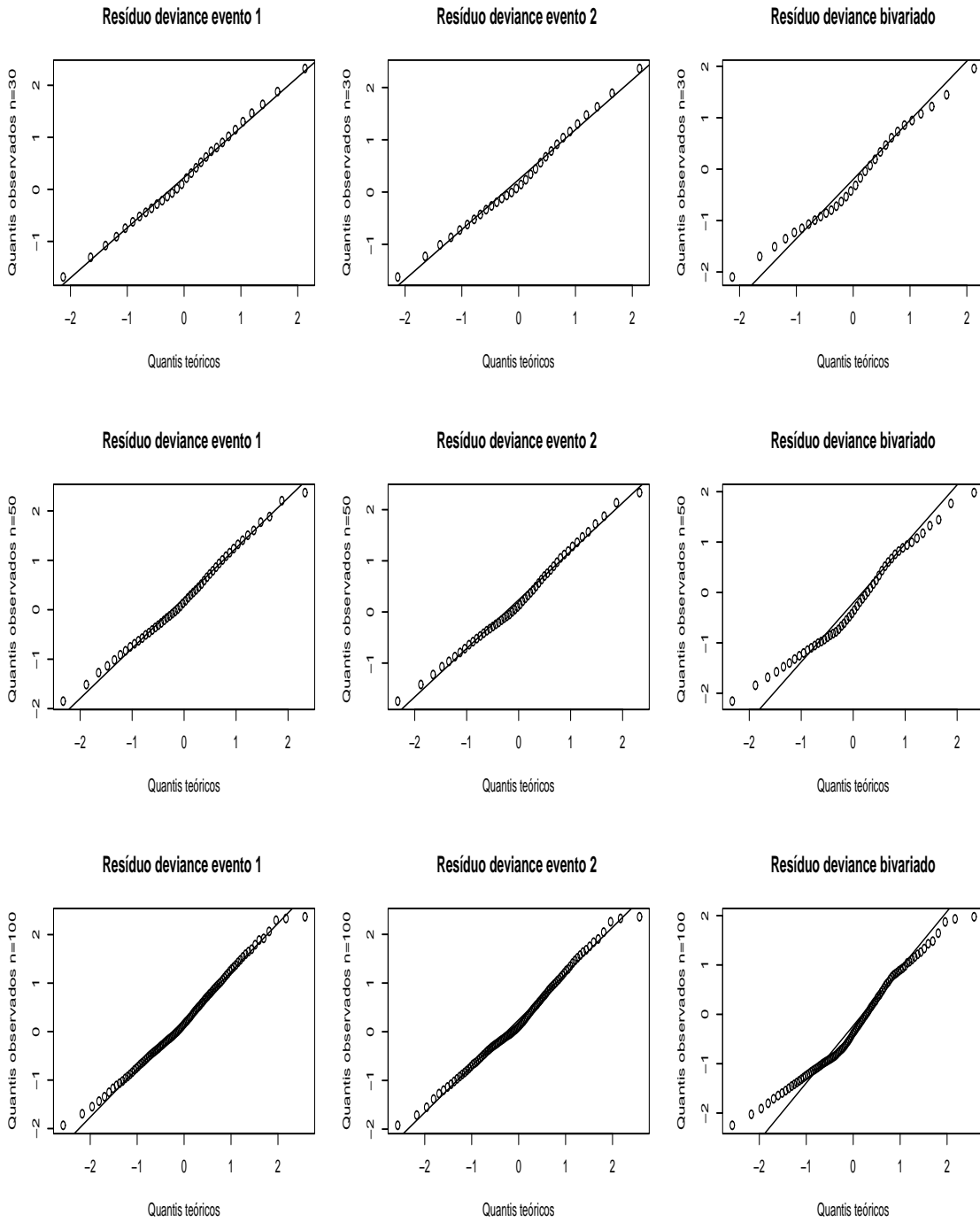


**Figura 6** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Frank com  $\alpha = 6$  e 30% de dados censurados

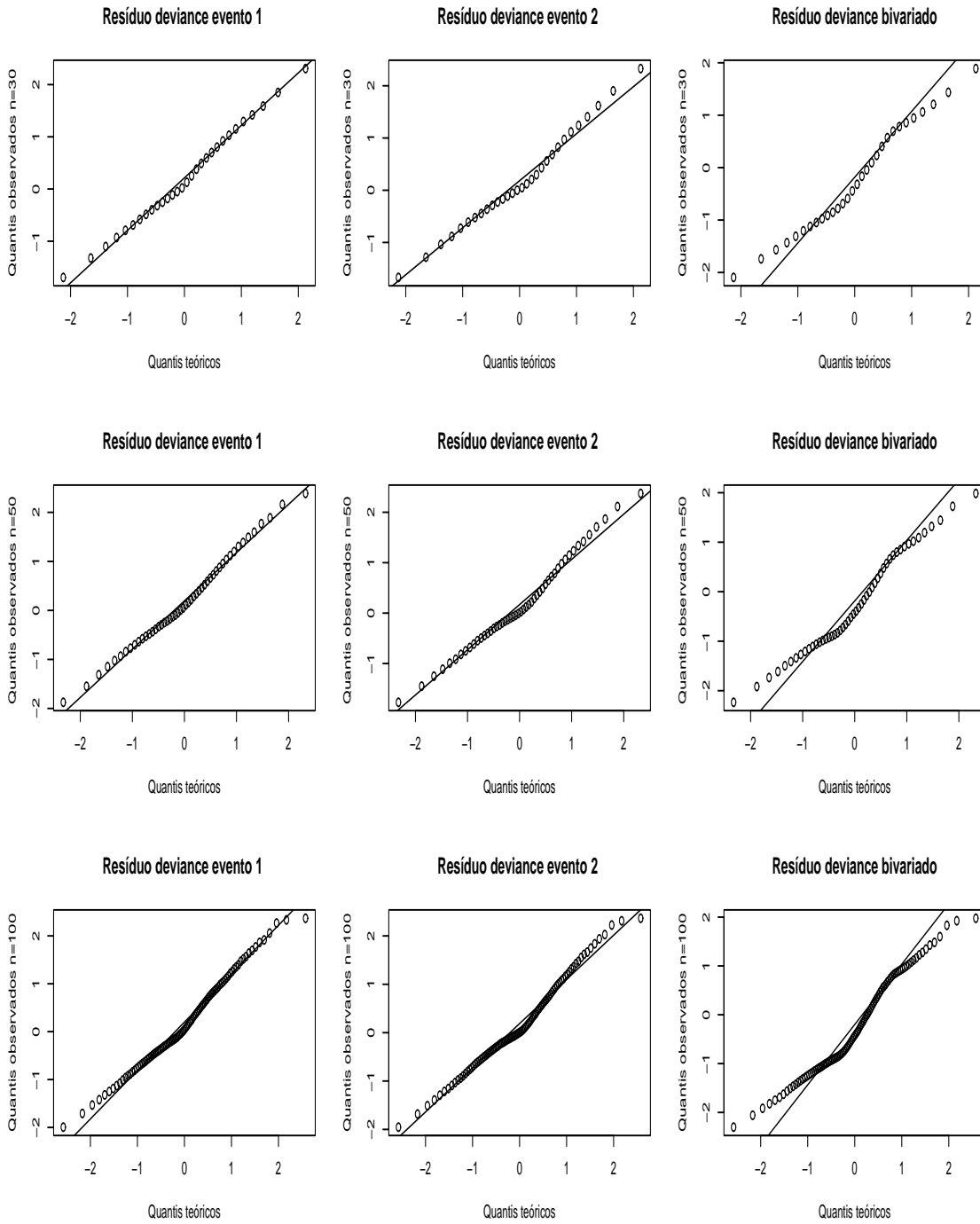
Pode-se notar, através da inspeção das Figuras 1 a 6 que as medidas de resíduo utilizadas seguem distribuições aproximadamente normais para o modelo de regressão bivariado com a cópula de Frank. A distribuição dos resíduos, entretanto, não é simétrica em torno de zero. Para as situações com maior percentual de censura nos dados nota-se um afastamento de normalidade, o que é esperado. O parâmetro  $\alpha$  de associação na cópula de Frank utilizado para geração das amostras também mostra-se como influente sobre a distribuição empírica das medidas de resíduo. Aparentemente, amostras geradas com associação forte entre os eventos tendem a apresentar resíduos bivariados com maiores afastamentos de normalidade.



**Figura 7** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 2$  e 10% de dados censurados

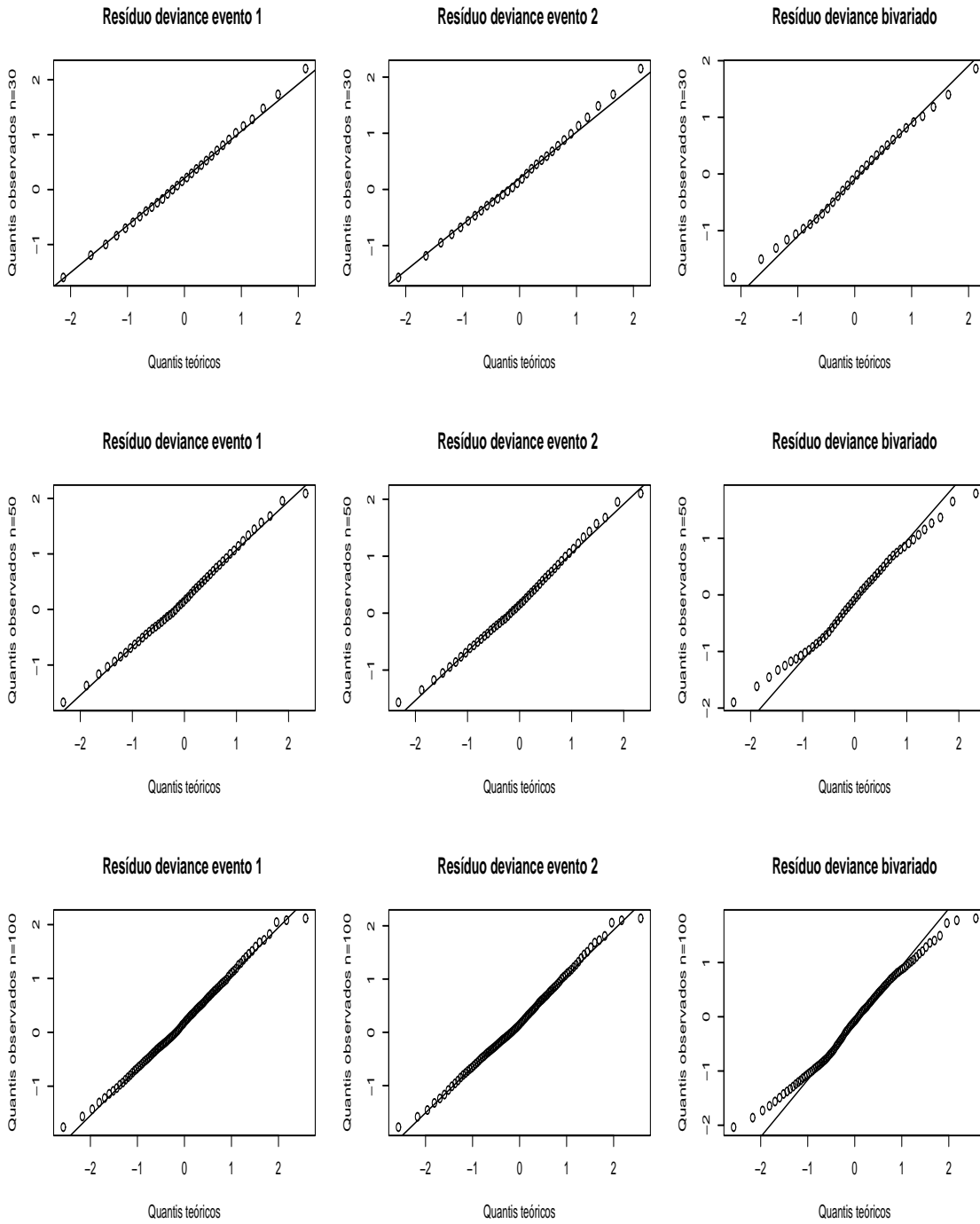


**Figura 8** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 2$  e 20% de dados censurados

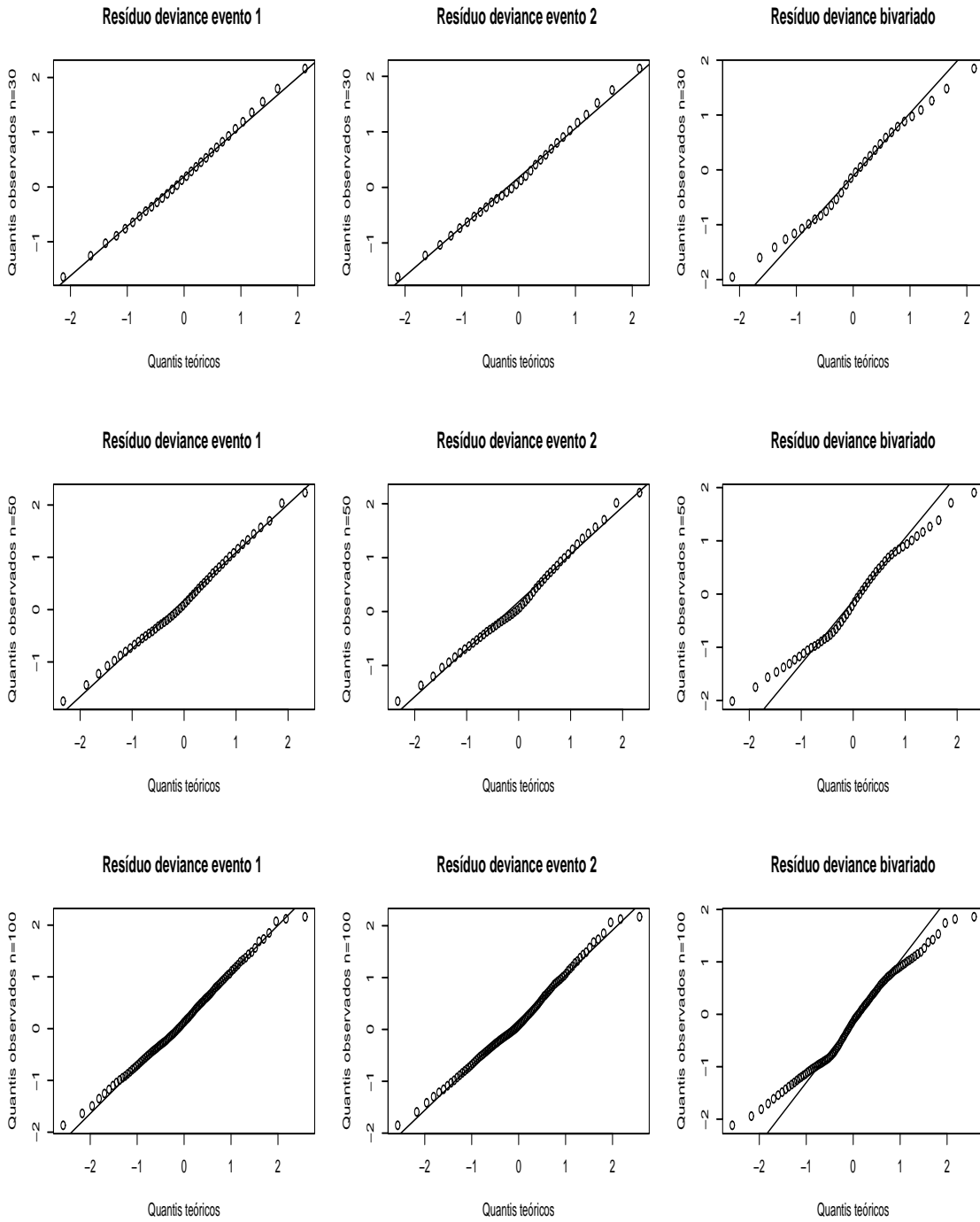


**Figura 9** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 2$  e 30% de dados censurados

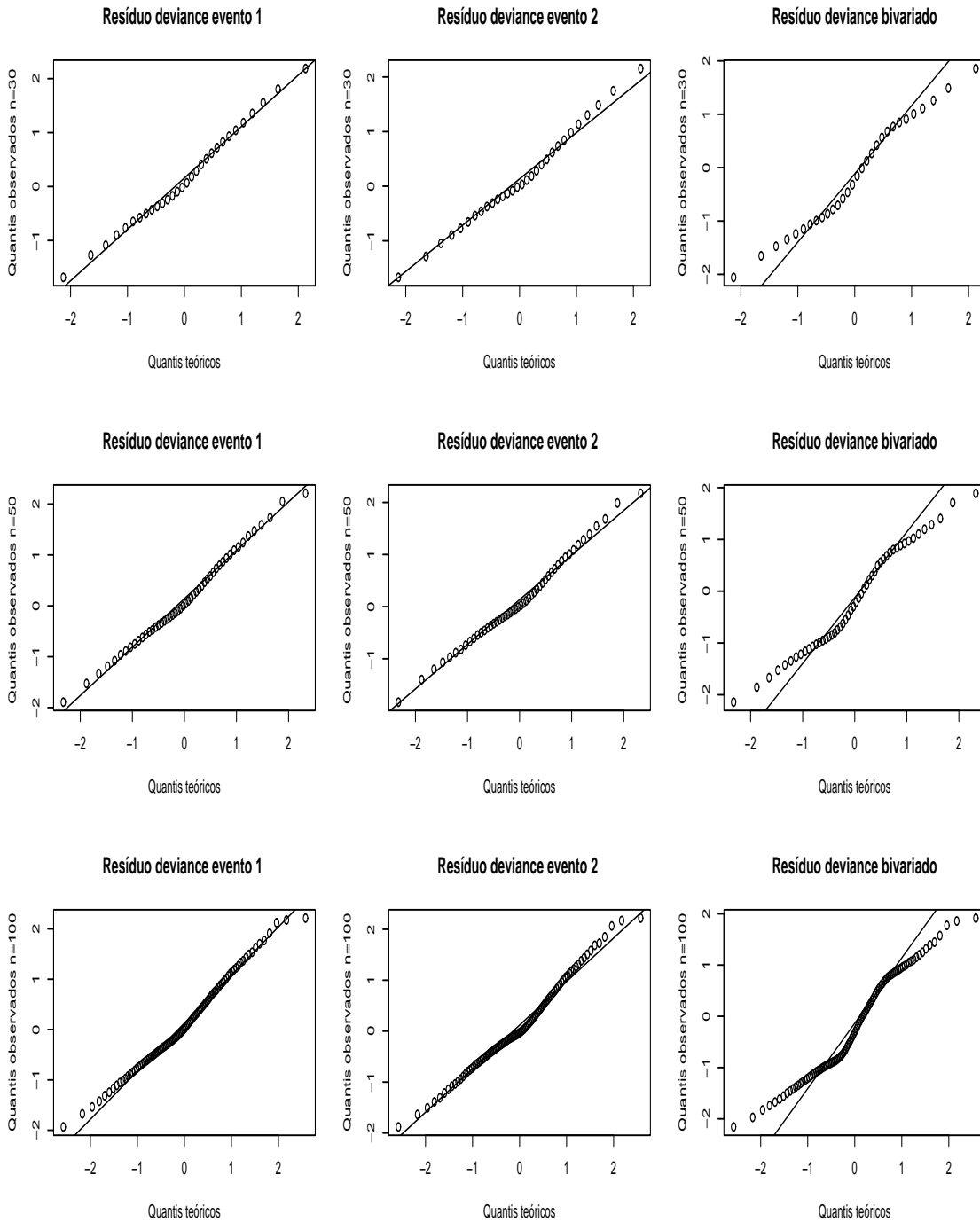




**Figura 10** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 7$  e 10% de dados censurados



**Figura 11** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_i^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 7$  e 20% de dados censurados



**Figura 12** – Gráficos normais de probabilidade para os resíduos deviance marginais  $Dev_{ik}$  e deviance bivariado  $Dev_{ik}^*$  para o modelo de regressão bivariado através da cópula de Clayton com  $\alpha = 7$  e 30% de dados censurados

Nas Figuras 7 a 12 são exibidos os gráficos normais de probabilidade para as medidas de resíduo marginais e bivariado quando são utilizados os modelos regressão bivariados com

a cópula de Clayton. Neste caso, os resíduos são aparentemente simétricos em torno de zero, mas apresentam alguns indícios de afastamento de normalidade principalmente com o aumento da presença de dados censurados.

## 4.2 Aplicação com os dados de pacientes diabéticos

Nas seguintes seções são apresentados os resultados da análise do conjunto de dados descrito em 3.1.1 sob diferentes abordagens. Na seção 4.2.1 estão os resultados de uma análise exploratória inicial dos dados. Na seção 4.2.2 os resultados referentes a utilização da cópula de Frank e na seção 4.2.3 aqueles referentes a utilização da cópula de Clayton.

### 4.2.1 Análise exploratória

Inicialmente foi realizada uma análise exploratória dos dados. A Figura 13 apresenta um gráfico de dispersão dos valores observados para os tempos, em anos, até o diagnóstico de retinopatia leve e grave nos pacientes. Por simplicidade de notação, os diagnósticos de retinopatia leve e grave são muitas vezes referidos neste texto por eventos 1 e 2 respectivamente. O coeficiente de correlação de Tau de Kendall calculado para os dados foi  $\tau = 0,64$ , indicando uma associação positiva entre os eventos, como também pode ser notado no gráfico de dispersão. Os percentuais de observações censuradas nos eventos 1 e 2 foram 53% e 83% respectivamente.

A Figura 14 apresenta gráficos com os ajustes marginais de modelos com distribuições Weibull para os tempos de falha e as estimativas não-paramétrica para os mesmos. As suposições feitas para as distribuições marginais das variáveis resposta parecem adequadas pela análise da Figura 14. Pode-se notar, entretanto, que a sobrevivência estimada para o evento 2 é imprópria, pois  $\lim_{t \rightarrow 0} \hat{S}(t) \neq 0$ . Tal fato pode ser explicado pela ocorrência de um percentual de curados na amostra. Neste trabalho não foram utilizados modelos específicos para abordar esta condição, detalhes para estes tipos de modelos podem ser consultados em Rizzato (2006).

As estimativas não paramétricas por grupos para as diferentes covariáveis podem ser vistas nas Figuras 15, 16, 17, 18 e 19. Os gráficos contidos nestas figuras podem ser utilizados como indicativos sobre a significância das covariáveis nos modelos de regressão. A análise destas figuras indica que os modelos de Cox não devem ser utilizados pois a suposição de riscos proporcionais não pode ser sustentada, uma vez que as curvas de sobrevivência para os diferentes grupos

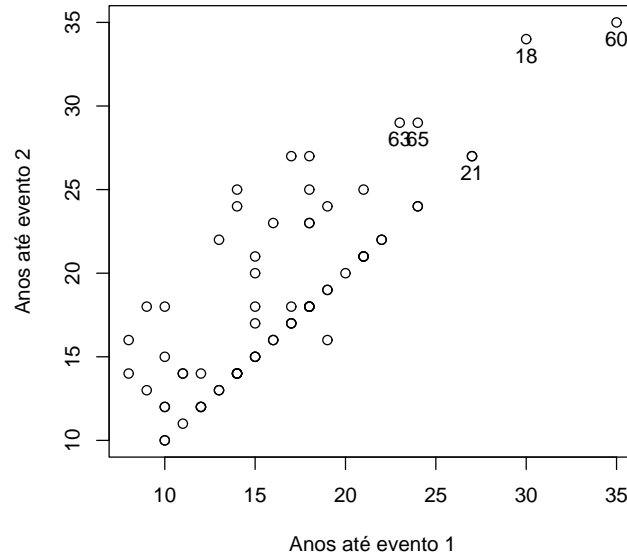


Figura 13 – Gráfico de dispersão das variáveis resposta dos dados de retinopatia

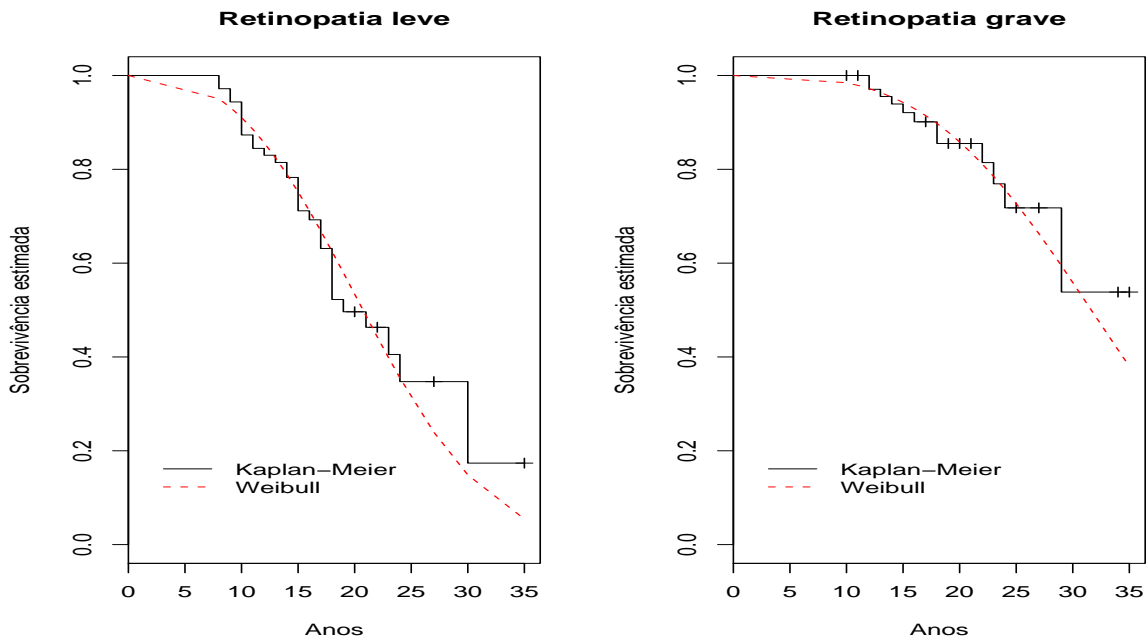
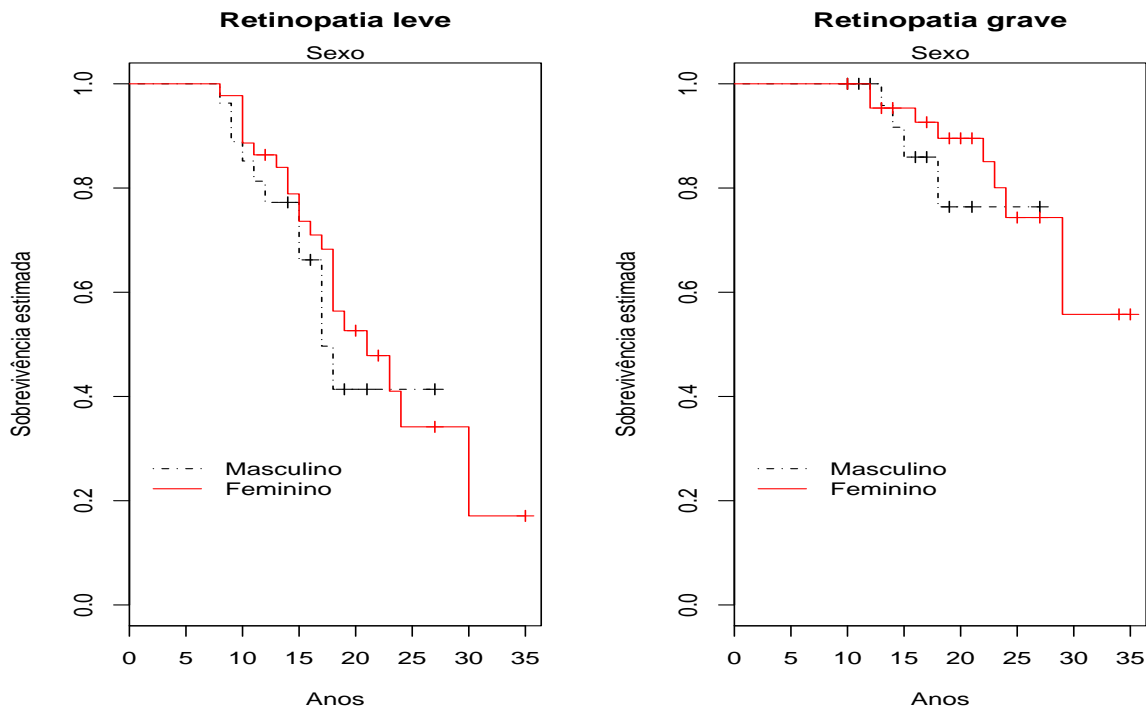


Figura 14 – Curvas de sobrevivência estimadas marginalmente por Kaplan-Meier e pela distribuição Weibull para os dados de retinopatia

se cruzam para diversas variáveis explicativas. É neste contexto que a utilização dos modelos de locação e escala na forma proposta se mostra vantajosa.



**Figura 15** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de retinopatia, para cada evento segundo sexo

#### 4.2.2 Modelo de regressão bivariado através da cópula de Frank

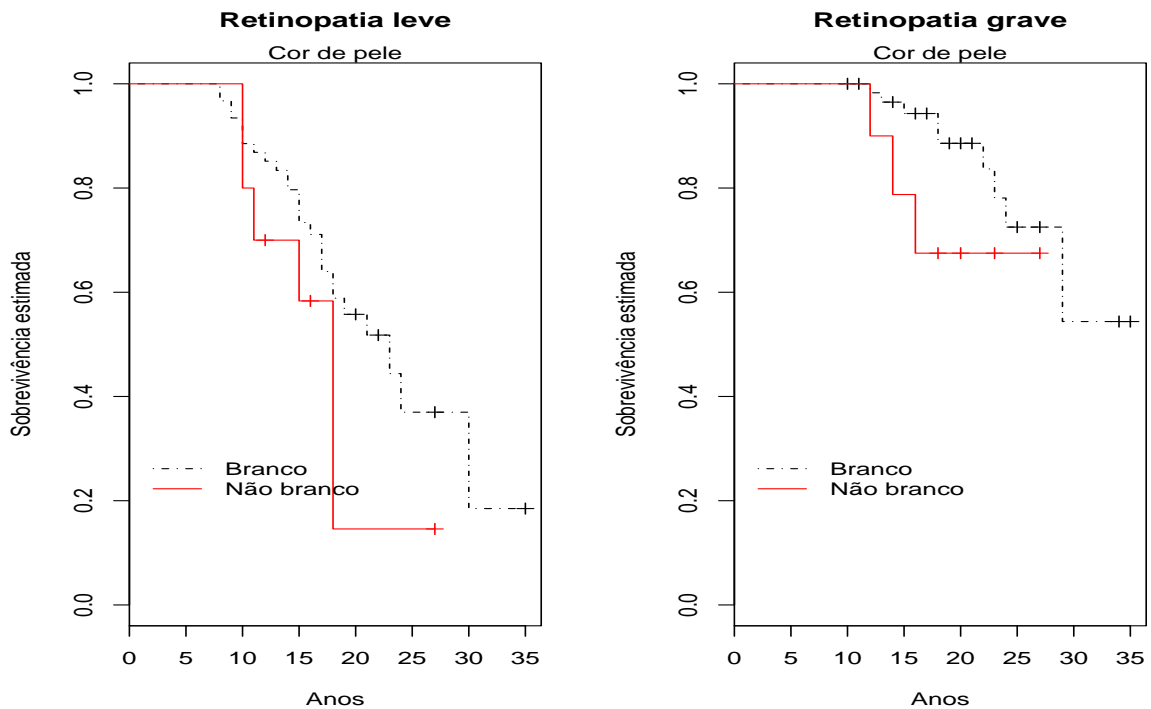
Primeiramente considerou-se o modelo de regressão bivariado através da cópula de Frank descrito na equação (8), reunindo modelos marginais dados por:

$$y_{ik} = \beta_{0k} + \beta_{1k}x_{i1k} + \beta_{2k}x_{i2k} + \beta_{3k}x_{i3k} + \beta_{4k}x_{i4k} + \beta_{5k}x_{i5k} + \sigma_k z_{ik}, i = 1, \dots, 71, k = 1, 2, \quad (10)$$

com funções de sobrevivência definidas em (2).

As estimativas para os parâmetros dos modelos foram obtidas pelos métodos de máxima verosimilhança e jackknife descritos na seção 3.4. Os resultados das estimativas assim como o erros padrões e significâncias encontram-se descritos na Tabela 1.

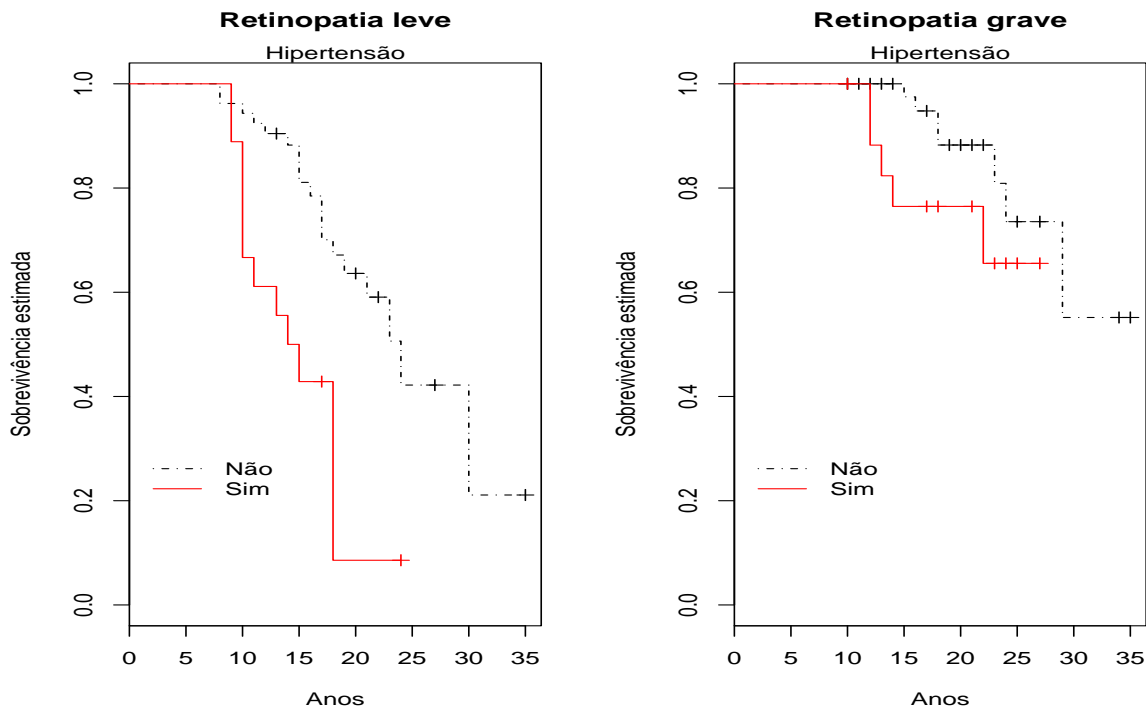
Para o evento 1 foram consideradas significativas as variáveis explicativas referentes a presença de hipertensão e neuropatia nos pacientes. Segundo o modelo ajustado ambas doenças aceleram a ocorrência de retinopatia leve nos pacientes diabéticos. Para o evento 2 foram consideradas significativas as variáveis explicativas referentes ao sexo e presença de neuropatia. Pelas estimativas do modelo os homens vivenciam o evento de retinopatia grave de forma mais precoce



**Figura 16** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de retinopatia, para cada evento segundo cor de pele

**Tabela 1** – Estimativas de máxima verossimilhança (EMV) e Jackknife dos parâmetros do modelo de regressão bivariado através da cópula de Frank para os dados de retinopatia

Parâmetro	EMV			Jackknife		
	Estimativa	Erro padrão	p-valor	Estimativa	Erro padrão	p-valor
$\beta_{01}$	3,2150	0,1166	< 0,0001	3,2555	0,1339	0,0000
$\beta_{11}$	0,1395	0,1064	0,1898	0,0782	0,1340	0,5595
$\beta_{21}$	0,0200	0,1150	0,8619	0,1801	0,2337	0,4409
$\beta_{31}$	-0,3985	0,1027	0,0001	-0,4570	0,1635	0,0052
$\beta_{41}$	-0,2632	0,1070	0,0139	-0,2444	0,1295	0,0591
$\beta_{51}$	0,0888	0,1106	0,4220	0,3117	0,0398	0,0000
$\sigma_1$	0,2815	0,0360	< 0,0001	0,1353	0,1593	0,3957
$\beta_{02}$	3,4820	0,1726	< 0,0001	3,4980	0,2621	0,0000
$\beta_{12}$	0,2349	0,1242	0,0586	0,1828	0,1958	0,3506
$\beta_{22}$	-0,0783	0,1271	0,5379	0,1012	0,2818	0,7195
$\beta_{32}$	-0,1615	0,1119	0,1489	-0,2481	0,2036	0,2230
$\beta_{42}$	-0,3407	0,1502	0,0233	-0,3643	0,1967	0,0640
$\beta_{52}$	0,0845	0,1306	0,5176	0,1936	0,2339	0,4078
$\sigma_2$	0,2513	0,0518	< 0,0001	0,2783	0,0663	0,0000
$\alpha$	10,7074	3,7235	0,0040	7,1264	5,2129	0,1716



**Figura 17** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de retinopatia, para cada evento segundo presença de hipertensão

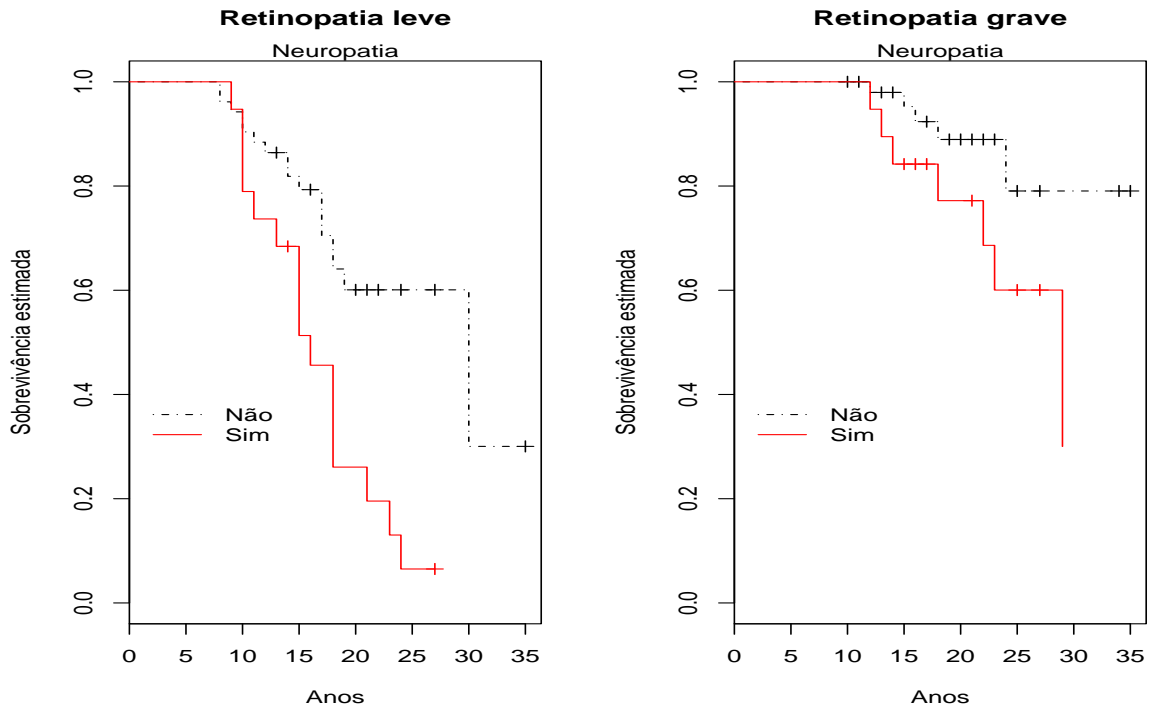
que as mulheres, e a presença de neuropatia foi identificada como fator que acelera a ocorrência de retinopatia grave. As estimativas pontuais obtidas pelos dois métodos são semelhantes.

#### 4.2.2.1 Análise de influência local

As estimativas apresentadas na Tabela 1 precisam ser validadas pelas metodologias de diagnóstico apresentadas na seção 3.4.3. Usando metodologia de influência local foram calculadas as medidas de curvatura máxima e os respectivos vetores correspondentes as direções de maior curvatura ( $\mathbf{d}_{max}$ ), assim como a medida de influência local total ( $C_i$ ). Essas medidas foram calculadas considerando dois esquemas de perturbação sendo eles ponderação de casos e perturbação da variável resposta. O terceiro esquema de perturbação apresentado em 3.4.3, que é referente a perturbações em variáveis explicativas do modelo, não foi utilizado nesta análise, pois todas as variáveis explicativas são categóricas.

As Figuras 20 e 21 apresentam gráficos das medidas de influência local contra os índices das observações para os dois esquemas de perturbação. As observações #24, #26, #44 e





**Figura 18** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de retinopatia, para cada evento segundo presença de neuropatia

#46 destacam-se das demais e podem ser considerados possíveis pontos influentes.

#### 4.2.2.2 Análise de influência global

Foram calculadas a distância generalizada de Cook ( $D_i$ ) e o afastamento da verossimilhança ( $l_i$ ) como medidas de influência global. Na Figura 22 são exibidos o gráfico do afastamento da verossimilhança versus o índice das observações e o gráfico de dispersão da distância de Cook marginal para os dois grupos de parâmetros referentes aos dois eventos. Nesta figura, destacam-se das demais as observações #44 e #46 que são considerados possíveis pontos influentes.

#### 4.2.2.3 Análise de resíduos

Análise de resíduos apresentada em 3.4.5 também foi considerada para validar o ajuste do modelo. A Figura 23 exibe os gráficos com os resíduos deviance para os eventos 1 e 2 versus o índice das observações. Algumas observações tem resíduos fora do intervalo  $[-2, 2]$

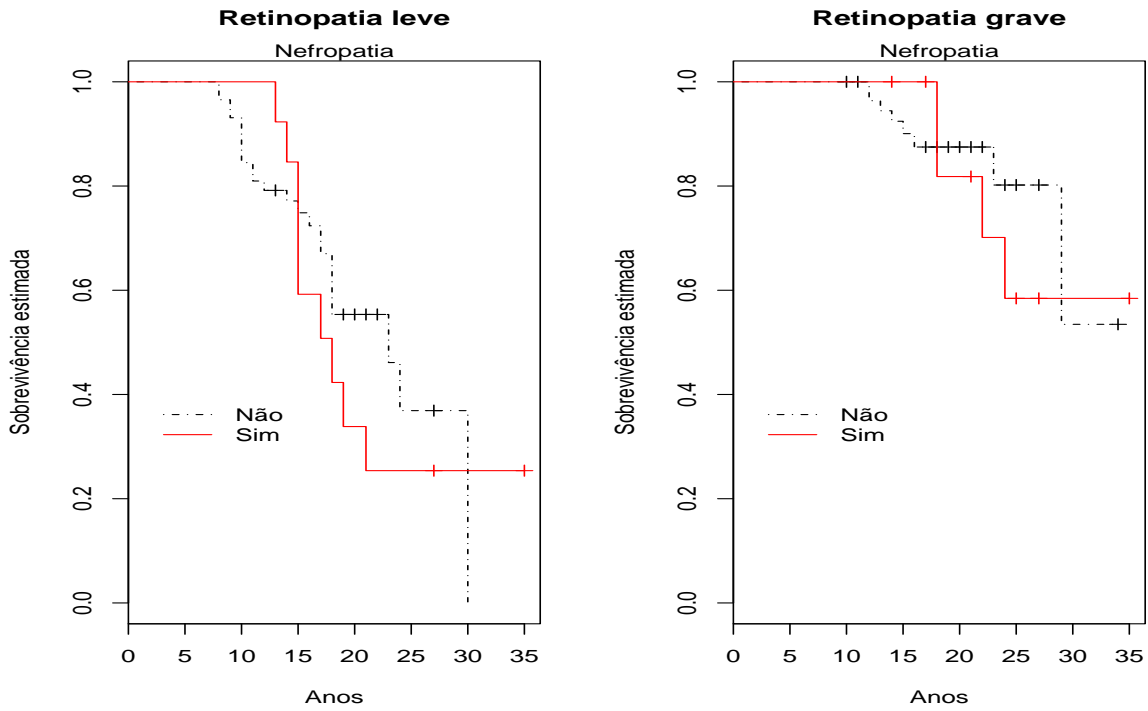


Figura 19 – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de retinopatia, para cada evento segundo presença de nefropatia

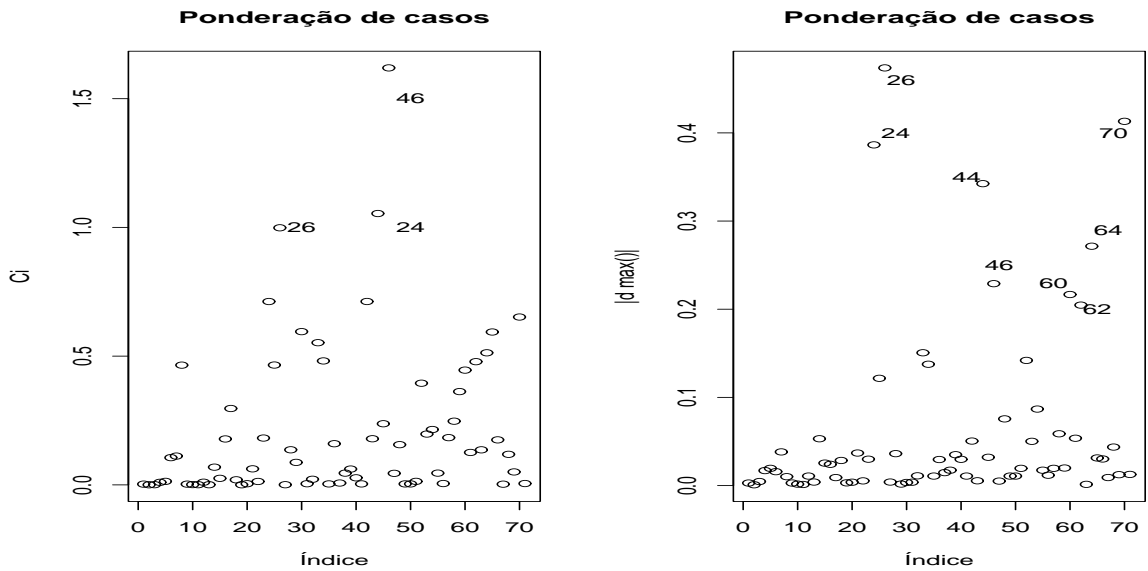
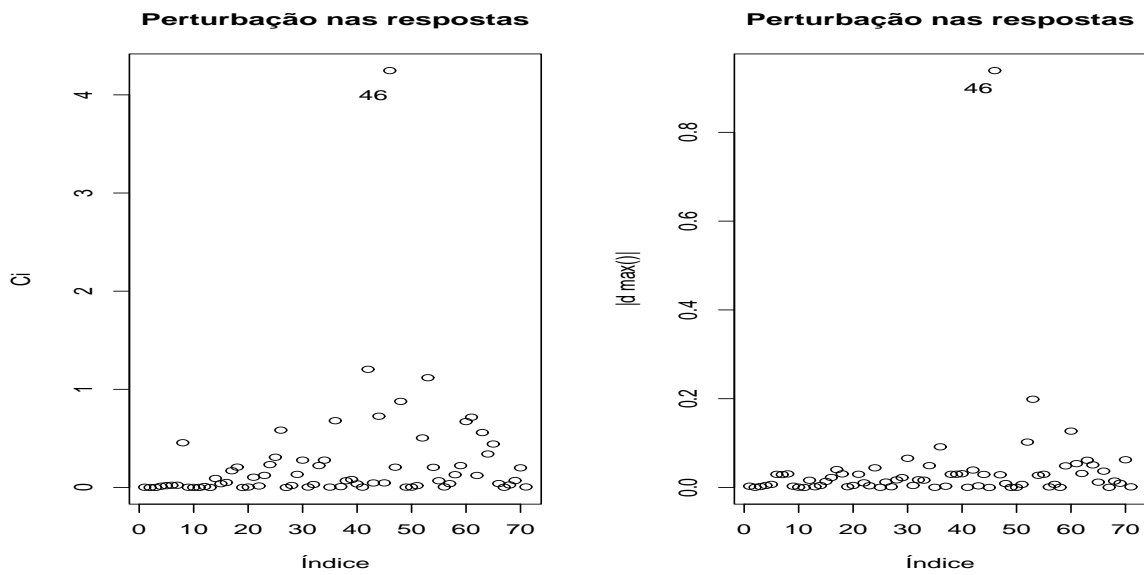


Figura 20 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Frank segundo esquema de ponderação de casos para dados de retinopatia

e merecem atenção especial. Pode-se notar que para o evento 1 o comportamento dos resíduos não apresenta problemas aparentes de aleatoriedade. Para os resíduos referentes ao evento 2



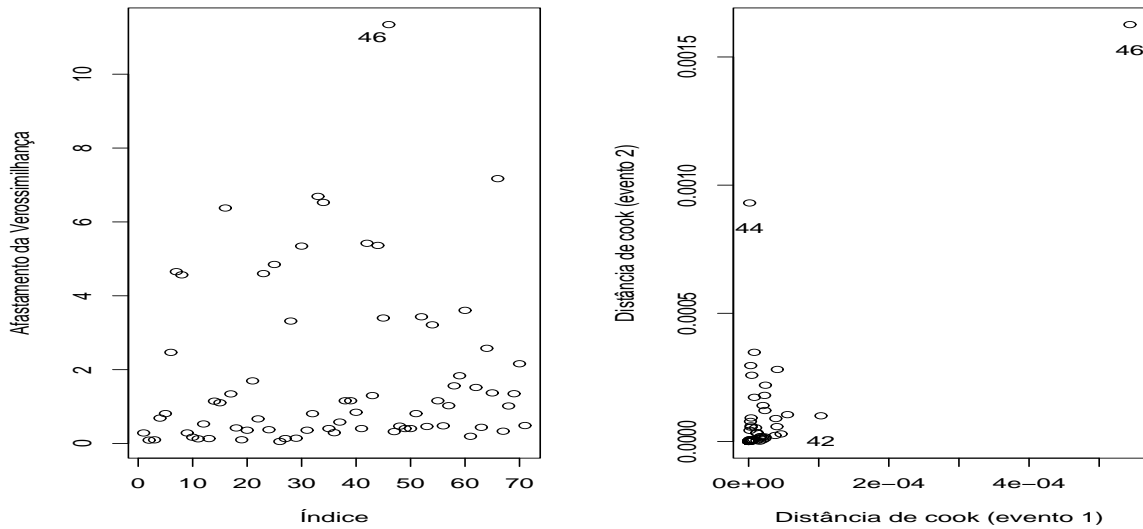
**Figura 21** – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Frank segundo esquema de perturbação de variáveis resposta para dados de retinopatia

percebe-se a formação de um certo padrão que pode ser decorrente da simplicidade do modelo utilizado que não considera uma possível proporção de indivíduos curados na amostra, como foi apontado na análise exploratória dos dados. Os resíduos bivariados detectaram os pontos #42 #46 como possíveis influentes assim como as medidas de influência utilizadas.

#### 4.2.2.4 Impacto das observações influentes

Por terem sido detectados como possíveis pontos influentes de forma recorrente pelas diferentes medidas de diagnóstico, as observações #24, #26, #44 e #46 foram investigadas visando detectar o impacto destas observações no modelo. Para isso foram estimados os parâmetros do modelo a partir de sub-amostras com a retirada dessas observações individualmente e em grupos. Estas estimativas estão apresentadas nas Tabelas 2 e 3 que contém as estimativas dos parâmetros, as mudanças relativas em relação às estimativas da amostra completa e os p-valores para cada sub-amostra, conforme mencionado na Seção 3.4.6

Mediante a análise das tabelas foi diagnosticada a observação 46 como ponto influente. Esta observação refere-se a um paciente do sexo feminino, branco, portador de hipertensão, neuropatia e nefropatia que apresentou observações censuradas para ambos eventos após decorridos 27 anos do diagnóstico de diabetes. Dada sua influência na modelagem proposta, esta



**Figura 22** – Gráficos de medidas de influência global para modelo de regressão bivariado através de cópula de Frank ajustado a dados de retinopatia

observação foi desconsiderada nas demais etapas da análise.

#### 4.2.2.5 Modelo final proposto utilizando a cópula de Frank

Mediante a eliminação da observação #46 os modelos completos para os eventos 1 e 2, como o descrito em (10) foram ajustados ao conjunto de dados. Foram selecionadas para compor os modelos finais as variáveis consideradas significativas a um nível de 10%. Desta forma os modelos finais propostos para o logaritmo dos tempos até o diagnóstico de retinopatia leve e grave em pacientes diabéticos podem ser expressos por:

$$y_{i1} = \beta_{01} + \beta_{11}x_{i11} + \beta_{21}x_{i21} + \beta_{31}x_{i31} + \beta_{41}x_{i41} + \sigma_1 z_{i1} \quad e$$

$$y_{i2} = \beta_{02} + \beta_{12}x_{i12} + \beta_{22}x_{i22} + \beta_{42}x_{i42} + \sigma_2 z_{i2}.$$

As estimativas para os parâmetros destes modelos e para o parâmetro de associação  $\alpha$  da cópula de Frank estão apresentadas na Tabela 4. Os resultados podem ser interpretados de forma que os coeficientes de regressão com sinal positivo retardam a ocorrência do evento a que se referem, enquanto os coeficientes negativos aceleram a ocorrência dos eventos.

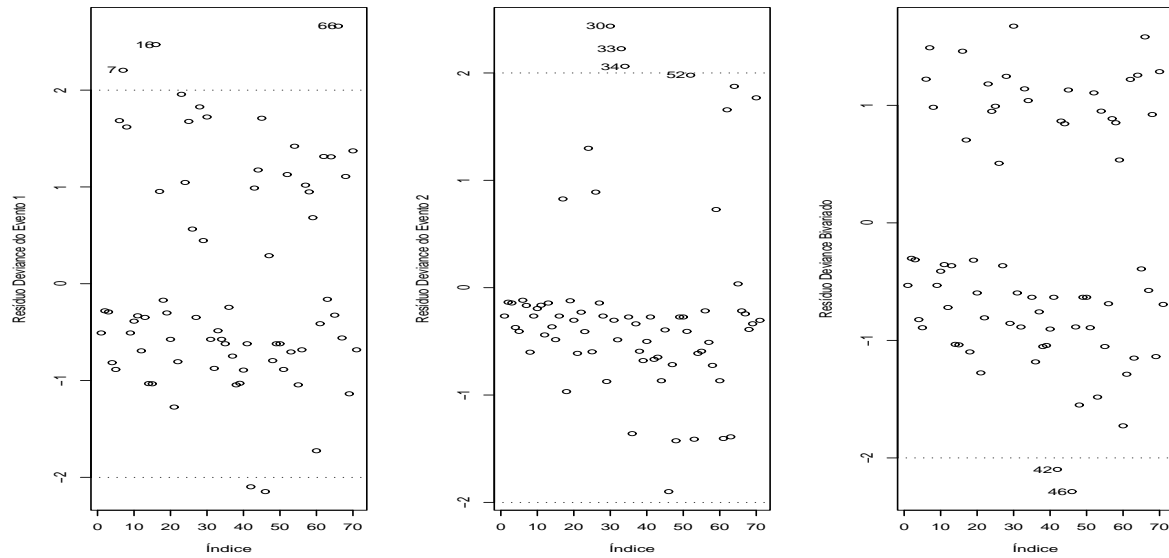
A medida  $\tau$  de Kendall calculada através da fórmula (3) e da estimativa do parâmetro  $\alpha$  da cópula de Frank foi 0.6385, valor bastante próximo do calculado na análise

**Tabela 2** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até retinopatia leve utilizando a cópula de Frank

Sub-amostra	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{31}$	$\beta_{41}$	$\beta_{51}$
A – (24)	3,242	0,098	0,006	-0,376	-0,245	0,065
	[-0,846]	[29,606]	[71,500]	[5,696]	[7,029]	[26,914]
	(0,000)	(0,394)	(0,963)	(0,000)	(0,026)	(0,572)
A – (26)	3,235	0,118	0,007	-0,383	-0,268	0,069
	[-0,622]	[15,197]	[63,500]	[3,915]	[-1,976]	[21,959]
	(0,000)	(0,286)	(0,952)	(0,000)	(0,014)	(0,541)
A – (44)	3,221	0,149	0,019	-0,412	-0,266	0,092
	[-0,196]	[-7,025]	[4,000]	[-3,388]	[-1,178]	[-3,829]
	(0,000)	(0,141)	(0,862)	(0,000)	(0,011)	(0,387)
A – (46)	3,152	0,222	-0,193	-0,271	-0,301	-0,016
	[1,950]	[-58,781]	[1066,000]	[31,945]	[-14,552]	[117,905]
	(0,000)	(0,015)	(0,110)	(0,008)	(0,002)	(0,875)
A – (24,26)	3,255	0,095	-0,006	-0,369	-0,264	0,061
	[-1,244]	[31,685]	[129,000]	[7,478]	[-0,114]	[31,081]
	(0,000)	(0,437)	(0,963)	(0,001)	(0,021)	(0,606)
A – (24,44)	3,244	0,112	0,006	-0,390	-0,248	0,071
	[-0,905]	[19,713]	[69,000]	[2,208]	[5,623]	[19,932]
	(0,000)	(0,312)	(0,958)	(0,000)	(0,022)	(0,523)
A – (24,46)	3,169	0,202	-0,213	-0,253	-0,293	-0,033
	[1,437]	[-44,516]	[1163,000]	[36,512]	[-11,360]	[137,613]
	(0,000)	(0,039)	(0,094)	(0,018)	(0,004)	(0,751)
A – (26,44)	3,230	0,123	0,039	-0,385	-0,259	0,067
	[-0,479]	[11,971]	[-96,500]	[3,338]	[1,482]	[24,775]
	(0,000)	(0,263)	(0,749)	(0,000)	(0,017)	(0,547)
A – (26,46)	3,166	0,221	-0,217	-0,269	-0,323	-0,018
	[1,521]	[-58,710]	[1187,000]	[32,472]	[-22,758]	[120,495]
	(0,000)	(0,023)	(0,084)	(0,009)	(0,002)	(0,860)
A – (44,46)	3,162	0,219	-0,166	-0,280	-0,302	-0,015
	[1,642]	[-56,918]	[929,500]	[29,837]	[-14,666]	[117,117]
	(0,000)	(0,013)	(0,147)	(0,006)	(0,002)	(0,878)
A – (24,26,44)	3,253	0,094	0,026	-0,364	-0,251	0,051
	[-1,173]	[32,903]	[-28,000]	[8,657]	[4,749]	[42,117]
	(0,000)	(0,439)	(0,841)	(0,001)	(0,025)	(0,655)
A – (24,26,46)	3,181	0,208	-0,230	-0,257	-0,323	-0,026
	[1,061]	[-48,889]	[1249,000]	[35,508]	[-22,758]	[128,829]
	(0,000)	(0,050)	(0,079)	(0,019)	(0,003)	(0,812)
A – (24,44,46)	3,177	0,201	-0,183	-0,260	-0,296	-0,032
	[1,179]	[-43,871]	[1015,000]	[34,680]	[-12,538]	[136,149]
	(0,000)	(0,032)	(0,131)	(0,015)	(0,003)	(0,755)
A – (26,44,46)	3,163	0,218	-0,178	-0,272	-0,308	-0,020
	[1,611]	[-56,272]	[992,500]	[31,669]	[-17,173]	[122,072]
	(0,000)	(0,025)	(0,161)	(0,009)	(0,002)	(0,848)
A – (24,26,44,46)	3,180	0,201	-0,192	-0,257	-0,307	-0,031
	[1,098]	[-43,799]	[1059,000]	[35,583]	[-16,641]	[134,685]
	(0,000)	(0,057)	(0,149)	(0,019)	(0,004)	(0,772)

**Tabela 3** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até retinopatia grave utilizando a cópula de Frank

Sub-amostra	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{32}$	$\beta_{42}$	$\beta_{52}$	$\alpha$
	3,525	0,146	-0,120	-0,107	-0,287	0,025	10,147
A - (24)	[-1,235] (0,000)	[37,633] (0,306)	[-52,746] (0,375)	[33,437] (0,366)	[15,850] (0,051)	[71,006] (0,853)	[5,233] (0,008)
	3,512	0,160	-0,018	-0,110	-0,306	0,020	10,884
A - (26)	[-0,862] (0,000)	[31,758] (0,250)	[77,395] (0,909)	[31,827] (0,353)	[10,273] (0,040)	[75,858] (0,874)	[-1,652] (0,007)
	3,604	0,192	-0,166	-0,202	-0,413	0,106	14,001
A - (44)	[-3,512] (0,000)	[18,476] (0,131)	[-112,388] (0,232)	[-25,325] (0,081)	[-21,309] (0,010)	[-25,562] (0,406)	[-30,763] (0,002)
	3,343	0,334	-0,313	-0,008	-0,311	-0,056	9,353
A - (46)	[4,003] (0,000)	[-42,146] (0,001)	[-299,872] (0,016)	[94,923] (0,940)	[8,835] (0,009)	[165,917] (0,635)	[12,646] (0,003)
	3,578	0,063	-0,076	-0,055	-0,269	-0,028	10,854
A - (24,26)	[-2,757] (0,000)	[73,223] (0,703)	[2,554] (0,626)	[65,820] (0,648)	[21,133] (0,053)	[132,663] (0,819)	[-1,375] (0,011)
	3,644	0,116	-0,218	-0,151	-0,369	0,055	13,369
A - (24,44)	[-4,661] (0,000)	[50,532] (0,420)	[-178,544] (0,148)	[6,192] (0,210)	[-8,453] (0,024)	[34,556] (0,675)	[-24,861] (0,005)
	3,375	0,295	-0,364	0,036	-0,295	-0,100	9,070
A - (24,46)	[3,079] (0,000)	[-25,585] (0,004)	[-365,134] (0,007)	[122,105] (0,752)	[13,531] (0,013)	[218,698] (0,388)	[15,288] (0,005)
	3,596	0,137	-0,108	-0,148	-0,360	0,049	12,262
A - (26,44)	[-3,262] (0,000)	[41,848] (0,359)	[-37,420] (0,507)	[8,359] (0,251)	[-5,547] (0,037)	[41,657] (0,721)	[-14,525] (0,005)
	3,388	0,270	-0,251	0,012	-0,301	-0,081	9,738
A - (26,46)	[2,694] (0,000)	[-14,985] (0,023)	[-220,945] (0,118)	[107,678] (0,916)	[11,652] (0,016)	[195,503] (0,491)	[9,057] (0,006)
	3,433	0,291	-0,362	-0,028	-0,362	-0,056	12,082
A - (44,46)	[1,413] (0,000)	[-23,968] (0,004)	[-362,963] (0,006)	[82,415] (0,797)	[-6,164] (0,005)	[165,917] (0,636)	[-12,843] (0,003)
	3,656	0,024	-0,154	-0,078	-0,304	-0,013	11,938
A - (24,26,44)	[-4,994] (0,000)	[89,911] (0,893)	[-96,807] (0,352)	[51,641] (0,531)	[10,889] (0,052)	[115,740] (0,916)	[-11,495] (0,009)
	3,432	0,218	-0,300	0,060	-0,288	-0,120	9,826
A - (24,26,46)	[1,427] (0,000)	[7,280] (0,107)	[-283,269] (0,059)	[137,090] (0,628)	[15,468] (0,018)	[241,657] (0,291)	[8,229] (0,010)
	3,461	0,257	-0,413	0,015	-0,349	-0,100	11,857
A - (24,44,46)	[0,617] (0,000)	[-9,493] (0,014)	[-427,203] (0,003)	[109,598] (0,889)	[-2,377] (0,007)	[218,225] (0,384)	[-10,735] (0,006)
	3,439	0,258	-0,326	-0,009	-0,337	-0,073	10,697
A - (26,44,46)	[1,221] (0,000)	[-9,621] (0,035)	[-316,220] (0,040)	[94,427] (0,940)	[1,145] (0,016)	[186,391] (0,550)	[0,094] (0,005)
	3,478	0,201	-0,361	0,041	-0,315	-0,117	10,598
A - (24,26,44,46)	[0,121] (0,000)	[14,517] (0,147)	[-360,920] (0,024)	[125,573] (0,733)	[7,661] (0,019)	[238,343] (0,316)	[1,016] (0,008)



**Figura 23** – Gráficos de resíduos Deviance obtidos para modelo de regressão bivariado através de cópula de Frank ajustado a dados de retinopatia

exploratória da Seção 4.2.1. Esta proximidade pode ser considerada como um indicativo de que a cópula de Frank modelou a associação corretamente.

Uma análise gráfica da qualidade do ajuste dos modelos propostos para os eventos em estudo pode ser realizada na Figura 24 onde estão exibidos os gráficos de envelopes simulados para os resíduos deviance dos eventos 1 e 2. Pode-se notar que os resíduos marginais apresentam afastamentos de normalidade, principalmente para o evento 2. Estes resíduos, entretanto, estão contidos nas bandas de confiança dos envelopes. A Figura 25 exhibe um gráfico de dispersão e um gráfico Normal de probabilidades para o resíduo do ajuste bivariado. Pode-se notar que no gráfico de dispersão existe uma formação de grupos e no gráfico normal tem-se um afastamento de normalidade. Estes fatos possivelmente podem ser explicados pela utilização de função de sobrevivência usual que não considera a proporção de indivíduos curados na amostra.

### 4.2.3 Modelo de regressão bivariado através da cópula de Clayton

Nesta seção são apresentados os resultados dos modelos de regressão bivariados através da cópula de Clayton. O modelo é referente aos logaritmos dos tempos até os diagnósticos de retinopatia leve e grave em pacientes diabéticos. Como visto nos resultados da análise exploratória, é razoável supor que os logaritmos dos tempos até os eventos 1 e 2 têm distribuição

**Tabela 4** – Estimativas para o modelo final com a cópula de Frank para os dados de retinopatia

Parâmetro	Estimativa	Erro padrão	p-valor
$\beta_{01}$	3,1454	0,0900	< 0,0000
$\beta_{11}$	0,2231	0,0898	0,0130
$\beta_{21}$	-0,1968	0,1073	0,0666
$\beta_{31}$	-0,2576	0,0762	0,0007
$\beta_{41}$	-0,3079	0,0940	0,0011
$\sigma_1$	0,2560	0,0327	< 0,0000
$\beta_{02}$	3,3201	0,1334	< 0,0000
$\beta_{12}$	0,3423	0,0971	0,0004
$\beta_{22}$	-0,2965	0,1033	0,0041
$\beta_{42}$	-0,3199	0,1164	0,0060
$\sigma_2$	0,2119	0,0438	< 0,0000
$\alpha$	9,0595	3,0701	0,0032

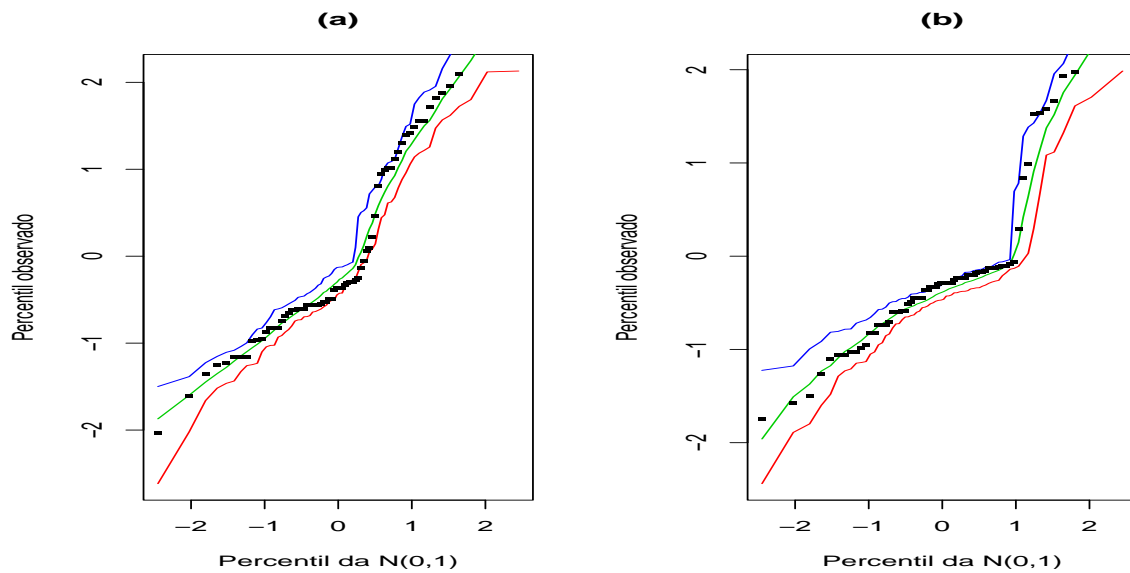
valor extremo. Os resultados da estimação dos parâmetros de regressão e do parâmetro de associação da cópula de Clayton estão apresentados na Tabela 5. Nesta tabela aparecem os resultados das estimativas, erros padrões e significâncias dos parâmetros obtidos pelos métodos de máxima verossimilhança e jackknife descritos na Seção 3.4 .

**Tabela 5** – Estimativas de máxima verossimilhança (EMV) e Jackknife dos parâmetros do modelo de regressão bivariado através da cópula de Clayton para os dados de retinopatia

Parâmetro	EMV			Jackknife		
	Estimativa	Erro padrão	p-valor	Estimativa	Erro padrão	p-valor
$\beta_{01}$	3,1994	0,1167	0,0000	3,2224	0,1349	0,0000
$\beta_{11}$	0,1597	0,1095	0,1448	0,1412	0,1483	0,3410
$\beta_{21}$	-0,0099	0,1135	0,9308	0,0931	0,2151	0,6652
$\beta_{31}$	-0,3980	0,1016	0,0001	-0,4736	0,1625	0,0036
$\beta_{41}$	-0,2634	0,1019	0,0097	-0,1982	0,1286	0,1233
$\beta_{51}$	0,0758	0,1111	0,4950	0,0806	0,1733	0,6421
$\sigma_1$	0,2838	0,0361	0,0000	0,3090	0,0387	0,0000
$\beta_{02}$	3,4017	0,1605	0,0000	3,3639	0,2532	0,0000
$\beta_{12}$	0,2682	0,1250	0,0319	0,2617	0,2250	0,2449
$\beta_{22}$	-0,1077	0,1207	0,3721	0,0539	0,2636	0,8381
$\beta_{32}$	-0,1407	0,1058	0,1835	-0,2561	0,2081	0,2185
$\beta_{42}$	-0,3515	0,1404	0,0123	-0,3947	0,1825	0,0305
$\beta_{52}$	0,1095	0,1274	0,3901	0,2851	0,2656	0,2831
$\sigma_2$	0,2406	0,0467	0,0000	0,2664	0,0734	0,0003
$\alpha$	7,8453	2,9314	0,0074	3,6416	3,8435	0,3434

Para o evento 1 foram consideradas significativas as variáveis explicativas referentes





**Figura 24** – Gráficos de envelopes simulados para resíduo Deviance marginal dos dados de retinopatia

a presença de hipertensão e neuropatia nos pacientes. Segundo o modelo ajustado ambas doenças aceleram a ocorrência de retinopatia leve nos pacientes diabéticos. Para o evento 2 foram consideradas significativas as variáveis explicativas referentes a sexo e presença de neuropatia. Pelas estimativas do modelo os homens vivenciam o evento de retinopatia grave de forma mais precoce que as mulheres, e a presença de neuropatia foi identificada como fator que acelera a ocorrência de retinopatia grave.

#### 4.2.3.1 Análise de influência local

Considerando os esquemas de ponderação de casos e perturbações nas variáveis resposta, foram calculadas as medidas de curvatura máxima e o respectivo vetor correspondente a direção de maior curvatura ( $\mathbf{d}_{max}$ ), assim como a medida de influência local total ( $C_i$ ). Gráficos destas medidas versus os índices das observações estão dispostos nas Figuras 26 e 27. Nestes gráficos as observações #24, #26, #46 e #65 se destacam das demais e, portanto, são possíveis pontos influentes.

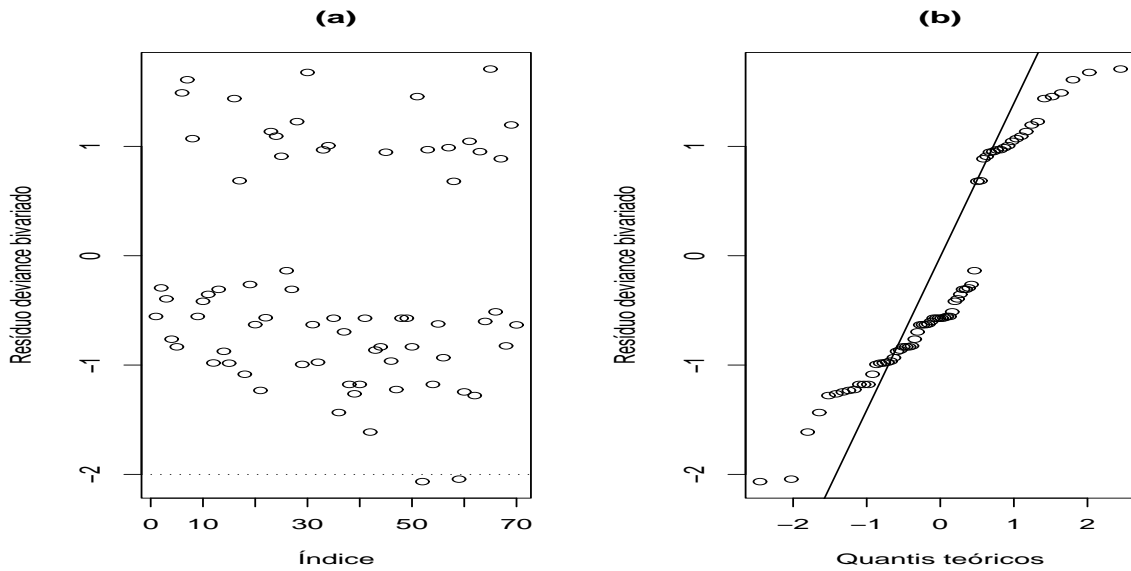


Figura 25 – (a) Gráfico do resíduo bivariado e (b) Gráfico Normal de probabilidades do resíduo bivariado com cópula de Frank nos dados de retinopatia.

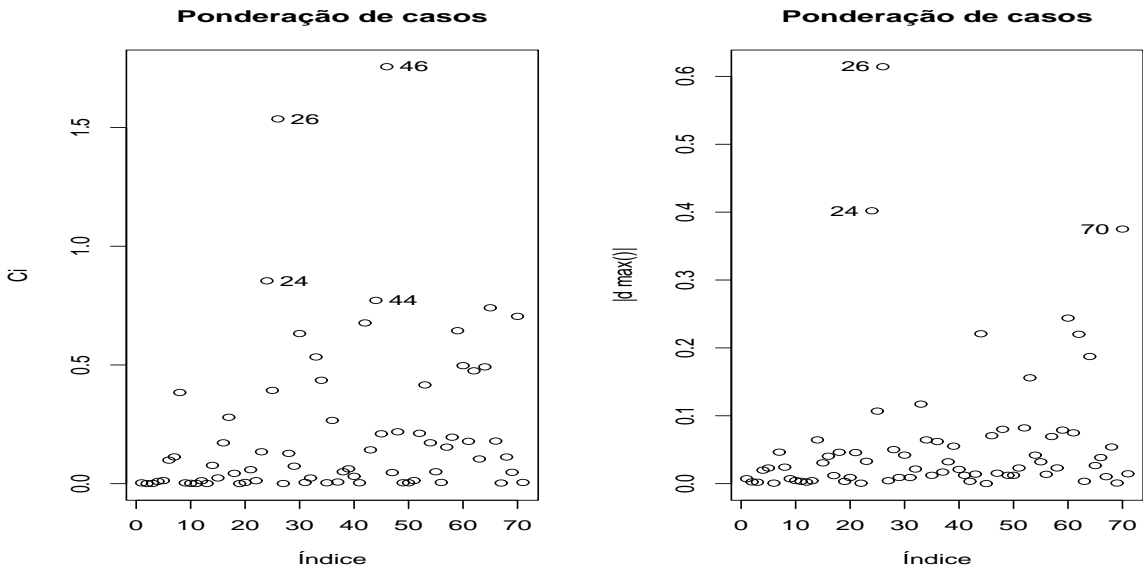
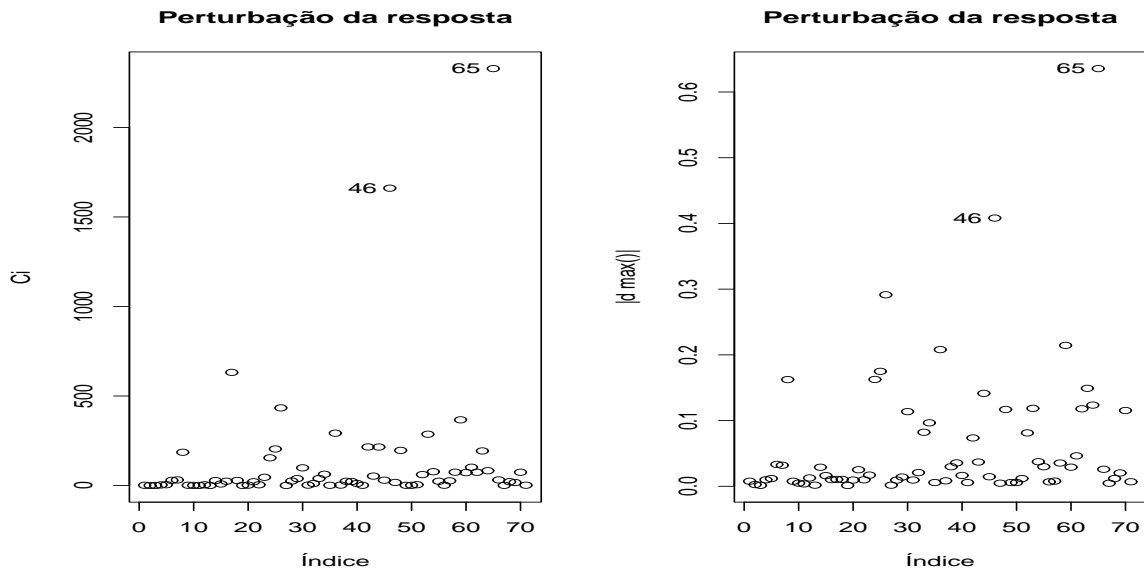


Figura 26 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Clayton segundo esquema de ponderação de casos para dados de retinopatia

### 4.2.3.2 Análise de influência global

Foram calculadas a distância generalizada de Cook ( $D_i$ ) e o afastamento da verossimilhança ( $l_i$ ) como medidas de influência global. Na Figura 28 são exibidos o gráfico do afastamento da verossimilhança versus o índice das observações e o gráfico de dispersão da dis-



**Figura 27** – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Clayton segundo esquema de perturbação de variáveis resposta para dados de retinopatia

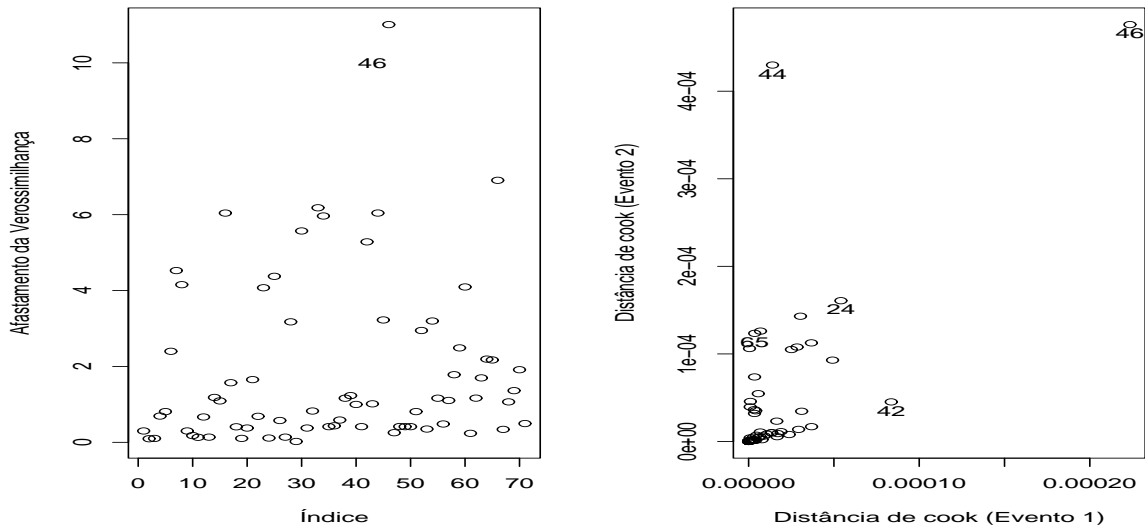
tância de cook para os dois grupos de parâmetros referentes aos dois eventos. Nesta figura, destacam-se das demais as observações #24, #44, #46 e #65 que são considerados possíveis pontos influentes.

#### 4.2.3.3 Análise de resíduos

Os gráficos exibidos na Figura 29 foram obtidos após o cálculo dos resíduos deviance para os eventos 1 e 2. Para o evento 2 pode-se notar um problema na aleatoriedade dos resíduos, que supostamente pode ser explicado pela ocorrência de uma fração de indivíduos que foram curados e não apresentarão retinopatia grave mesmo para longo período de estudo. As observações que têm resíduos fora do intervalo  $[-2, 2]$  foram investigadas em conjunto com as outras técnicas de diagnóstico previamente citadas.

#### 4.2.3.4 Impacto das observações influentes

Detectadas as observações #24, #44, #46 e #65 como possivelmente influentes foi realizada uma análise do impacto destas observações sobre o modelo utilizado. Nesta análise foram estimados os parâmetros do modelo a partir de sub-amostras com a retirada destas observações individualmente e em grupos. Os resultados desta análise estão exibidos nas Tabelas 6 e



**Figura 28** – Gráficos de medidas de influência global para modelo de regressão bivariado através de cópula de Clayton ajustado a dados de retinopatia

7. A observação #46 foi diagnosticada como influente e por isso foi desconsiderada na obtenção do modelo final.

#### 4.2.3.5 Modelo final proposto utilizando a cópula de Clayton

Mediante a eliminação da observação #46 os modelos completos para os eventos 1 e 2, como o descrito em (10) foram ajustados ao conjunto de dados. Foram selecionadas para compor os modelos finais as variáveis consideradas significativas a um nível de 10%. Desta forma os modelos finais propostos para o logaritmo dos tempos até o diagnóstico de retinopatia leve e grave em pacientes diabéticos podem ser expressos por:

$$y_{i1} = \beta_{01} + \beta_{11}x_{i11} + \beta_{21}x_{i21} + \beta_{31}x_{i31} + \beta_{41}x_{i41} + \sigma_1 z_{i1} \quad e$$

$$y_{i2} = \beta_{02} + \beta_{12}x_{i12} + \beta_{22}x_{i22} + \beta_{42}x_{i42} + \sigma_2 z_{i2}.$$

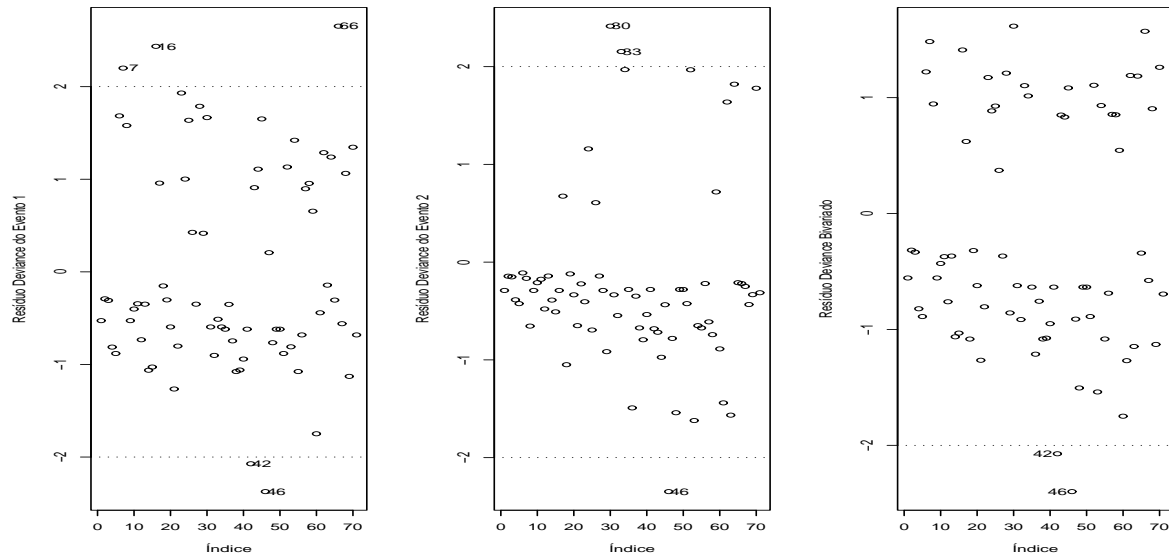
As estimativas para os parâmetros destes modelos e para o parâmetro de associação  $\alpha$  da cópula de Clayton estão apresentadas na Tabela 8. Os resultados podem ser interpretados de forma que os coeficientes de regreção com sinal positivo retardam a ocorrência do evento a que se referem, enquanto os coeficientes negativos aceleram a ocorrência dos eventos.

**Tabela 6** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até retinopatia leve utilizando a cópula de Clayton

Sub-amostra	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{31}$	$\beta_{41}$	$\beta_{51}$
A – (24)	3, 238	0, 099	-0, 021	-0, 362	-0, 242	0, 038
	[-1, 200]	[38, 259]	[-112, 245]	[8, 945]	[7, 973]	[50, 000]
	(0, 000)	(0, 396)	(0, 864)	(0, 001)	(0, 020)	(0, 739)
A – (26)	3, 204	0, 177	-0, 013	-0, 421	-0, 260	0, 074
	[-0, 150]	[-10, 958]	[-30, 612]	[-5, 754]	[1, 139]	[2, 507]
	(0, 000)	(0, 082)	(0, 907)	(0, 000)	(0, 009)	(0, 491)
A – (44)	3, 149	0, 227	-0, 194	-0, 281	-0, 297	-0, 016
	[1, 563]	[-42, 079]	[-1883, 673]	[29, 372]	[-12, 642]	[121, 504]
	(0, 000)	(0, 013)	(0, 099)	(0, 005)	(0, 002)	(0, 875)
A – (46)	3, 195	0, 155	-0, 006	-0, 384	-0, 289	0, 086
	[0, 131]	[3, 006]	[39, 796]	[3, 643]	[-9, 871]	[-13, 325]
	(0, 000)	(0, 150)	(0, 960)	(0, 000)	(0, 009)	(0, 457)
A – (24, 26)	3, 233	0, 130	-0, 027	-0, 392	-0, 241	0, 043
	[-1, 063]	[18, 785]	[-178, 571]	[1, 482]	[8, 580]	[43, 008]
	(0, 000)	(0, 248)	(0, 818)	(0, 000)	(0, 017)	(0, 698)
A – (24, 44)	3, 172	0, 195	-0, 211	-0, 256	-0, 283	-0, 043
	[0, 856]	[-22, 041]	[-2054, 082]	[35, 578]	[-7, 365]	[157, 124]
	(0, 000)	(0, 049)	(0, 097)	(0, 015)	(0, 004)	(0, 689)
A – (24, 46)	3, 229	0, 104	-0, 022	-0, 357	-0, 260	0, 045
	[-0, 916]	[35, 003]	[-120, 408]	[10, 302]	[1, 291]	[40, 106]
	(0, 000)	(0, 380)	(0, 864)	(0, 001)	(0, 025)	(0, 707)
A – (26, 44)	3, 159	0, 234	-0, 185	-0, 295	-0, 299	-0, 020
	[1, 272]	[-46, 399]	[-1783, 673]	[25, 930]	[-13, 554]	[126, 649]
	(0, 000)	(0, 008)	(0, 105)	(0, 003)	(0, 001)	(0, 840)
A – (26, 46)	3, 210	0, 150	-0, 011	-0, 383	-0, 297	0, 089
	[-0, 334]	[6, 074]	[-12, 245]	[3, 668]	[-12, 680]	[-17, 678]
	(0, 000)	(0, 144)	(0, 924)	(0, 000)	(0, 006)	(0, 429)
A – (44, 46)	3, 144	0, 225	-0, 204	-0, 262	-0, 319	-0, 019
	[1, 735]	[-41, 014]	[-1980, 612]	[34, 070]	[-21, 147]	[125, 462]
	(0, 000)	(0, 013)	(0, 092)	(0, 011)	(0, 002)	(0, 858)
A – (24, 26, 44)	3, 179	0, 207	-0, 198	-0, 272	-0, 292	-0, 042
	[0, 631]	[-29, 430]	[-1922, 449]	[31, 709]	[-10, 972]	[154, 749]
	(0, 000)	(0, 029)	(0, 107)	(0, 010)	(0, 003)	(0, 695)
A – (24, 26, 46)	3, 236	0, 116	-0, 028	-0, 366	-0, 283	0, 065
	[-1, 153]	[27, 176]	[-186, 735]	[8, 040]	[-7, 517]	[13, 984]
	(0, 000)	(0, 301)	(0, 816)	(0, 001)	(0, 012)	(0, 571)
A – (24, 44, 46)	3, 164	0, 201	-0, 229	-0, 245	-0, 296	-0, 053
	[1, 097]	[-25, 986]	[-2238, 776]	[38, 467]	[-12, 491]	[170, 317]
	(0, 000)	(0, 041)	(0, 077)	(0, 026)	(0, 005)	(0, 636)
A – (26, 44, 46)	3, 161	0, 219	-0, 192	-0, 263	-0, 340	-0, 010
	[1, 210]	[-37, 383]	[-1863, 265]	[33, 920]	[-29, 119]	[113, 588]
	(0, 000)	(0, 013)	(0, 098)	(0, 012)	(0, 001)	(0, 921)
A – (24, 26, 44, 46)	3, 181	0, 198	-0, 209	-0, 247	-0, 329	-0, 036
	[0, 584]	[-24, 170]	[-2030, 612]	[37, 965]	[-24, 753]	[147, 230]
	(0, 000)	(0, 038)	(0, 096)	(0, 026)	(0, 002)	(0, 739)

**Tabela 7** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até retinopatia grave utilizando a cópula de Clayton

Sub-amostra	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{32}$	$\beta_{42}$	$\beta_{52}$	$\alpha$
A – (24)	3,460	0,154	-0,134	-0,073	-0,294	0,033	7,251
	[-1,714]	[42,692]	[-23,955]	[48,188]	[16,302]	[69,954]	[7,578]
	(0,000)	(0,306)	(0,310)	(0,539)	(0,036)	(0,812)	(0,015)
A – (26)	3,488	0,238	-0,173	-0,175	-0,389	0,115	10,073
	[-2,528]	[11,260]	[-60,631]	[-24,449]	[-10,725]	[-4,749]	[-28,393]
	(0,000)	(0,044)	(0,182)	(0,082)	(0,004)	(0,316)	(0,006)
A – (44)	3,288	0,355	-0,306	-0,004	-0,336	-0,015	7,443
	[3,348]	[-32,364]	[-184,587]	[97,370]	[4,381]	[113,333]	[5,125]
	(0,000)	(0,000)	(0,013)	(0,972)	(0,004)	(0,904)	(0,005)
A – (46)	3,523	0,299	-0,182	-0,180	-0,369	0,022	9,137
	[-3,566]	[-11,372]	[-69,266]	[-28,287]	[-4,865]	[79,543]	[-16,471]
	(0,000)	(0,036)	(0,239)	(0,179)	(0,046)	(0,889)	(0,015)
A – (24, 26)	3,546	0,152	-0,220	-0,130	-0,357	0,069	9,451
	[-4,230]	[43,326]	[-104,457]	[7,747]	[-1,593]	[37,352]	[-20,463]
	(0,000)	(0,269)	(0,135)	(0,254)	(0,014)	(0,588)	(0,011)
A – (24, 44)	3,323	0,309	-0,358	0,046	-0,318	-0,068	7,146
	[2,316]	[-15,063]	[-232,126]	[132,480]	[9,474]	[162,009]	[8,911]
	(0,000)	(0,004)	(0,006)	(0,680)	(0,007)	(0,580)	(0,007)
A – (24, 46)	3,592	0,193	-0,236	-0,128	-0,299	-0,064	8,436
	[-5,606]	[28,225]	[-118,849]	[9,097]	[15,050]	[157,991]	[-7,527]
	(0,000)	(0,265)	(0,183)	(0,416)	(0,139)	(0,727)	(0,035)
A – (26, 44)	3,361	0,314	-0,344	-0,024	-0,362	-0,019	9,314
	[1,208]	[-17,040]	[-219,406]	[83,156]	[-2,987]	[117,808]	[-18,727]
	(0,000)	(0,001)	(0,006)	(0,824)	(0,002)	(0,868)	(0,005)
A – (26, 46)	3,655	0,229	-0,266	-0,188	-0,430	0,028	12,727
	[-7,435]	[14,728]	[-147,168]	[-33,333]	[-22,248]	[74,612]	[-62,228]
	(0,000)	(0,118)	(0,128)	(0,186)	(0,023)	(0,864)	(0,009)
A – (44, 46)	3,422	0,396	-0,420	-0,037	-0,351	-0,131	8,577
	[-0,585]	[-47,614]	[-289,879]	[73,490]	[0,256]	[219,178]	[-9,331]
	(0,000)	(0,001)	(0,021)	(0,790)	(0,036)	(0,462)	(0,009)
A – (24, 26, 44)	3,397	0,273	-0,399	0,023	-0,354	-0,069	9,221
	[0,150]	[-1,603]	[-270,009]	[116,489]	[-0,654]	[162,557]	[-17,541]
	(0,000)	(0,007)	(0,002)	(0,829)	(0,003)	(0,554)	(0,007)
A – (24, 26, 46)	3,767	0,137	-0,363	-0,149	-0,417	-0,034	13,157
	[-10,739]	[48,919]	[-236,769]	[-5,757]	[-18,720]	[131,050]	[-67,703]
	(0,000)	(0,403)	(0,065)	(0,304)	(0,039)	(0,836)	(0,015)
A – (24, 44, 46)	3,480	0,353	-0,511	0,009	-0,317	-0,221	8,218
	[-2,296]	[-31,767]	[-374,745]	[106,468]	[9,758]	[301,644]	[-4,747]
	(0,000)	(0,008)	(0,014)	(0,951)	(0,068)	(0,267)	(0,013)
A – (26, 44, 46)	3,546	0,328	-0,480	-0,035	-0,426	-0,124	11,946
	[-4,236]	[-22,446]	[-345,311]	[75,124]	[-21,195]	[213,242]	[-52,271]
	(0,000)	(0,009)	(0,010)	(0,805)	(0,017)	(0,460)	(0,008)
A – (24, 26, 44, 46)	3,604	0,282	-0,565	0,018	-0,413	-0,201	11,966
	[-5,956]	[-5,295]	[-424,234]	[112,722]	[-17,639]	[283,470]	[-52,520]
	(0,000)	(0,035)	(0,004)	(0,900)	(0,024)	(0,247)	(0,011)



**Figura 29** – Gráficos de resíduos Deviance obtidos para modelo de regressão bivariado através de cópula de Clayton ajustado a dados de retinopatia

Uma análise gráfica da qualidade do ajuste dos modelos propostos para os eventos em estudo pode ser realizada na Figura 30 onde estão exibidos os gráficos de envelopes simulados para os resíduos deviance para os eventos 1 e 2, e também na Figura 31 que exibe um gráfico de dispersão e um gráfico Normal de probabilidades para o resíduo do ajuste bivariado. Pode-se notar que existem problemas na modelagem que devem ser decorrentes do problema de fração de curados na amostra. A função de Clayton, entretanto, parece modelar a correlação de forma satisfatória e a estimativa para o  $\tau$  de Kendall igual a 0,78 obtida através da estimativa do parâmetro  $\alpha$  é próxima da medida calculada na análise exploratória dos dados.

### 4.3 Aplicação com os dados de infecção em pacientes com insuficiência renal

Nas seguintes seções são apresentados os resultados da análise do conjunto de dados descrito na Seção 3.1.2 sob diferentes abordagens. Na seção 4.3.1 estão os resultados de uma análise exploratória inicial dos dados. Na seção 4.3.2 os resultados referentes a utilização da cópula de Frank e na seção 4.3.3 aqueles referentes a utilização da cópula de Clayton.

**Tabela 8** – Estimativas para o modelo final com a cópula de Clayton para os dados de retinopatia

Parâmetro	Estimativa	Erro padrão	p-valor
$\beta_{01}$	3,1417	0,0891	< 0,0001
$\beta_{11}$	0,2295	0,0892	0,0101
$\beta_{21}$	-0,1895	0,1046	0,0700
$\beta_{31}$	-0,2778	0,0685	0,0000
$\beta_{41}$	-0,2999	0,0917	0,0011
$\sigma_1$	0,2626	0,0335	< 0,0001
$\beta_{02}$	3,2812	0,1252	< 0,0001
$\beta_{12}$	0,3576	0,0926	0,0001
$\beta_{22}$	-0,3027	0,0969	0,0018
$\beta_{42}$	-0,3393	0,1121	0,0025
$\sigma_2$	0,2060	0,0401	< 0,0001
$\alpha$	7,3909	2,5712	0,0040

### 4.3.1 Análise exploratória

Inicialmente foi realizada uma análise exploratória dos dados. A Figura 32 apresenta um gráfico de dispersão dos valores observados para os tempos, em semanas, até a ocorrência de dois eventos de infecção nos pacientes. Por simplicidade de notação, os distintos eventos de infecção são muitas vezes referidos por eventos 1 e 2. O coeficiente de correlação de Tau de Kendall calculado para os dados foi  $\tau = 0,031$ , indicando ausência de associação entre os eventos tal como mostra a figura 32. Os percentuais de observações censuradas nos eventos 1 e 2 foram 29% e 18% respectivamente.

A Figura 33 apresenta gráficos com os ajustes marginais de modelos com distribuições Weibull para os tempos de falha e as estimativas não-paramétrica para os mesmos. As estimativas não paramétricas por grupos para as diferentes covariáveis podem ser vistas nas Figuras 34, 35, 36 e 37. Estes gráficos podem ser utilizados como indicativos sobre a significância das covariáveis nos modelos de regressão. A análise destas figuras indica que os modelos de Cox não devem ser utilizados pois a suposição de riscos proporcionais não pode ser sustentada, uma vez que as curvas de sobrevivência para os diferentes grupos se cruzam para diversas variáveis explicativas. É neste contexto que a utilização dos modelos de locação e escala na forma proposta se mostra vantajosa.



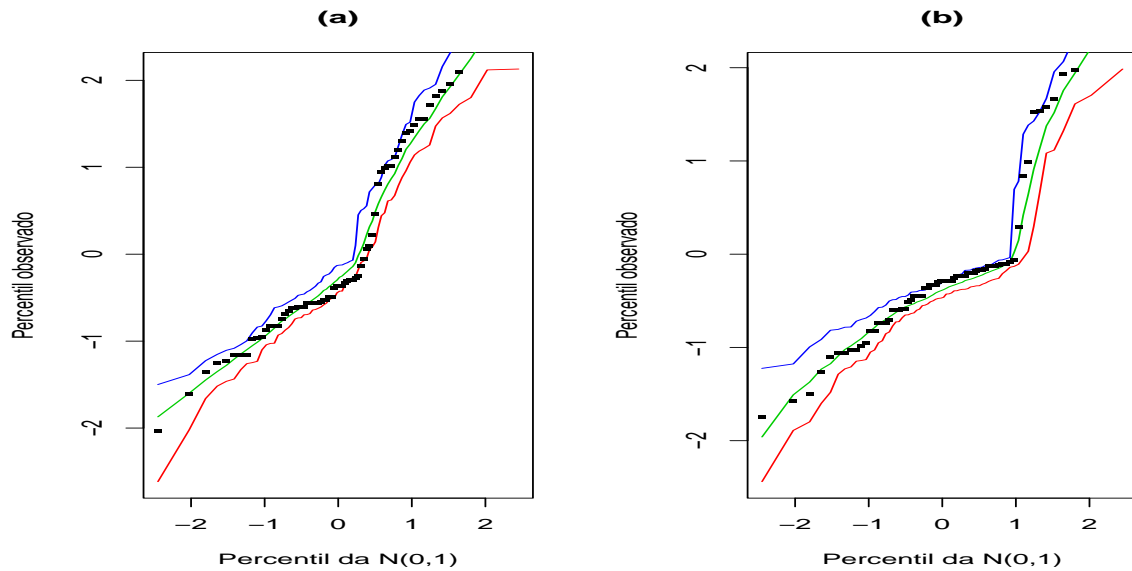


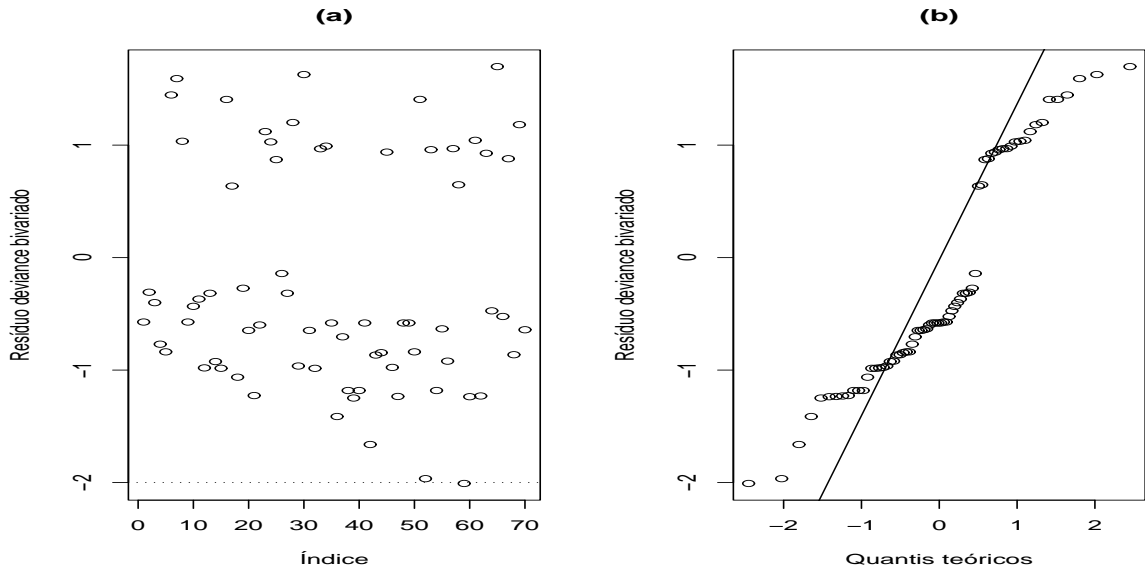
Figura 30 – Gráficos de envelopes simulados do resíduo Deviance marginal dos dados de retinopatia

### 4.3.2 Modelo de regressão bivariado através da cópula de Frank

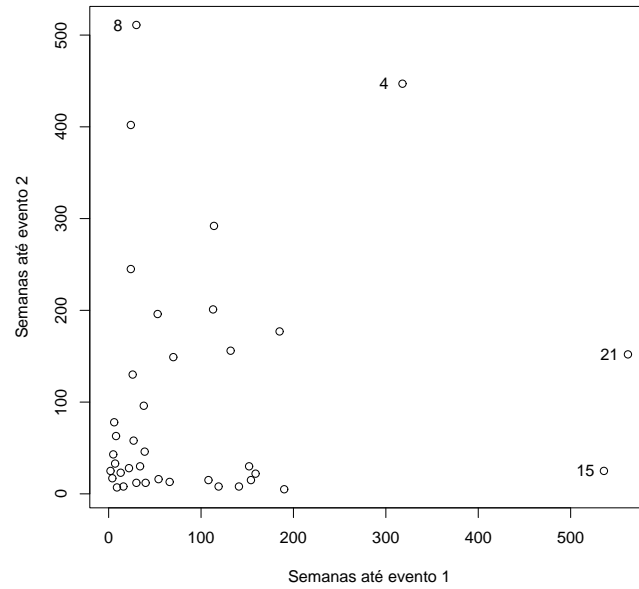
Nesta seção são apresentados os resultados da utilização da cópula de Frank para reunir dois modelos como o descrito em (10). Os modelos são referentes aos logaritmos dos tempos até a ocorrência de dois eventos de infecção em pacientes com insuficiência renal. Como visto nos resultados da análise exploratória, é razoável supor que os logaritmos dos tempos até os eventos 1 e 2 têm distribuição valor extremo. Os resultados da estimação dos parâmetros de regressão e do parâmetro de associação da cópula de Frank estão apresentados na Tabela 9. Nesta tabela aparecem os resultados das estimativas, erros padrões e significâncias dos parâmetros obtidos pelos métodos de máxima verossimilhança e Jackknife descritos na Seção 3.4. Para o evento 1 foram consideradas significativas as covariáveis referentes ao sexo do paciente e a presença de doença do tipo AN e PKD. Para o evento 2 as covariáveis referentes a sexo e PKD foram significativas, isto indica que existem diferenças significativas entre os sexos em relação ao tempo até ocorrência de infecção.

#### 4.3.2.1 Análise de influência local

As Figuras 38, 39 e 40 apresentam gráficos das medidas de influência local contra os índices das observações para os esquemas de perturbação descritos na Seção 3.4.3. As observações

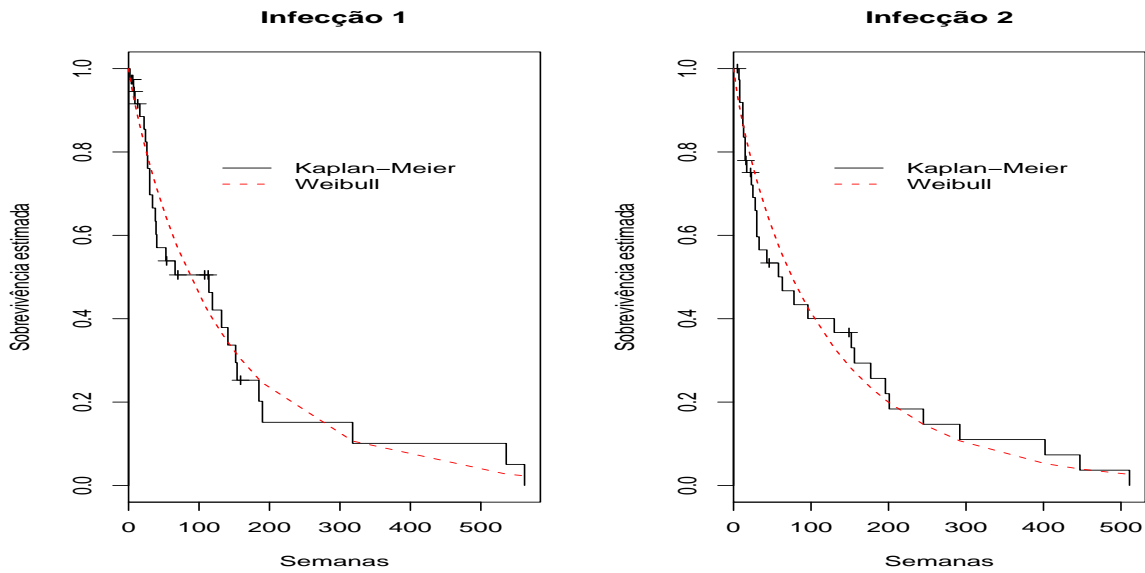


**Figura 31** – (a) Gráfico do resíduo bivariado e (b) Gráfico Normal de probabilidades do resíduo bivariado com cópula de Clayton nos dados de retinopatia.

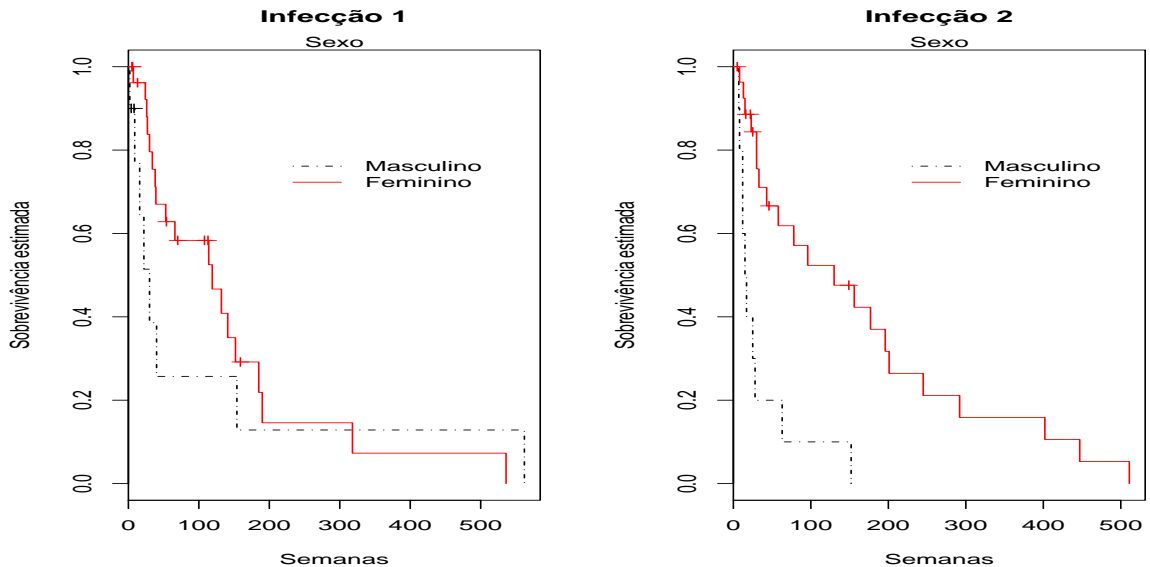


**Figura 32** – Gráfico de dispersão das variáveis resposta nos dados de infecção

#8, #10, #15 e #21 destacam-se das demais e são possíveis pontos influentes.



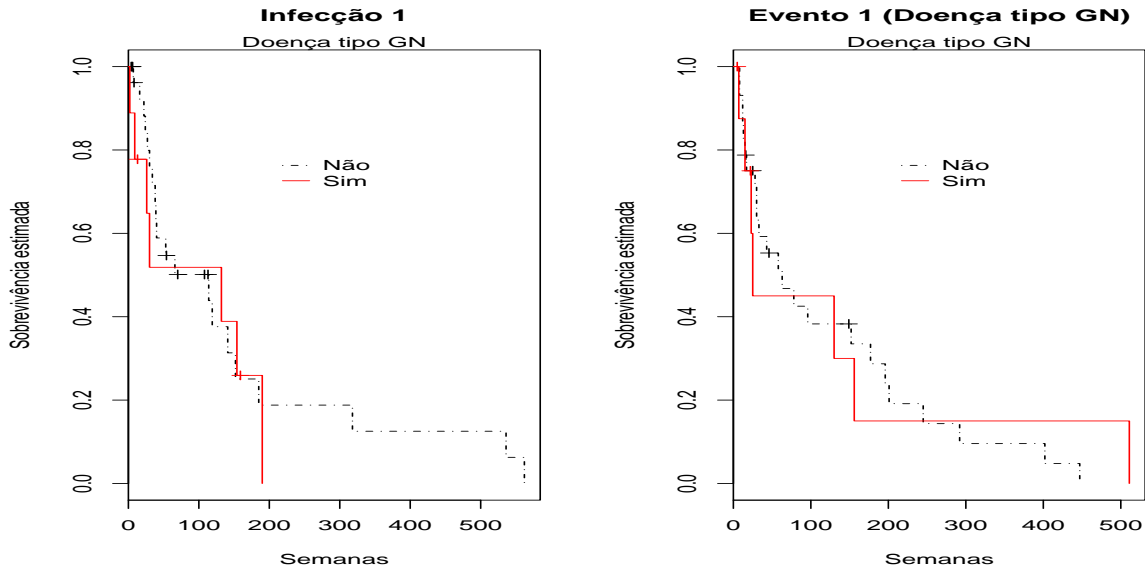
**Figura 33** – Curvas de sobrevivência estimadas marginalmente por Kaplan-Meier e pela distribuição Weibull para os dados de infecção



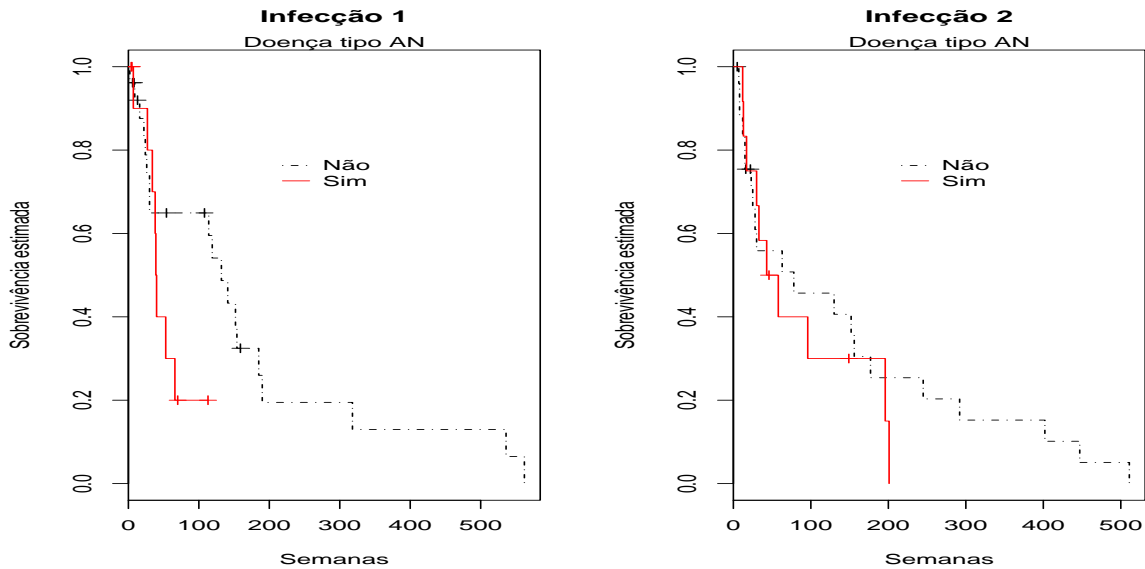
**Figura 34** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de infecção, para cada evento segundo sexo

#### 4.3.2.2 Análise de influência global

Foram calculadas a distância generalizada de Cook ( $D_i$ ) e o afastamento da verossimilhança ( $l_i$ ) como medidas de influência global. Na Figura 41 são exibidos o gráfico do afastamento da verossimilhança versus o índice das observações e o gráfico de dispersão

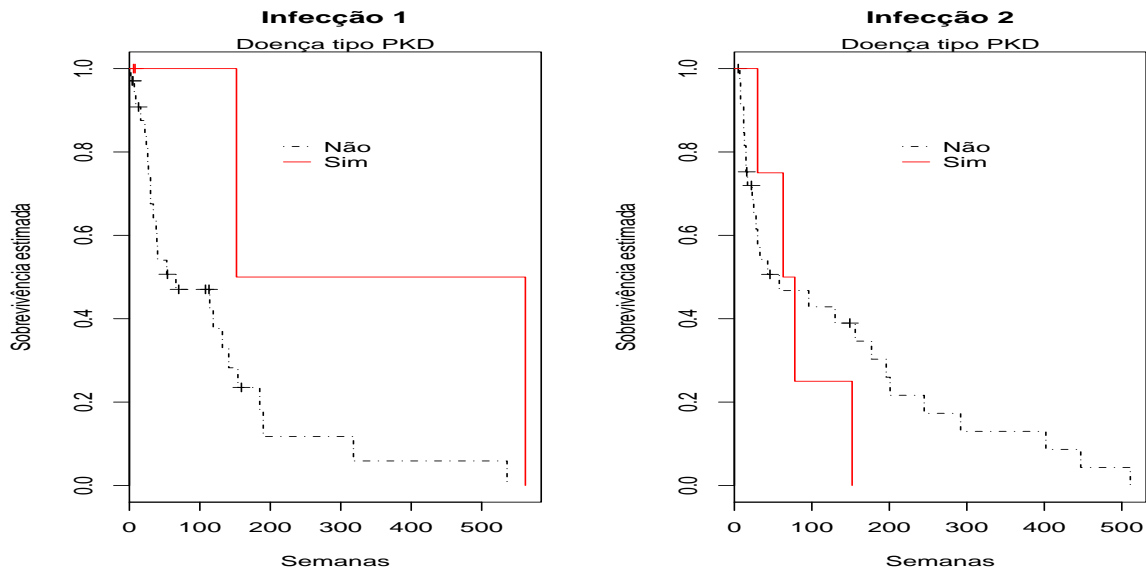


**Figura 35** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de infecção, para cada evento segundo presença de doença tipo GN



**Figura 36** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de infecção, para cada evento segundo presença de doença tipo AN

das distâncias de cook calculadas para dois grupos de parâmetros  $\theta_1 = (\beta_{01}, \beta_{11}, \dots, \beta_{51}, \sigma_1)$  e  $\theta_2 = (\beta_{02}, \beta_{12}, \dots, \beta_{52}, \sigma_2)$  referentes aos dois eventos. Nesta figura, destacam-se das demais as observações #10, #21 e #31 que podem ser considerados possíveis pontos influentes.



**Figura 37** – Estimativas de Kaplan-Meier das funções de sobrevivência para os dados de infecção, para cada evento segundo presença de doença tipo PKD

**Tabela 9** – Estimativas de máxima verossimilhança (EMV) e Jackknife dos parâmetros do modelo de regressão bivariado através da cópula de Frank para os dados de infecção

Parâmetro	EMV			Jackknife		
	Estimativa	Erro padrão	p-valor	Estimativa	Erro padrão	p-valor
$\beta_{01}$	4,2292	0,6238	< 0,0000	4,5006	0,8701	< 0,0000
$\beta_{11}$	0,0002	0,0138	0,9895	0,0025	0,0170	< 0,8821
$\beta_{21}$	1,0091	0,4270	0,0181	0,6625	0,7669	0,3877
$\beta_{31}$	-0,2403	0,5108	0,6381	-0,1855	0,5181	< 0,7203
$\beta_{41}$	-1,0107	0,5102	0,0476	-1,1429	0,5426	< 0,0352
$\beta_{51}$	1,4697	0,7218	0,0417	2,0193	1,5772	< 0,2004
$\sigma_1$	0,8648	0,1279	< 0,0000	1,0460	0,1414	< 0,0000
$\beta_{02}$	3,2404	0,5896	< 0,0000	3,2762	0,7038	< 0,0000
$\beta_{12}$	0,0030	0,0134	0,8205	0,0033	0,0134	< 0,8067
$\beta_{22}$	2,1221	0,3870	0,0000	2,1357	0,4225	0,0000
$\beta_{32}$	-0,1603	0,4936	0,7453	-0,1264	0,5680	< 0,8239
$\beta_{42}$	-0,8166	0,4732	0,0844	-0,9024	0,4346	< 0,0379
$\beta_{52}$	0,7314	0,6232	0,2406	1,0533	1,0328	0,3078
$\sigma_2$	0,8256	0,1208	< 0,0000	0,9222	0,1369	< 0,0000
$\alpha$	1,9741	1,2387	0,1110	2,1908	1,5982	< 0,1704

#### 4.3.2.3 Análise de resíduos

Os gráficos exibidos na figura 42 foram obtidos após o cálculo dos resíduos de-viance marginais para os eventos 1 e 2 e o resíduo bivariado. Pela análise da figura não há

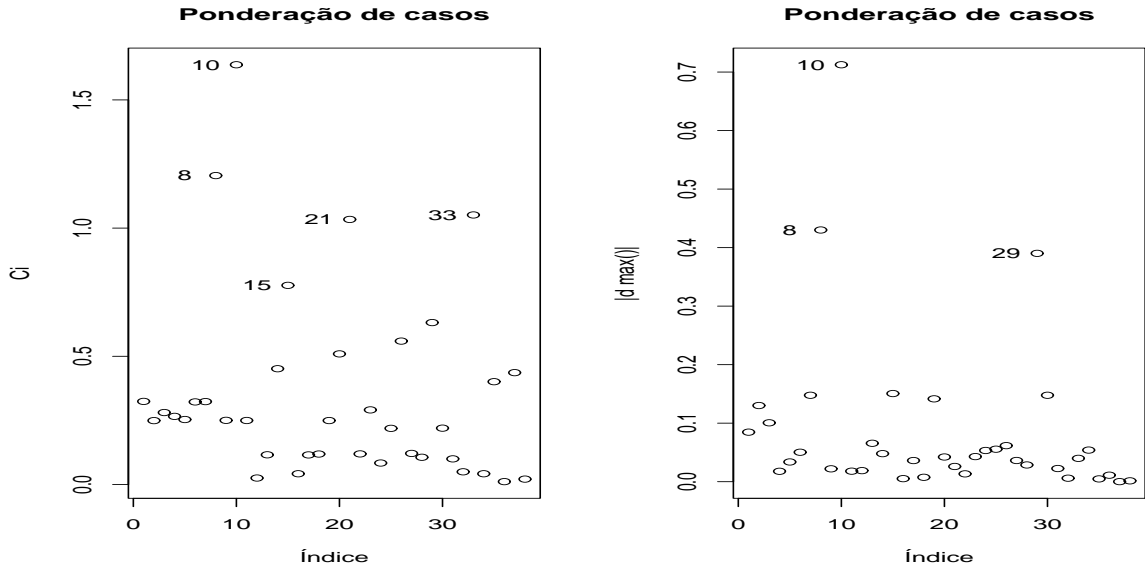


Figura 38 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Frank segundo esquema de ponderação de casos para dados de infecção

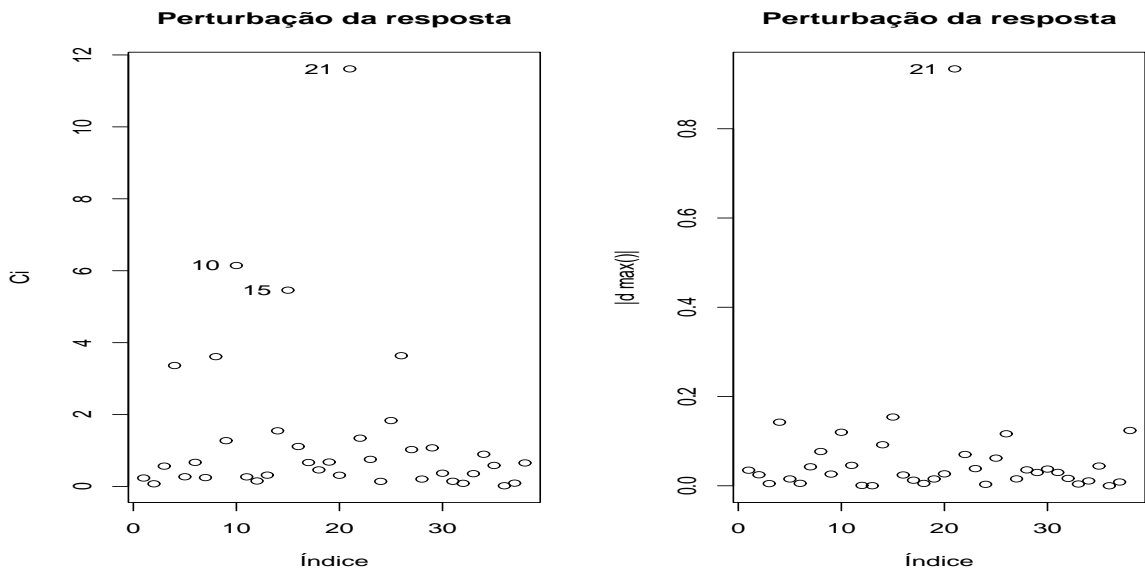
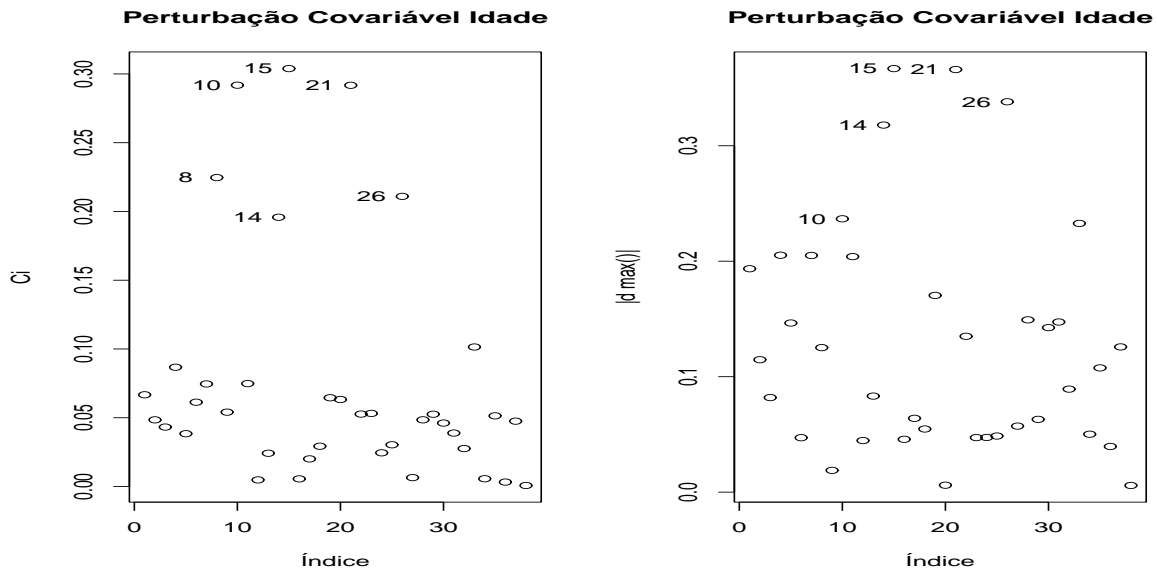


Figura 39 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Frank segundo esquema de ponderação de casos para dados de infecção



**Figura 40** – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Frank segundo esquema de perturbação da covariável idade para dados de infecção

evidências contra as suposições adotadas para os modelos marginais, uma vez que os resíduos têm comportamento aparentemente aleatório e simétrico. As observações que têm resíduos fora do intervalo  $[-2, 2]$  foram investigadas em conjunto com as outras técnicas de diagnóstico previamente citadas. Pode-se notar que o resíduo bivariado proposto tem poder para detectar como possivelmente influentes os mesmos pontos que as análises de influência local e global.

#### 4.3.2.4 Impacto das observações influentes

Detectadas as observações #8, #10, #15 e #21 como possíveis pontos influentes é de interesse detectar o impacto destas observações no modelo. Para isso foram estimados os parâmetros do modelo a partir de sub-amostras com a retirada dessas observações individualmente e em grupos. Estas estimativas estão apresentadas nas Tabelas 10 e 11 que contém as estimativas dos parâmetros, as mudanças relativas em relação às estimativas da amostra completa e os p-valores para cada sub-amostra.

Mediante a análise das tabelas foram diagnosticadas as observações #8 e #15 como pontos influentes. A observação #8 é referente a paciente do sexo feminino, com 55 anos de idade e portador da doença do tipo GN cujas infecções ocorreram após 30 e 511 semanas de tratamento. A observação #15 é referente a paciente do sexo feminino, com 17 anos de idade,

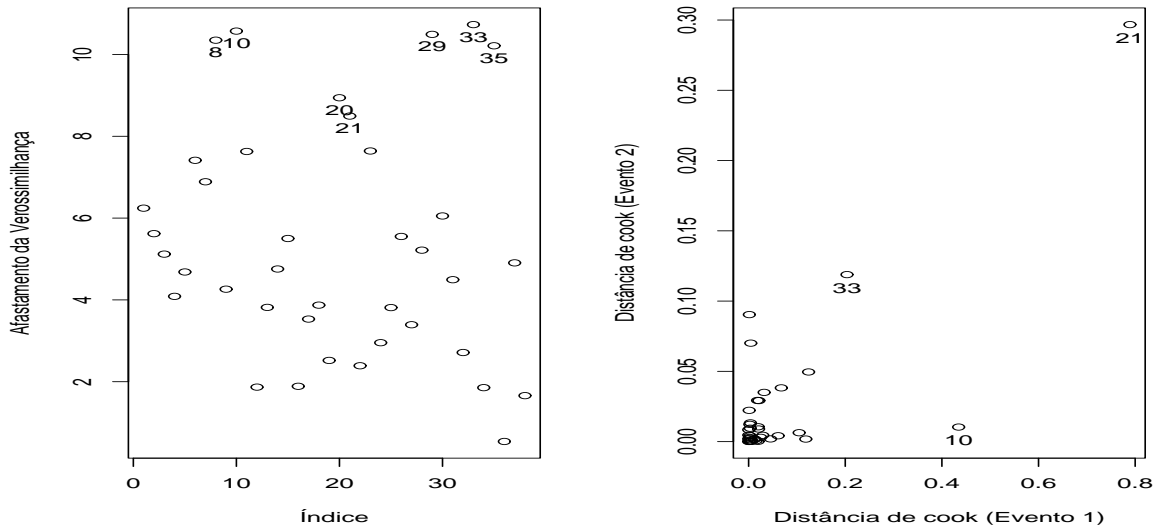
**Tabela 10** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até infecção 1

Sub-amostra	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{31}$	$\beta_{41}$	$\beta_{51}$
	4,198	0,002	1,009	-0,183	-1,051	1,425
$A - (8)$	[0,724] (0,000)	[-800,000] (0,897)	[-0,020] (0,018)	[24,138] (0,724)	[-3,986] (0,037)	[3,075] (0,044)
	4,042	-0,010	1,610	-0,553	-0,722	2,129
$A - (10)$	[4,427] (0,000)	[5300,000] (0,416)	[-59,512] (0,000)	[-129,580] (0,206)	[28,632] (0,135)	[-44,887] (0,002)
	3,717	0,010	0,903	-0,081	-0,933	1,542
$A - (15)$	[12,105] (0,000)	[-5050,000] (0,424)	[10,513] (0,020)	[66,265] (0,853)	[7,685] (0,035)	[-4,940] (0,019)
	4,011	0,003	1,158	-0,218	-1,039	0,127
$A - (21)$	[5,162] (0,000)	[-1350,000] (0,832)	[-14,734] (0,005)	[9,348] (0,670)	[-2,779] (0,041)	[91,366] (0,908)
	4,025	-0,008	1,561	-0,504	-0,791	2,039
$A - (8,10)$	[4,833] (0,000)	[4150,000] (0,531)	[-54,707] (0,000)	[-109,223] (0,261)	[21,808] (0,103)	[-38,743] (0,003)
	3,690	0,012	0,891	-0,019	-0,976	1,492
$A - (8,15)$	[12,755] (0,000)	[-5950,000] (0,335)	[11,732] (0,020)	[92,189] (0,965)	[3,521] (0,023)	[-1,504] (0,018)
	3,958	0,004	1,184	-0,144	-1,082	0,144
$A - (8,21)$	[6,410] (0,000)	[-2150,000] (0,744)	[-17,301] (0,004)	[40,008] (0,784)	[-7,012] (0,032)	[90,222] (0,894)
	3,692	-0,002	1,452	-0,420	-0,707	2,101
$A - (10,15)$	[12,689] (0,000)	[950,000] (0,896)	[-43,886] (0,001)	[-74,366] (0,306)	[30,056] (0,109)	[-42,982] (0,001)
	3,823	-0,009	1,795	-0,581	-0,679	0,382
$A - (10,21)$	[9,596] (0,000)	[4450,000] (0,478)	[-77,844] (0,000)	[-141,504] (0,163)	[32,845] (0,145)	[74,022] (0,689)
	3,498	0,013	1,040	-0,043	-0,941	0,096
$A - (15,21)$	[17,274] (0,000)	[-6350,000] (0,295)	[-3,092] (0,004)	[82,302] (0,920)	[6,963] (0,029)	[93,461] (0,922)
	3,669	0,001	1,381	-0,347	-0,775	2,002
$A - (8,10,15)$	[13,251] (0,000)	[-450,000] (0,930)	[-36,821] (0,001)	[-44,121] (0,403)	[23,331] (0,075)	[-36,218] (0,002)
	3,802	-0,007	1,765	-0,521	-0,733	0,344
$A - (8,10,21)$	[10,102] (0,000)	[3550,000] (0,579)	[-74,901] (0,000)	[-116,369] (0,225)	[27,554] (0,120)	[76,560] (0,717)
	3,445	0,015	1,053	0,031	-0,987	0,096
$A - (8,15,21)$	[18,541] (0,000)	[-7300,000] (0,219)	[-4,300] (0,004)	[112,754] (0,942)	[2,413] (0,018)	[93,448] (0,920)
	3,479	0,000	1,618	-0,427	-0,657	0,332
$A - (10,15,21)$	[17,737] (0,000)	[100,000] (0,998)	[-60,305] (0,000)	[-77,441] (0,265)	[34,992] (0,117)	[77,438] (0,713)
	3,446	0,002	1,567	-0,346	-0,714	0,284
$A - (8,10,15,21)$	[18,522] (0,000)	[-1100,000] (0,839)	[-55,222] (0,000)	[-43,955] (0,374)	[29,423] (0,087)	[80,649] (0,749)



**Tabela 11** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até à infecção 2

Sub-amostra	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{32}$	$\beta_{42}$	$\beta_{52}$	$\alpha$
	3,588	-0,003	1,962	-0,554	-0,734	0,692	2,339
$A - (8)$	[-10,744] (0,000)	[183,333] (0,842)	[7,543] (0,000)	[-246,096] (0,247)	[10,176] (0,103)	[5,584] (0,236)	[-18,531] (0,073)
	3,244	0,003	2,113	0,029	-0,768	0,734	2,510
$A - (10)$	[-0,114] (0,000)	[10,000] (0,850)	[0,466] (0,000)	[117,989] (0,958)	[5,890] (0,115)	[-0,218] (0,250)	[-27,207] (0,065)
	3,098	0,006	2,122	-0,181	-0,823	0,732	1,792
$A - (15)$	[4,395] (0,000)	[-96,667] (0,654)	[0,019] (0,000)	[-13,117] (0,704)	[-0,796] (0,075)	[0,068] (0,235)	[9,192] (0,129)
	2,966	0,009	2,189	-0,252	-0,897	-0,148	1,201
$A - (21)$	[8,448] (0,000)	[-196,667] (0,471)	[-3,128] (0,000)	[-57,527] (0,578)	[-9,833] (0,046)	[120,248] (0,827)	[39,120] (0,369)
	3,524	-0,002	1,983	-0,358	-0,698	0,706	2,720
$A - (8,10)$	[-8,778] (0,000)	[163,333] (0,888)	[6,558] (0,000)	[-123,673] (0,507)	[14,560] (0,136)	[3,591] (0,249)	[-37,808] (0,066)
	3,449	0,000	1,955	-0,568	-0,738	0,690	2,150
$A - (8,15)$	[-6,444] (0,000)	[86,667] (0,975)	[7,892] (0,000)	[-254,778] (0,214)	[9,576] (0,090)	[5,762] (0,230)	[-8,959] (0,081)
	3,326	0,003	2,028	-0,601	-0,795	-0,096	1,553
$A - (8,21)$	[-2,660] (0,000)	[10,000] (0,818)	[4,452] (0,000)	[-275,453] (0,173)	[2,621] (0,065)	[113,135] (0,879)	[21,303] (0,266)
	3,110	0,006	2,093	-0,002	-0,786	0,714	2,281
$A - (10,15)$	[4,006] (0,000)	[-93,333] (0,680)	[1,409] (0,000)	[99,063] (0,998)	[3,723] (0,099)	[2,499] (0,260)	[-15,577] (0,081)
	2,968	0,009	2,169	-0,121	-0,857	-0,152	1,534
$A - (10,21)$	[8,389] (0,000)	[-196,667] (0,499)	[-2,181] (0,000)	[24,485] (0,806)	[-4,923] (0,061)	[120,767] (0,827)	[22,251] (0,285)
	2,876	0,011	2,192	-0,262	-0,897	-0,162	1,073
$A - (15,21)$	[11,244] (0,000)	[-256,667] (0,385)	[-3,265] (0,000)	[-63,523] (0,553)	[-9,858] (0,043)	[122,133] (0,810)	[45,627] (0,398)
	3,403	0,001	1,954	-0,387	-0,712	0,686	2,458
$A - (8,10,15)$	[-5,040] (0,000)	[66,667] (0,937)	[7,915] (0,000)	[-141,849] (0,456)	[12,870] (0,118)	[6,363] (0,258)	[-24,582] (0,079)
	3,279	0,003	2,039	-0,481	-0,748	-0,079	1,656
$A - (8,10,21)$	[-1,194] (0,000)	[3,333] (0,815)	[3,929] (0,000)	[-200,625] (0,333)	[8,401] (0,091)	[110,800] (0,905)	[16,069] (0,284)
	3,228	0,005	2,030	-0,603	-0,796	-0,110	1,406
$A - (8,15,21)$	[0,358] (0,000)	[-53,333] (0,697)	[4,372] (0,000)	[-276,889] (0,157)	[2,572] (0,060)	[115,074] (0,860)	[28,747] (0,287)
	2,881	0,011	2,158	-0,138	-0,869	-0,185	1,382
$A - (10,15,21)$	[11,090] (0,000)	[-263,333] (0,402)	[-1,682] (0,000)	[13,616] (0,773)	[-6,417] (0,054)	[125,273] (0,789)	[29,953] (0,314)
	3,198	0,005	2,025	-0,491	-0,759	-0,110	1,500
$A - (8,10,15,21)$	[1,293] (0,000)	[-63,333] (0,699)	[4,589] (0,000)	[-206,933] (0,306)	[7,029] (0,082)	[115,033] (0,867)	[23,994] (0,307)



**Figura 41** – Gráficos de medidas de influência global para modelo de regressão bivariado através de cópula de Frank ajustado a dados de infecção

que não apresenta nenhuma das doenças investigadas e teve uma infecção após 536 semanas e teve a observação da outra infecção censurada após 25 semanas do tratamento. Para ambas as pacientes foram observados tempos elevados para um evento e tempos curtos para o outro. Dada sua influência na modelagem proposta, estas observações foram desconsideradas nas demais etapas da análise.

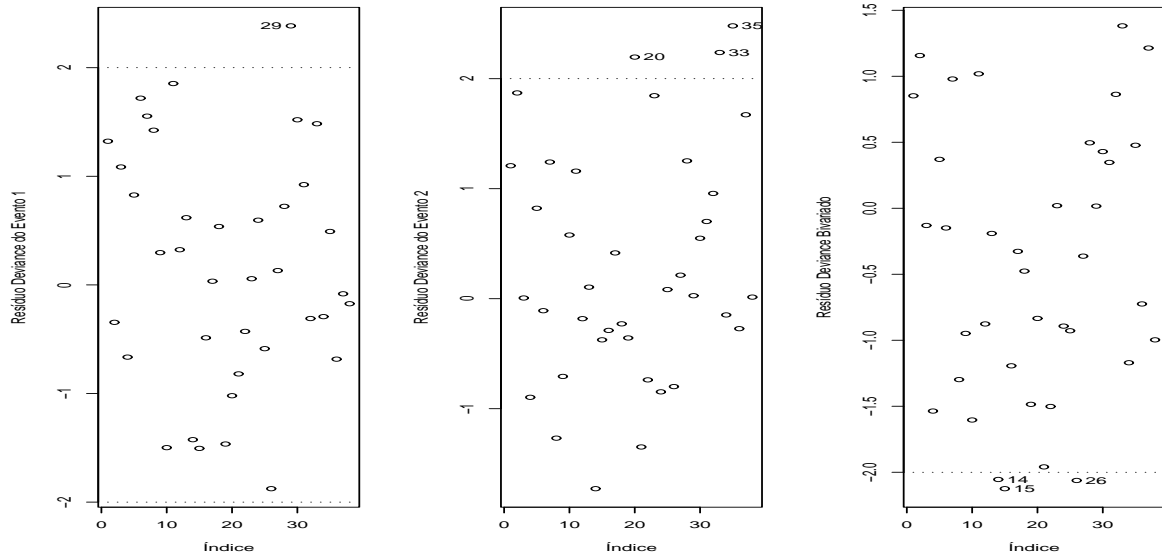
#### 4.3.2.5 Modelo final proposto utilizando a cópula de Frank

Mediante a eliminação das observações #8 e #15 os modelos completos para os eventos 1 e 2, como o descrito em (10) foram ajustados ao conjunto de dados. Foram selecionadas para compor os modelos as variáveis consideradas significativas a um nível de 10%. Desta forma os modelos finais propostos para o logaritmo dos tempos até as infecções 1 e 2 nos pacientes com insuficiência renal podem ser expressos por:

$$y_{i1} = \beta_{01} + \beta_{21}x_{i2} + \beta_{31}x_{i3} + \beta_{41}x_{i4} + \beta_{51}x_{i5} + \sigma_1 z_{i1} \quad e$$

$$y_{i2} = \beta_{02} + \beta_{22}x_{i2} + \beta_{32}x_{i3} + \beta_{42}x_{i4} + \beta_{52}x_{i5} + \sigma_2 z_{i2}.$$

As estimativas para os parâmetros destes modelos e para o parâmetro de associação  $\alpha$  da cópula de Frank estão apresentadas na Tabela 12. Os resultados podem ser interpretados



**Figura 42** – Gráficos de resíduos Deviance obtidos para modelo de regressão bivariado através de cópula de Frank ajustado a dados de infecção

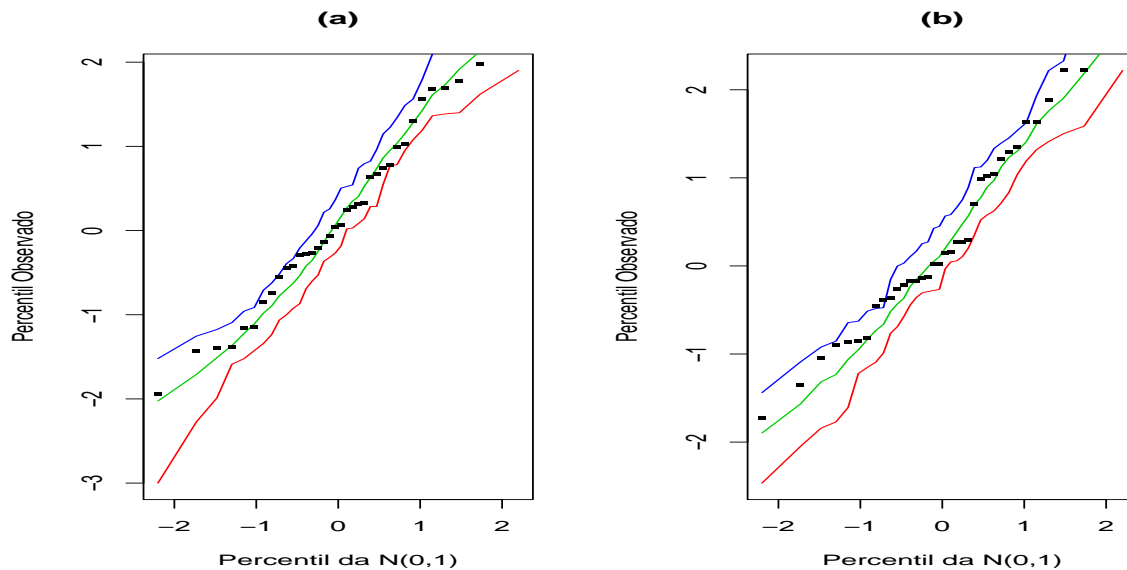
de forma que os coeficientes de regreção com sinal positivo retardam a ocorrência do evento a que se referem, enquanto os coeficientes negativos aceleram a mesma. O  $\tau = 0,22$  obtido através das estimativas da cópula detectou uma certa associação entre os eventos, diferente do que foi apontado na análise exploratória.

**Tabela 12** – Esimativas para o modelo final com a cópula de Frank para os dados de infecção

Parâmetro	Estimativa	Erro padrão	p-valor
$\beta_{01}$	4,0691	0,4427	0,0000
$\beta_{21}$	0,9355	0,3966	0,0183
$\beta_{31}$	0,0821	0,4434	0,8532
$\beta_{41}$	-0,7779	0,3802	0,0408
$\beta_{51}$	1,6746	0,6174	0,0067
$\sigma_1$	0,7862	0,1234	0,0000
$\beta_{02}$	3,4473	0,4100	0,0000
$\beta_{22}$	1,9570	0,3872	0,0000
$\beta_{32}$	-0,5162	0,4594	0,2612
$\beta_{42}$	-0,7156	0,3483	0,0399
$\beta_{52}$	0,7098	0,5494	0,1964
$\sigma_2$	0,7854	0,1173	0,0000
$\alpha$	2,1286	1,2328	0,0842

Uma análise gráfica da qualidade do ajuste dos modelos propostos para os eventos

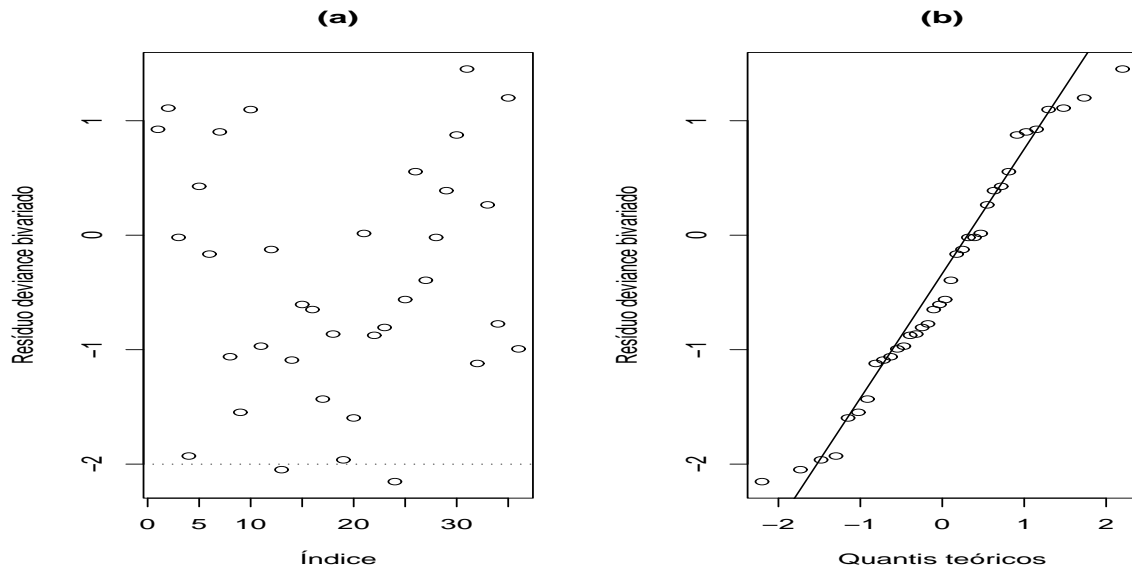
em estudo pode ser realizada na Figura 43 onde estão exibidos os gráficos de envelopes simulados para os resíduos deviance para os eventos 1 e 2, e também na Figura 44. Aparentemente o ajuste do modelo foi adequado pois os resíduos marginais tem comportamento normal e estão contidos nos envelopes simulados e os resíduos bivariados têm comportamento aleatório e aproximadamente normal.



**Figura 43** – Gráficos de envelopes simulados para os resíduos deviance calculados marginalmente para o tempo até as infecções 1 e 2.

### 4.3.3 Modelo de regressão bivariado através da cópula de Clayton

Nesta seção são apresentados os resultados da utilização da cópula de Clayton para reunir dois modelos como o descrito em (10). Os modelos são referentes aos logaritmos dos tempos até a ocorrência de dois eventos de infecção em pacientes com insuficiência renal. Como visto nos resultados da análise exploratória, é razoável supor que os logaritmos dos tempos até os eventos 1 e 2 têm distribuição valor extremo. Os resultados da estimação dos parâmetros de regressão e do parâmetro de associação da cópula de Clayton estão apresentados na Tabela 13. Nesta tabela aparecem os resultados das estimativas, erros padrões e significâncias dos parâmetros obtidos pelos métodos de máxima verossimilhança e jackknife descritos na Seção 3.4.



**Figura 44** – (a) Gráfico do resíduo bivariado e (b) Gráfico Normal de probabilidades do resíduo bivariado com cópula de Frank nos dados de infecção.

**Tabela 13** – Estimativas de máxima verossimilhança e Jackknife dos parâmetros do modelo de regressão bivariado através da cópula de Clayton para os dados de infecção

Parâmetro	Estimativa	EMV			Jackknife		
		Erro padrão	p-valor	Estimativa	Erro padrão	p-valor	
$\beta_{01}$	4,1882	0,6169	< 0,0000	4,4646	0,8275	0,0000	
$\beta_{11}$	0,0070	0,0146	0,6312	0,0117	0,0189	0,5356	
$\beta_{21}$	0,8047	0,3975	0,0430	0,3135	0,7663	0,6824	
$\beta_{31}$	-0,0981	0,5152	0,8489	0,2147	0,5839	0,7131	
$\beta_{41}$	-1,1724	0,5401	0,0300	-1,3843	0,6776	0,0411	
$\beta_{51}$	0,6388	0,6828	0,3495	0,2238	1,0800	0,8359	
$\sigma_1$	0,8806	0,1305	< 0,0000	1,0374	0,1425	0,0000	
$\beta_{02}$	2,9849	0,5268	< 0,0000	2,8892	0,6711	0,0000	
$\beta_{12}$	0,0112	0,0138	0,4198	0,0144	0,0147	0,3272	
$\beta_{22}$	2,2329	0,3665	0,0000	2,2756	0,4269	0,0000	
$\beta_{32}$	-0,0961	0,4964	0,8466	-0,0172	0,6351	0,9784	
$\beta_{42}$	-1,0593	0,5193	0,0414	-1,1773	0,4759	0,0134	
$\beta_{52}$	0,3594	0,6049	0,5524	0,5203	0,8446	0,5379	
$\sigma_2$	0,8431	0,1241	< 0,0000	0,9540	0,1436	0,0000	
$\alpha$	1,2714	0,6938	0,0669	0,8714	0,9546	0,3614	

#### 4.3.3.1 Análise de influência local

As Figuras 45, 46 e 47 apresentam gráficos das medidas de influência local contra os índices das observações para os esquemas de perturbação descritos em 3.4.3. As observações

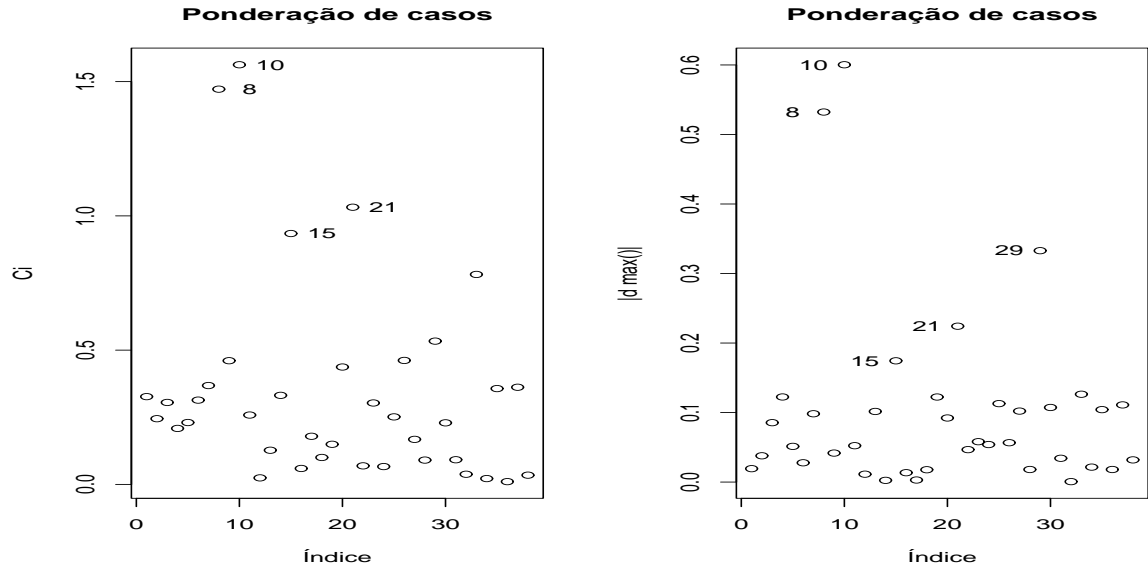


Figura 45 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Clayton segundo esquema de ponderação de casos para dados de infecção

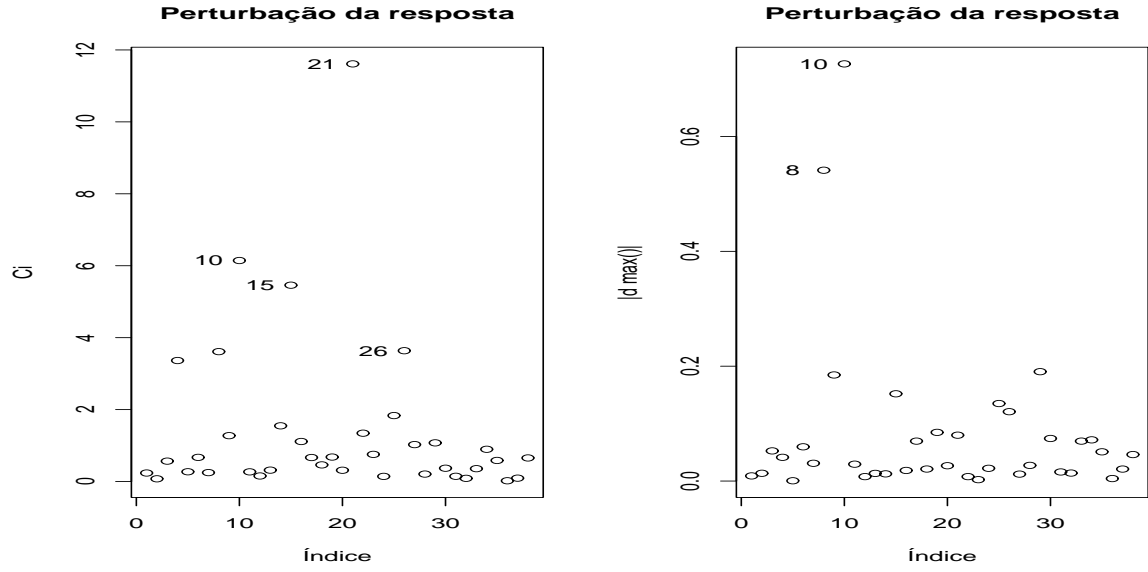
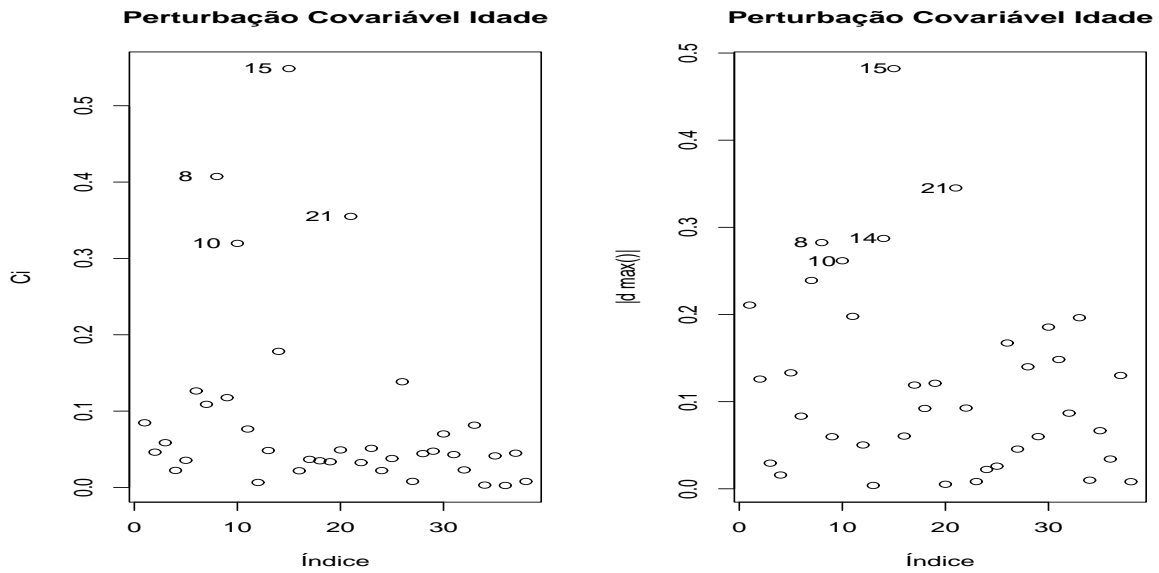


Figura 46 – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Clayton segundo esquema de perturbação da variável resposta para dados de infecção



**Figura 47** – Gráficos de medidas de influência para  $\Psi$  do modelo de regressão bivariado com cópula de Clayton segundo esquema de perturbação da covariável idade para dados de infecção

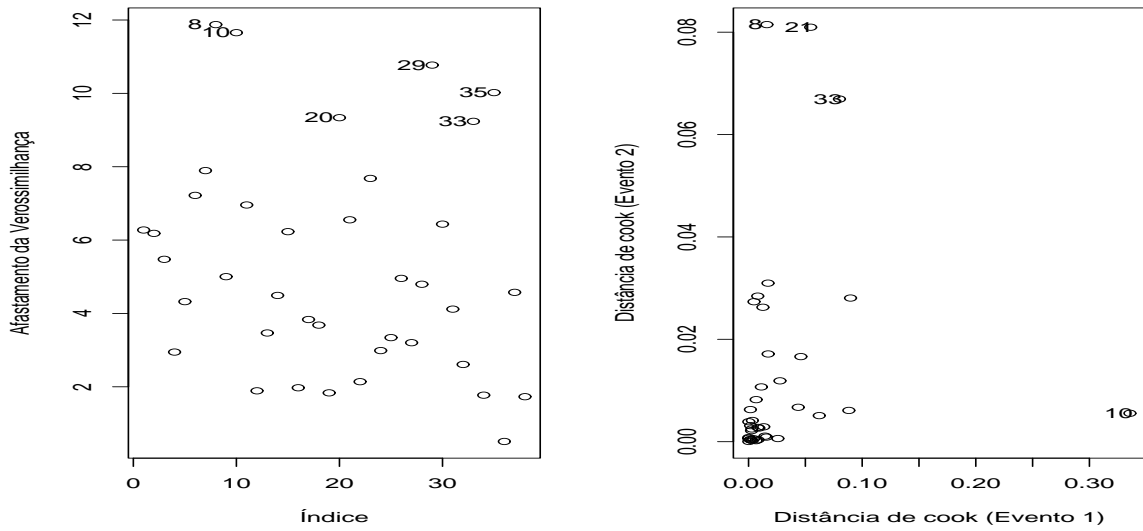
#8, #10, #15 e #21 destacam-se das demais e são possíveis pontos influentes.

#### 4.3.3.2 Análise de influência global

Foram calculadas a distância generalizada de Cook e o afastamento da verossimilhança como medidas de influência global. Na Figura 48 (a) é exibido o gráfico do afastamento da verossimilhança versus o índice das observações e em 48 (b) o gráfico de dispersão da distância de cook para os dois grupos de parametros referentes aos dois eventos. Nesta figura, destacam-se das demais as observações #8 #21 que são considerados possíveis pontos influentes.

#### 4.3.3.3 Análise de resíduos

Análise de resíduos apresentada em 3.4.5 também foi considerada para validar o ajuste do modelo. A Figura 49 exhibe os gráficos com os resíduos deviance para os eventos 1 e 2 e o resíduo deviance bivariado versus o índice das observações. O comportamento das medidas de resíduo é aparentemente aleatório mas um pouco assimétrico. O resíduo bivariado detecta pontos semelhantes aos detectados pelas outras técnicas de diagnóstico e, portanto, mostra-se como uma ferramenta útil neste contexto.



**Figura 48** – Gráficos de medidas de influência global para modelo de regressão bivariado através de cópula de Clayton ajustado a dados de infecção

#### 4.3.3.4 Impacto das observações influentes

Detectados alguns possíveis pontos influentes é de interesse detectar o impacto destas observações no modelo. Para isso foram estimados os parâmetros do modelo a partir de sub-amostras com a retirada dessas observações individualmente e em grupos. Estas estimativas estão apresentadas nas Tabelas 14 e 15 que contém as estimativas dos parâmetros, as mudanças relativas em relação às estimativas da amostra completa e os p-valores para cada sub-amostra.

#### 4.3.3.5 Modelo final proposto utilizando a cópula de Clayton

Mediante a eliminação das observações #8, #10 e #15 os modelos completos para os eventos 1 e 2, como o descrito em (10) foram ajustados ao conjunto de dados. Foram selecionadas para compor os modelos as variáveis consideradas significativas a um nível de 10%. Desta forma os modelos finais propostos para o logaritmo dos tempos até as infecções 1 e 2 nos pacientes com insuficiência renal podem ser expressos por:

$$y_{i1} = \beta_{01} + \beta_{21}x_{i2} + \beta_{31}x_{i3} + \beta_{41}x_{i4} + \beta_{51}x_{i5} + \sigma_1 z_{i1} \quad e$$

$$y_{i2} = \beta_{02} + \beta_{22}x_{i2} + \beta_{32}x_{i3} + \beta_{42}x_{i4} + \beta_{52}x_{i5} + \sigma_2 z_{i2}.$$

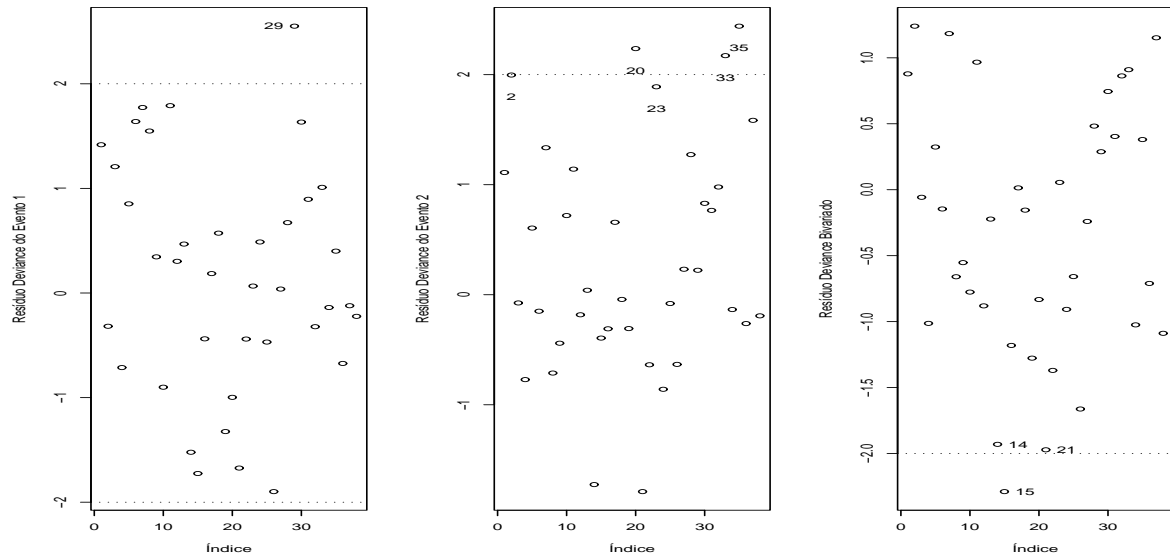


**Tabela 14** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até infecção 1

Sub-amostra	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{31}$	$\beta_{41}$	$\beta_{51}$
	4,188	0,010	0,701	-0,153	-1,211	0,520
A – (8)	[-0,055] (0,000)	[-45,070] (0,521)	[13,055] (0,078)	[-58,903] (0,785)	[-3,117] (0,032)	[18,340] (0,472)
	4,064	-0,006	1,430	-0,551	-0,675	1,412
A – (10)	[2,922] (0,000)	[191,549] (0,577)	[-77,501] (0,001)	[-470,497] (0,186)	[42,485] (0,153)	[-121,494] (0,029)
	3,702	0,016	0,746	0,019	-1,096	0,764
A – (15)	[11,561] (0,000)	[-122,535] (0,223)	[7,372] (0,035)	[120,186] (0,963)	[6,676] (0,019)	[-19,878] (0,215)
	4,034	0,007	0,968	-0,132	-1,126	0,139
A – (21)	[3,622] (0,000)	[7,042] (0,655)	[-20,117] (0,027)	[-36,542] (0,802)	[4,096] (0,037)	[78,208] (0,889)
	4,092	-0,002	1,247	-0,548	-0,758	1,220
A – (8, 10)	[2,243] (0,000)	[133,803] (0,849)	[-54,703] (0,004)	[-467,702] (0,220)	[35,417] (0,123)	[-91,481] (0,077)
	3,700	0,020	0,630	-0,068	-1,163	0,614
A – (8, 15)	[11,599] (0,000)	[-180,282] (0,126)	[21,755] (0,065)	[29,503] (0,873)	[0,996] (0,012)	[3,687] (0,317)
	4,033	0,010	0,870	-0,129	-1,172	0,166
A – (8, 21)	[3,641] (0,000)	[-36,620] (0,551)	[-7,905] (0,062)	[-33,747] (0,825)	[0,221] (0,039)	[73,925] (0,867)
	3,746	0,000	1,313	-0,387	-0,661	1,438
A – (10, 15)	[10,514] (0,000)	[94,366] (0,971)	[-62,968] (0,001)	[-301,035] (0,324)	[43,728] (0,138)	[-125,526] (0,020)
	3,868	-0,006	1,621	-0,566	-0,633	0,449
A – (10, 21)	[7,583] (0,000)	[190,141] (0,590)	[-101,104] (0,000)	[-486,335] (0,170)	[46,053] (0,171)	[29,495] (0,605)
	3,501	0,017	0,879	0,015	-1,046	0,105
A – (15, 21)	[16,355] (0,000)	[-136,620] (0,185)	[-9,034] (0,015)	[115,321] (0,972)	[10,960] (0,022)	[83,480] (0,905)
	3,745	0,006	1,097	-0,377	-0,772	1,214
A – (8, 10, 15)	[10,519] (0,000)	[14,085] (0,626)	[-36,163] (0,007)	[-290,166] (0,348)	[34,259] (0,090)	[-90,446] (0,060)
	3,886	-0,004	1,488	-0,526	-0,688	0,414
A – (8, 10, 21)	[7,155] (0,000)	[149,296] (0,787)	[-84,723] (0,001)	[-444,306] (0,242)	[41,437] (0,156)	[35,002] (0,634)
	3,497	0,021	0,772	-0,033	-1,121	0,070
A – (8, 15, 21)	[16,460] (0,000)	[-191,549] (0,108)	[4,170] (0,035)	[65,942] (0,939)	[4,547] (0,013)	[89,096] (0,935)
	3,518	0,002	1,470	-0,398	-0,625	0,392
A – (10, 15, 21)	[15,963] (0,000)	[73,239] (0,872)	[-82,465] (0,000)	[-312,215] (0,289)	[46,785] (0,140)	[38,500] (0,630)
	3,513	0,006	1,299	-0,357	-0,710	0,330
A – (8, 10, 15, 21)	[16,066] (0,000)	[11,268] (0,603)	[-61,256] (0,001)	[-269,358] (0,359)	[39,521] (0,102)	[48,211] (0,680)

**Tabela 15** – Estimativas, [mudanças relativas] e (p-valores) dos parâmetros de regressão para o tempo até infecção  
2

Sub-amostra	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{32}$	$\beta_{42}$	$\beta_{52}$	$\alpha$
A – (8)	3,348 [-12,052] (0,000)	0,006 [45,946] (0,641)	2,032 [8,993] (0,000)	-0,567 [-487,461] (0,201)	-0,935 [11,771] (0,063)	0,311 [12,324] (0,584)	1,549 [-21,196] (0,049)
A – (10)	2,926 [2,065] (0,000)	0,012 [-7,207] (0,410)	2,282 [-2,221] (0,000)	0,037 [138,342] (0,945)	-1,039 [1,981] (0,047)	0,322 [9,222] (0,595)	1,564 [-22,369] (0,031)
A – (15)	2,784 [6,834] (0,000)	0,015 [-33,333] (0,284)	2,184 [2,195] (0,000)	-0,103 [-6,632] (0,834)	-1,019 [3,905] (0,044)	0,371 [-4,540] (0,533)	1,143 [10,555] (0,056)
A – (21)	2,842 [4,893] (0,000)	0,013 [-15,315] (0,319)	2,266 [-1,500] (0,000)	-0,195 [-102,280] (0,666)	-1,048 [1,151] (0,034)	-0,271 [176,311] (0,702)	0,745 [41,687] (0,289)
A – (8,10)	3,306 [-10,656] (0,000)	0,007 [34,234] (0,579)	2,018 [9,607] (0,000)	-0,442 [-358,238] (0,347)	-0,901 [15,044] (0,068)	0,226 [36,266] (0,699)	1,739 [-36,069] (0,031)
A – (8,15)	3,143 [-5,174] (0,000)	0,010 [9,910] (0,399)	1,968 [11,882] (0,000)	-0,590 [-510,984] (0,152)	-0,904 [14,714] (0,050)	0,317 [10,603] (0,556)	1,420 [-11,095] (0,034)
A – (8,21)	3,221 [-7,815] (0,000)	0,007 [37,838] (0,570)	2,085 [6,637] (0,000)	-0,583 [-504,249] (0,159)	-0,934 [11,941] (0,051)	-0,255 [171,968] (0,702)	1,057 [17,276] (0,191)
A – (10,15)	2,791 [6,600] (0,000)	0,014 [-26,126] (0,334)	2,220 [0,573] (0,000)	0,059 [160,933] (0,912)	-0,975 [8,055] (0,058)	0,357 [-0,564] (0,551)	1,381 [-8,028] (0,032)
A – (10,21)	2,767 [7,390] (0,000)	0,014 [-28,829] (0,289)	2,316 [-3,731] (0,000)	-0,067 [30,155] (0,889)	-1,071 [-0,981] (0,031)	-0,294 [182,939] (0,677)	1,017 [20,460] (0,155)
A – (15,21)	2,663 [10,873] (0,000)	0,016 [-46,847] (0,217)	2,233 [0,004] (0,000)	-0,193 [-100,207] (0,669)	-1,028 [3,028] (0,033)	-0,285 [180,400] (0,682)	0,715 [44,042] (0,211)
A – (8,10,15)	3,168 [-6,021] (0,000)	0,010 [10,811] (0,427)	1,938 [13,203] (0,000)	-0,456 [-372,642] (0,310)	-0,846 [20,166] (0,069)	0,252 [28,990] (0,655)	1,532 [-19,826] (0,027)
A – (8,10,21)	3,146 [-5,288] (0,000)	0,008 [26,126] (0,523)	2,098 [6,033] (0,000)	-0,449 [-364,767] (0,314)	-0,909 [14,233] (0,057)	-0,267 [175,268] (0,691)	1,123 [12,143] (0,154)
A – (8,15,21)	3,022 [-1,124] (0,000)	0,011 [1,802] (0,349)	2,031 [9,016] (0,000)	-0,599 [-520,311] (0,130)	-0,911 [14,054] (0,041)	-0,272 [176,622] (0,667)	0,996 [22,041] (0,121)
A – (10,15,21)	2,635 [11,831] (0,000)	0,017 [-50,450] (0,217)	2,260 [-1,205] (0,000)	-0,064 [33,679] (0,895)	-1,022 [3,603] (0,035)	-0,291 [182,092] (0,677)	0,886 [30,655] (0,139)
A – (8,10,15,21)	3,005 [-0,576] (0,000)	0,011 [0,901] (0,372)	2,031 [9,051] (0,000)	-0,468 [-384,560] (0,278)	-0,869 [18,044] (0,055)	-0,273 [176,875] (0,674)	1,017 [20,444] (0,114)



**Figura 49** – Gráficos de resíduos Deviance obtidos para modelo de regressão bivariado através de cópula de Clayton ajustado a dados de infecção

As estimativas para os parâmetros destes modelos e para o parâmetro de associação  $\alpha$  da cópula de Frank estão apresentadas na Tabela 16. Os resultados podem ser interpretados de forma que os coeficientes de regreção com sinal positivo retardam a ocorrência do evento a que se referem, equanto os coeficientes negativos aceleram a mesma.

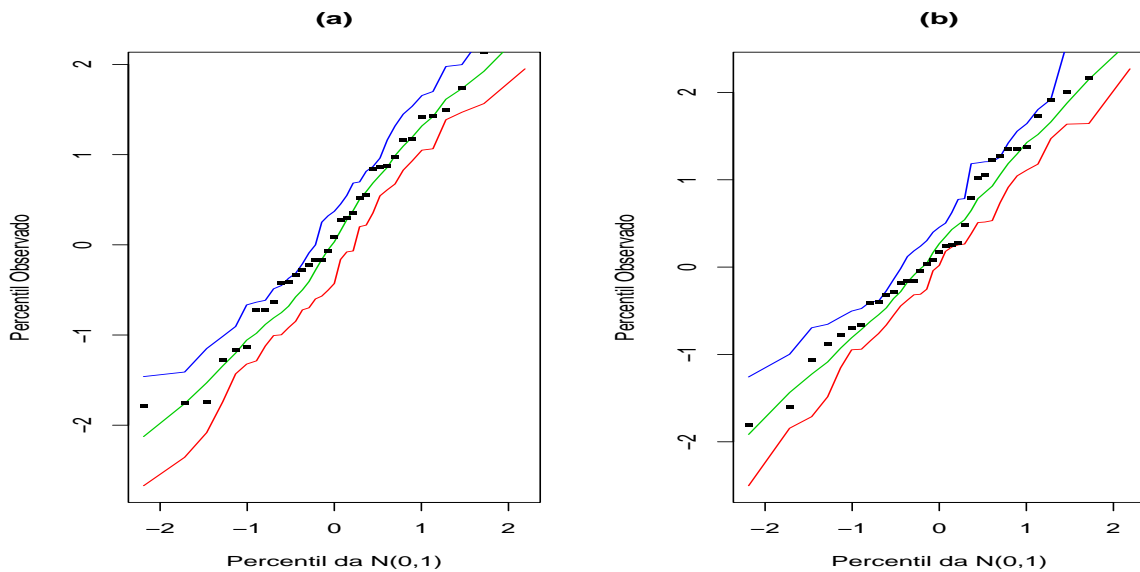
**Tabela 16** – Esimativas para o modelo final com a cópula de Frank para os dados de infecção

Parâmetro	Estimativa	Erro padrão	p-valor
$\beta_{01}$	3,8667	0,4193	0,0000
$\beta_{21}$	1,1527	0,3766	0,0022
$\beta_{31}$	-0,2647	0,3773	0,4829
$\beta_{41}$	-0,6144	0,3552	0,0837
$\beta_{51}$	1,4178	0,5300	0,0075
$\sigma_1$	0,7589	0,1166	0,0000
$\beta_{02}$	3,4800	0,4144	0,0000
$\beta_{22}$	1,8953	0,4010	0,0000
$\beta_{32}$	-0,2929	0,4325	0,4982
$\beta_{42}$	-0,5769	0,3258	0,0766
$\beta_{52}$	0,4173	0,5289	0,4301
$\sigma_2$	0,8070	0,1199	0,0000
$\alpha$	1,4128	0,6337	0,0258

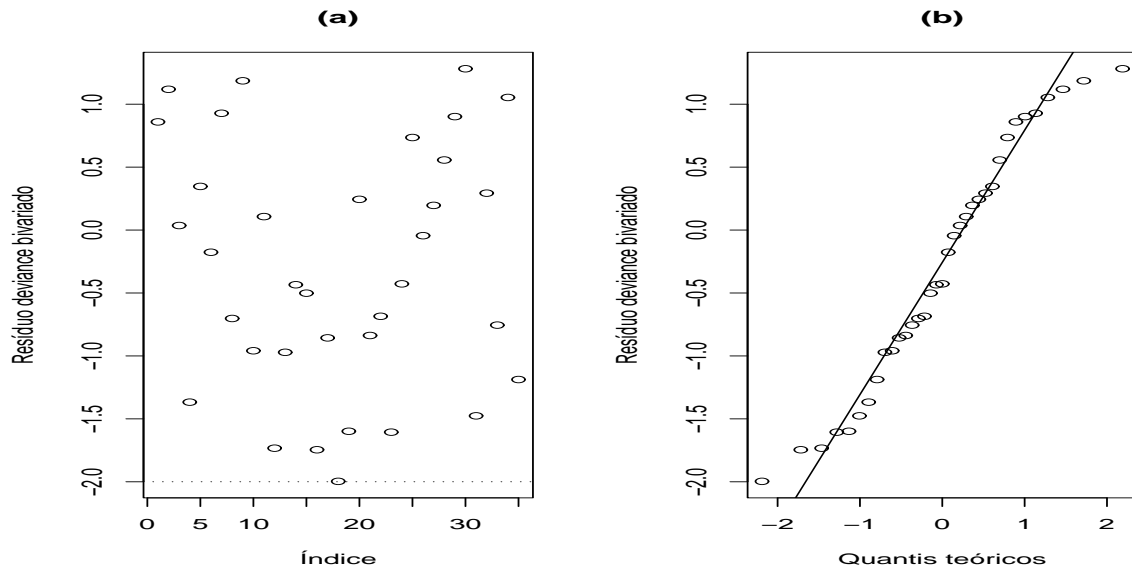
Uma análise gráfica da qualidade do ajuste dos modelos propostos para os eventos

em estudo pode ser realizada na Figura 50 onde estão exibidos os gráficos de envelopes simulados para os resíduos deviance para os eventos 1 e 2, e também na Figura 51 que exibe um gráfico Normal de probabilidades para o resíduo do ajuste bivariado e um gráfico de dispersão dos mesmos. Mediante a análise das figuras pode-se concluir que o modelo proposto se ajusta ao conjunto de dados de forma adequada pois os resíduos marginais estão contidos dentro dos envelopes simulados e o resíduo bivariado apresenta distribuição aproximadamente normal.

A associação estimada pelo modelo foi de  $\tau = 0,44$ , significativamente superior à obtida na análise exploratória. A forma de condução do experimento buscou uma independência entre os eventos, mas como os eventos distintos ocorrem no mesmo indivíduo, não se pode assegurar esta independência. De fato, para os modelos usando ambas as cópulas a hipótese de independência entre os eventos foi rejeitada.



**Figura 50** – Gráficos de envelopes simulados para os resíduos deviance calculados marginalmente para o tempo até as infecções 1 e 2.



**Figura 51** – (a) Gráfico do resíduo bivariado e (b) Gráfico Normal de probabilidades do resíduo bivariado com cópula de Clayton nos dados de infecção.

## 5 CONSIDERAÇÕES FINAIS

O objetivo inicial deste trabalho visava estudar um modelo de regressão bivariado com a presença de dados censurados por meio da teoria de cópulas. As cópulas de Frank e Clayton utilizadas neste trabalho apresentaram um bom desempenho para modelar dados de sobrevivência bivariados e apresentaram resultados semelhantes.

Foram apresentados os modelos de regressão com respostas bivariadas e as metodologias de estimação (máxima verossimilhança e jackknife), assim como técnicas de diagnóstico sob a abordagem de influência local, local total e global. Estas técnicas mostraram-se ferramentas importantes para avaliar a robustez dos modelos de regressão bivariados.

Foram utilizadas medidas de resíduos comumente usadas em análise de sobrevivência de forma marginal e também foi proposta uma medida de resíduos conjunta para modelos de regressão com respostas bivariadas, baseada nos resíduos Martingale e Deviance univariados. Para avaliar a distribuição empírica destes resíduos foi desenvolvido um estudo de simulação de Monte Carlo considerando diferentes tamanhos de amostra, porcentagens de censura e configurações paramétricas. Nas aplicações com os conjuntos de dados, a medida de resíduo proposta mostrou-se capaz de detectar pontos influentes condizentes aos detectados pelas técnicas de influência local.

Por meio das simulações foi observado que os resíduos propostos apresentam distribuição empírica aproximadamente Normal padrão. E como esperado na medida em que aumenta o percentual de censuras estas distribuições se afastam da normalidade.

Os resultados obtidos para a modelagem dos dados referente aos pacientes diabéticos revelam que este conjunto de dados demanda utilização de modelos marginais mais complexos do que os utilizados, pois apresenta uma fração de indivíduos curados que não foram considerados devido ao uso da suposição que ambas as variáveis resposta seguem uma distribuição Weibull. O elevado percentual de censuras na amostra também deve ser considerado neste conjunto de dados. As diferentes cópulas não apresentaram diferenças significativas entre si e também apresentaram resultados, para a estimação da associação entre as respostas, semelhantes ao obtido em por técnica não-paramétrica.

Na aplicação com os dados de pacientes com insuficiência renal os modelos pro-

postos mostraram-se adequados. As medidas utilizadas na análise de resíduos apresentaram comportamento semelhante ao esperado após o estudo de simulação realizado. Em análise exploratória dos dados, não foi detectada associação entre as as variáveis resposta, mas para ambas as cópulas utilizadas a associação foi considerada significativa no modelo de regressão.

## 5.1 Trabalhos futuros

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

1. Utilizar modelos marginais mais complexos incorporando a possibilidade de modelar dados que contenham uma fração de indivíduos curados.
2. Utilizar diferentes funções de cópulas
3. Estudar metodologia para seleção de função de cópula adequada
4. Estender os modelos de regressão para casos com mais de duas variáveis resposta.

## REFERÊNCIAS

- Achcar, J. A. (2005). Uma abordagem Bayesiana para dados multivariados de sobrevivência e binários na presença de covariáveis. *Revista de Estatística Matemática* 23, 79-94.
- ATKINSON, A. C. **Plots, Transformations, and Regression**. Oxford: University Press, 1985. 280p.
- CHATTERJEE, N.; SHIH, J. A bivariate cure-mixture approach for modeling familial association in diseases. **Biometrics**, Washington, v.57, p.779-786, 2001.
- COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blucher, 2006. 392p.
- COOK, R. D. Detection of influential observations in linear regression. **Technometrics**, Alexandria, v.19, p.15-18, 1977.
- COOK, R.D. Assessment of Local Influence. **Journal of the Royal Statistical Society**, London, v.48, p.133-169, 1986.
- COOK, R. D.; PEÑA, D.; WEISBERG, S. The likelihood displacement: A unifying principle for influence. **Communications in Statistics, Theory and Methods**, New York, v.17, p.623-640, 1988.
- COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. New York: Chapman & Hill, 1982. 280p.
- COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society B**, London, v.30, p.248-275, 1968.
- DOORNIK, J.A. **Ox: an Object-oriented matrix Programming Language**. Disponível em: <<http://www.nuff.ox.ac.uk/User/Doonik/>>. Acesso em: 4 jun. 2007.



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

GENEST, C. Frank's family of bivariate distributions. **Biometrika**, Cambridge, v.74, p.549-555, 1987.

GENEST, C.; RIVEST, L. P. Statistical inference procedures for bivariate archimedean copulas. **Journal of the American Statistical Association**, Alexandria, v.88, p.1034-1042, 1993.

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, p.63-78, 2005.

HUSTER, W.J.; BROOKMEYER, R.; SELF S.G. Modeling paired survival data with covariates. **Biometrics**, Washington, v.45, p.145-156, 1989.

KAPLAN, E.L.; MEIER, P., Nonparametric Estimation from Incomplete Observations. **Journal of the American Statistical Association**, Boston, v.53, p.457-481, 1958.

LAWLESS, J.F. **Statistical Models and Methods for Lifetime Data**. United States: John Wiley & Sons, 2003. 350p.

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

MILLER, R.G. The jackknife—A review. **Biometrika**, Cambridge, v.61, p.1-15, 1974.

Núñez, J. S. R. (2005). Modelagem Bayesiana para dados de sobrevivência bivariados através de cópulas. Tese de Doutorado, IME/USP.

OAKES, D.; RITZ, J. Regression in a bivariate copula model. **Biometrika**, Cambridge, v.87, p.345-352, 2000.

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

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

RIZATO, F.B. **Modelos de Regressão log-gama generalizado com fração de cura**. 2007. 74 p. Dissertação ( Mestrado em Estatística e Experimentação Agronômica )- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

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

Tibaldi, F. S. **Modeling of Correlated Data and Multivariate Survival Data**. 2004. 160 p. Teste de Doutorado, Universidade de Hasselt, Hasselt, 2004.

WANG, W; WELLS, M.T. Model selection and semiparametric inference for bivariate failure-time data. **Journal of the American Statistical Association**, Alexandria, v.95, p.62,72, 2000.

WANG, W; WELLS, M.T. Nonparametric estimators of the bivariate survival function under simplified censoring conditions. **Biometrika**, Cambridge, v.84, p.863-880, 1997.

ZHU, H.; ZHANG, H. A diagnostic procedure based on local influence. **Biometrika**, Cambridge, v.91, p.597-589, 2004.

**ANEXOS**

Exemplo de programa em R utilizado para simulações com cópula de Frank.

```

require(copula)
require(MASS)
require(survival)
set.seed(577)
banco<-rpois(27,98577234)
cont<-1

expr <- quote({

  verosimul<-function(param){
    a<-param[1]
    b01<-param[2]
    b02<-param[3]
    b11<-param[4]
    b12<-param[5]
    d1<-param[6]
    d2<-param[7]

    s12<-(-(1/a))*log(1+((exp((-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1));

    s2<-(-exp((y1-(b01*xx0+b11*xx1))/d1)*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))*
    exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/
    (d1*(exp(-a)-1)*(1+(exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1)))

    s1<-(-exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    exp((y2-(b02*xx0+b12*xx1))/d2)*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))*
    exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2)))/(d2*(exp(-a)-1)*
    (1+(exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1)))

    f12<-(-exp((y1-(b01*xx0+b11*xx1))/d1)*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))*
    exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))*a*exp((y2-(b02*xx0+b12*xx1))/d2)*
    exp(-exp((y2-(b02*xx0+b12*xx1))/d2))*exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2)))/
    (d1*d2*(exp(-a)-1)*(1+(exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1)))
    +
    exp((y1-(b01*xx0+b11*xx1))/d1)*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))*
    exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)*
    (exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    a*exp((y2-(b02*xx0+b12*xx1))/d2)*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))
    *exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2)))/(d1*(exp(-a)-1)^2*
    (1+(exp(-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1))))-1)*
    (exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1)^2*d2))

    verosim<-(-1)*(((c1*c2)*log(f12))+(c1*(1-c2)*log(-s2))+((1-c1)*c2*log(-s1))+((1-c1)*(1-c2)*log(s12)))
    sum(verosim)
    #browser()
  }

  Cens<-matrix(0,ssize,2*RepNo)
  Y<-matrix(0,ssize,2*RepNo)
  estima<-matrix(0,7,RepNo)
  conver<-rep(33,RepNo)
  xx0<-rep(1,ssize)
  xx1<-runif(ssize)
  set.seed(banco[cont])
  cont<-cont+1
  conver<-rep(33,RepNo)
  frk<-frankCopula(Copula)
  distr<-mvdc(frk, c("weibull", "weibull"),
    list(list(shape= 3.5, scale = 6), list(shape=4, scale = 20)))
  for (i in 1:RepNo-1){
    tempo<-rmvdc(distr,ssize)

```

```

suras1<-runif(ssize,1,CProb*max(tempo[,1]))
suras2<-runif(ssize,1,CProb*max(tempo[,2]))
c1<-ifelse(tempo[,1]>suras1,0,1)
c2<-ifelse(tempo[,2]>suras2,0,1)
Cens[, (2*i)+1]<-c1
Cens[, (2*i)+2]<-c2
y1<-ifelse(tempo[,1]>suras1,log(suras1),log(tempo[,1]))
y2<-ifelse(tempo[,2]>suras2,log(suras2),log(tempo[,2]))
ajust1<-survreg(Surv(y1,c1)~xx1,dist='extreme')
ajust2<-survreg(Surv(y2,c2)~xx1,dist='extreme')
if(i<2){feit <- c(CProb+1, ajust1$coefficients[1],
ajust2$coefficients[1],ajust1$coefficients[2],
ajust2$coefficients[2], ajust1$scale,
ajust2$scale)}
else {feit<-apply(estima[,1:i],1,mean)}
relatorio<-optim(feit,verosimul,method="BFGS")
estima[,i+1]<-relatorio$par
conver[i+1]<-relatorio$convergence
Y[, (2*i)+1]<-y1
Y[, (2*i)+2]<-y2
}
conver
pct30<-1-(mean(apply(Cens,2,sum))/30)
apply(estima,1,mean)

passa<-numeric(RepNo)
dev1<-matrix(0,ssize,RepNo)
dev2<-matrix(0,ssize,RepNo)
dev<-matrix(0,ssize,RepNo)
martin3<-matrix(0,ssize,RepNo)
martin1<-matrix(0,ssize,RepNo)
martin2<-matrix(0,ssize,RepNo)
schap1<-matrix(0,ssize,RepNo)
schap2<-matrix(0,ssize,RepNo)
schap12<-matrix(0,ssize,RepNo)
for(i in 1:RepNo-1){
a<-estima[1,i+1]
b01<-estima[2,i+1]
b02<-estima[3,i+1]
b11<-estima[4,i+1]
b12<-estima[5,i+1]
d1<-estima[6,i+1]
d2<-estima[7,i+1]
y1<-Y[, (2*i+1)]
y2<-Y[, (2*i+2)]

schap12[,i+1]<-(-(1/a)*log(1+((exp((-a*exp(-exp((y1-(b01*xx0+b11*xx1))/d1)))))-1)*
(exp(-a*exp(-exp((y2-(b02*xx0+b12*xx1))/d2))))-1)/(exp(-a)-1));
schap1[,i+1]<-exp(-exp((y1-(b01*xx0+b11*xx1))/d1))
schap2[,i+1]<-exp(-exp((y2-(b02*xx0+b12*xx1))/d2))
martin1[,i+1]<- Cens[, (2*i+1)]+log(schap1[,i+1])
martin2[,i+1]<- Cens[, (2*i+2)]+log(schap2[,i+1])
dev1[,i+1]<-sign(martin1[,i+1])*(-2*(martin1[,i+1]+Cens[, (2*i+1)]*
log(Cens[, (2*i+1)]-martin1[,i+1])))^(1/2)

dev2[,i+1]<-sign(martin2[,i+1])*(-2*(martin2[,i+1]+Cens[, (2*i+2)]*
log(Cens[, (2*i+2)]-martin2[,i+1])))^(1/2)

martin3[, i + 1] <- apply(Cens[, (2 * i + 1):(2 * i + 2)],1,sum)/2 + log(schap12[, i + 1])
dev[, i + 1] <- sign(martin3[, i + 1]) *
(-2 * (martin3[,i + 1] + apply(Cens[, (2 * i + 1):(2 * i + 2)],1, sum)/2 *
log(apply(Cens[, (2 * i + 1):(2 * i + 2)],1, sum)/2 - martin3[, i + 1])))^(1/2)
if (ks.test(dev1[,i+1],"pnorm")$p.value>0.05) passa[i]<-1
}
sum(passa)

resid1<-apply(apply(dev1,2,sort),1,mean)
resid2<-apply(apply(dev2,2,sort),1,mean)
resid3<-apply(apply(dev,2,sort),1,mean)
postscript(file = paste("residuo", Copula,CProb, "%03d.ps", sep="_"))
par(mfrow=c(3,3))

```

```

qqnorm(resid1,xlab="Quantis teóricos",ylab="Quantis observados n=ssize",
main="Resíduo deviance evento 1 ")
qqline(resid1)
qqnorm(resid2,xlab="Quantis teóricos",ylab="Quantis observados n=ssize",
main="Resíduo deviance evento 2 ")
qqline(resid2)
qqnorm(resid3,xlab="Quantis teóricos",ylab="Quantis observados n=ssize",
main="Resíduo deviance bivariado")
qqline(resid3)
dev.off()
})

subst <- function(Command, ...)
  do.call("substitute", list(Command, list(...)))

sink("Copula_Batch_file.R")
for(ssize in c(30,50,100))
  for(CProb in c(5,3,2))
    for(Copula in c(-2,0.3,6))

      {
        thisExpr <-
          subst(expr,ssize=ssize,CProb = CProb, RepNo = 100, Copula = Copula)
        cat("\n#####\n### ", "Copula =", Copula, "Censura=", CProb,
            "\n#####\n")
        for(j in 2:length(thisExpr)) print(thisExpr[[j]])
      }

sink()

```

Exemplo de programa em Ox utilizado para realizar estimativas de máxima verossimilhança e Jackknife nos modelos com cópula de Clayton.

```

#include<oxstd.h>
#include<oxdraw.h>
#include<oxfloat.h>
#include<maximize.h>
#include<simula.h>
#pragma link("maximize.oxo")
static decl g_mY;
static decl g_mX;
log_vero(const vP,const adFunc, const avScore,const amHessian)
{
  decl n=rows(g_mY);
  decl cont,t1, t2, c1, c2, xx0, xx1, xx2, xx3, xx4,xx5;
  decl uns=ones(n,1);
  decl vero=zeros(1,n);
  decl a1=fabs(vP[0][0]); //alpha
  decl d1=fabs(vP[1][0]); // sigma 1
  decl d2=fabs(vP[2][0]); // sigma 2
  decl b01=vP[3][0]; //beta 01
  decl b11=vP[4][0]; //beta 11
  decl b21=vP[5][0]; //beta 21
  decl b31=vP[6][0]; //beta 31
  decl b41=vP[7][0]; //beta 41
  decl b51=vP[8][0]; //beta 51
  decl b02=vP[9][0]; //beta 02
  decl b12=vP[10][0]; //beta 12
  decl b22=vP[11][0]; //beta 22
  decl b32=vP[12][0]; //beta 32
  decl b42=vP[13][0]; //beta 42
  decl b52=vP[14][0]; //beta 52

  for(cont=0;cont<n;++cont)
  {
    t1=g_mY[cont][0];
    t2=g_mY[cont][1];
    c1=g_mY[cont][2];

```

```

c2=g_mY[cont][3];
xx0=g_mY[cont][4];
xx1=g_mY[cont][5];
xx2=g_mY[cont][6];
xx3=g_mY[cont][7];
xx4=g_mY[cont][8];
xx5=g_mY[cont][9];

decl s12=((exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)+
(exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)-1)^(-1/a1));

decl s2=(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)*
exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)))*
exp(-a1*exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)))*
(exp(-a1*exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)))-1)/
(d1*(exp(-a1)-1)*(1+(exp(-a1*exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)))-1)*
(exp(-a1*exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)))-1)/(exp(-a1)-1))));

decl s1=(-(exp(-a1*exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)))-1)*
exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)*
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)))*
exp(-a1*exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)))/
(d2*(exp(-a1)-1)*(1+(exp(-a1*exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)))-1)*
(exp(-a1*exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)))-1)/(exp(-a1)-1))));

decl f12=(exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)+
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)-1)^(-1/a1)*
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)*exp((log(t2)-
(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)*exp(-exp((log(t1)-
(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)*
exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)/
(d2*(exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)+
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)-1)^2*d1)+
(exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)+
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)-1)^(-1/a1)*
exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)*
exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1)*
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)*a1*
exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2)/
(d1*(exp(-exp((log(t1)-(b01*xx0+b11*xx1+b21*xx2+b31*xx3+b41*xx4+b51*xx5))/d1))^-a1)+
exp(-exp((log(t2)-(b02*xx0+b12*xx1+b22*xx2+b32*xx3+b42*xx4+b52*xx5))/d2))^-a1)-1)^2*d2);

vero[0][cont]=((c1*c2)*log(f12))+(c1*(1-c2)*log(-s2))
+((1-c1)*c2*log(-s1))+((1-c1)*(1-c2)*log(s12))

;
}
adFunc[0]=double(vero*uns);
return 1;
}
main()
{
g_mY = loadmat("a:klein_bivariado.txt");
decl nc=rows(g_mY);
decl dfunc;
decl vP=< 0.2;0.87;0.84;4.20;-0.01885;1.02;-0.19;-0.95;1.60;
3.09;0.002856;2.14;-0.19;-0.715;0.8526 >;
log_vero(vP,&dfunc,0,0);
MaxControl(-1,20);
decl mhess;
decl ir, var, var1,var11,var12,desv;
ir=MaxBFGS(log_vero,&vP,&dfunc,0,1);
Num2Derivative(log_vero,vP,&mhess);
var1=invertsym((-1)*mhess);
print(" os valores das variâncias 1 \n",var1);
desv=diagonal(var1);
print("desvios",desv);
decl logorig=dfunc;
decl parfull=vP;
decl parfull1=zeros(7,1);
decl parfull2=zeros(7,1);

```

```

parfull1[0][0]=vP[1][0];
parfull2[0][0]=vP[2][0];
parfull1[1:6][0]=vP[3:8][0];
parfull2[1:6][0]=vP[9:14][0];
print(" os valores dos parametros \n", vP);
decl neu=rows(vP-1);
decl test2=zeros(neu,1);
decl cont1;
for(cont1=0;cont1<neu;++cont1)
{
test2[cont1][0]=vP[cont1][0]/sqrt(var1[cont1][cont1]);
}
decl ep= sqrt(diagonal(-1/mhess));
print("\n Estimativa,Erro padrão p_valor:", vP~ep'~(2*(1-probn(fabs(vP'./ep)))'));

print("t calculado",test2);
decl hessi1=zeros(7,7);
decl hessi2=zeros(7,7);
hessi1[0][0]=mhess[1][1];
hessi1[0][1:6]=mhess[1][3:8];
hessi1[1:6][0]=mhess[3:8][1];
hessi1[1:6][1:6]=mhess[3:8][3:8];
hessi2[0][0]=mhess[2][2];
hessi2[0][1:6]=mhess[2][9:14];
hessi2[1:6][0]=mhess[9:14][2];
hessi2[1:6][1:6]=mhess[9:14][9:14];
var11=invertsym((-1)*hessi1);
var12=invertsym((-1)*hessi2);

decl matpseud=zeros(15,nc);
decl veros=zeros(nc,1);
decl cook=zeros(nc,1);
decl matpar=zeros(15,nc);
decl varjack=zeros(15,nc);
decl cook1=zeros(nc,1);
decl cook2=zeros(nc,1);
decl i;
decl nmat=g_mY;
/// Técnicas de Jackknife e Deleção de casos
for (i=0; i<nc; ++i)
{
g_mY=nmat;
g_mY=dropr(g_mY,i);
decl nc=rows(g_mY);
decl dfunc;

decl vP=< 0.2;0.87;0.84;4.20;-0.01885;1.02;-0.19;-0.95;1.60;
3.09;0.002856;2.14;-0.19;-0.715;0.8526 >;
log_vero(vP,&dfunc,0,0);
println("vero=" ,dfunc);

MaxControl(-1,20);
decl ir, vari ;
decl vP1=zeros(7,1);
decl vP2=zeros(7,1);
ir=MaxBFGS(log_vero,&vP,&dfunc,0,1);
Num2Derivative(log_vero,vP,&vari);
varjack[i]=(diagonal(invertsym((-1)*vari)))';
vP1[0][0]=vP[1][0];
vP2[0][0]=vP[2][0];
vP1[1:6][0]=vP[3:8][0];
vP2[1:6][0]=vP[9:14][0];
veros[i][0]=dfunc;
matpar[i]=vP[i][0];
matpseud[i]=(nc*parfull)-((nc-1)*vP);
cook[i][0]=(vP-parfull)'*(var1)*(vP-parfull);
cook1[i][0]=(vP1-parfull1)'*(var11)*(vP1-parfull1);
cook2[i][0]=(vP2-parfull2)'*(var12)*(vP2-parfull2);
}
decl jack=zeros(15,1);
decl jackvar=zeros(15,1);

```



```
jack=meanr(matpseud);
jackvar=sqrt(sumr((matpseud-jack).^2)/(nc*(nc-1)));

decl LD=2*(logorig-veros);
print("cook1",cook1);
print("cook2",cook2);
print("LD",LD);
print("olha o jack ai gente",jack);
print("variânciasjack",jackvar);
}
```

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)