



Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

ÉRICA VIEIRA NOGUEIRA

INFLUÊNCIA LOCAL EM MODELOS ELÍPTICOS MULTINÍVEIS

Recife

2019

ÉRICA VIEIRA NOGUEIRA

INFLUÊNCIA LOCAL EM MODELOS ELÍPTICOS MULTINÍVEIS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

Área de Concentração: Ciências Exatas e da Terra

Orientadora: Profa. Dra. Audrey Helen Mariz de Aquino Cysneiros

Coorientador: Prof. Dr. Roberto Ferreira Manghi

Recife

2019

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

N778i Nogueira, Érica Vieira
Influência local em modelos elípticos multiníveis / Érica Vieira Nogueira. –
2019.
75 f.: il., fig., tab.

Orientadora: Audrey Helen Mariz de Aquino Cysneiros.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN,
Estatística, Recife, 2019.
Inclui referências e apêndices.

1. Estatística. 2. Influência local. 3. Modelos elípticos. I. Cysneiros, Audrey
Helen Mariz de Aquino (orientador). II. Título.

310

CDD (23. ed.)

UFPE- MEI 2019-131

ÉRICA VIEIRA NOGUEIRA

INFLUÊNCIA LOCAL DE MODELOS ELÍPTICOS MULTINÍVEIS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 31 de julho de 2019.

BANCA EXAMINADORA

Prof.(º) Roberto Ferreira Manghi
UFPE

Prof.(º) Francisco José de Azevedo Cysneiros
UFPE

Prof.(º) Juvêncio Santos Nobre
UFC

À minha Mãe, o bem mais precioso que Deus
poderia me dar. Exemplo de garra e dedicação.

AGRADECIMENTOS

Primeiramente a Deus por ter me dado saúde e força para superar as dificuldades.

À minha mãe **Marina** por acreditar em meus objetivos.

À minha irmã **Fernanda**, aos meus tios e tias pelo carinho e palavras de ânimo.

À minha amiga **Nayara** e a meu amigo **Mizael** por estarem ao meu lado me dando suporte emocional nos momentos em que eu pensava que não conseguiria chegar até o fim deste trabalho.

À minha orientadora **Audrey Helen Mariz de Aquino Cysneiros** e ao meu coorientador **Roberto Ferreira Manghi** pela oportunidade de trabalhar com pessoas de tanta qualidade, inteligência e por todo o suporte que eu precisei.

À todos os professores do **Departamento de Estatística do CCEN** nos quais tive a oportunidade de conhecer e aprender um pouco mais.

Aos colegas da Pós-Graduação.

Pelo apoio financeiro da **CAPES**.

A linguagem secreta da Estatística, tão atraente em uma cultura voltada para os fatos, é empregada para apelar, inflar, confundir e levar a simplificações exageradas. Métodos e termos estatísticos são necessários para relatar dados de tendências sociais e econômicas, condições dos negócios, pesquisas de opinião e censos. No entanto, sem redatores que usem as palavras com honestidade e conhecimento, e sem leitores que saibam o que significam, o resultado só pode ser um absurdo semântico.

(HUFF, 2016, p.16)

RESUMO

A classe dos modelos elípticos multiníveis tem sido amplamente utilizada na modelagem de dados que apresentam estrutura de hierarquia. Esta classe propõe o uso de distribuições de probabilidade pertencentes à classe elíptica, que envolve todas as distribuições simétricas contínuas, sendo a normal um caso particular. O principal objetivo da presente dissertação é o desenvolvimento de técnicas de diagnóstico, mais especificamente a análise de influência local via curvatura normal para os modelos elípticos multiníveis. Para tanto, quatro esquemas de perturbação são considerados: perturbação de casos, perturbação na matriz de variância e covariância, perturbação na resposta e perturbação na variável explicativa contínua. As técnicas desenvolvidas foram utilizadas segundo ajustes para dois conjuntos de dados reais, especificamente sob os modelos normal, t-Student e exponencial potência e os resultados obtidos foram discutidos. O modelo sob a distribuição t-Student se adequou melhor ao exemplo de dois níveis e a distribuição exponencial potência, com o exemplo de três níveis, conseguiu adequar-se de forma satisfatória ao conjunto de dados analisado. Podemos ver que os modelos elípticos multiníveis se mostram muito eficientes para a modelagem de dados com estrutura de hierarquia e com observações discrepantes.

Palavras-chave: Influência local. Modelos elípticos. Modelos multiníveis.

ABSTRACT

The class of elliptical multilevel models has been widely used in data sets, which present some hierarchical structure. This class proposes the use of probability distributions in the elliptical class, which involves all symmetric continuous distributions, including the normal distribution as a particular case. The main objective of this dissertation is the development of diagnostic techniques, more specifically, the analysis of local influence via normal curvature in the elliptical multilevel models. Therefore, four perturbation schemes are considered: case-weight perturbation, scale matrix perturbation, response variable and continuous explanatory variable perturbation. The techniques developed were used according to fit two real datasets, specifically under models normal, Student-t and exponential power models. The model under the t-Student distribution was better suited to the two-level example and the exponential power distribution, with the three-level example, was able to satisfactorily fit the data set analyzed. We can see that multilevel elliptical models are very efficient for hierarchical structure data modeling with discrepant observations.

Keywords: Elliptical models. Local influence. Multilevel models.

LISTA DE GRÁFICOS

Gráfico 1	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.	42
Gráfico 2	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.	43
Gráfico 3	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.	43
Gráfico 4	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição t-Student com o banco de dados dos filhotes de ratos	45
Gráfico 5	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição t-Student aplicado ao banco de dados dos filhotes de ratos.	46
Gráfico 6	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição t-Student aplicado ao banco de dados dos filhotes de ratos.	46
Gráfico 7	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos	48
Gráfico 8	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos.	49
Gráfico 9	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos.	49
Gráfico 10	– Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição normal aplicado ao banco de dados do produto interno bruto.	54

Gráfico 11 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição normal aplicado ao banco de dados do produto interno bruto.	54
Gráfico 12 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 $Lwater$, para a distribuição normal aplicado ao banco de dados do produto interno bruto.	55
Gráfico 13 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição normal aplicado ao banco de dados do produto interno bruto.	55
Gráfico 14 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.	56
Gráfico 15 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.	57
Gráfico 16 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 $Lwater$, para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.	57
Gráfico 17 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.	58
Gráfico 18 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.	60
Gráfico 19 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.	60

Gráfico 20 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 $Lwater$, para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.	61
Gráfico 21 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.	61

LISTA DE TABELAS

Tabela 1	– Descrição das variáveis pertencentes aos experimentos dos ratos	39
Tabela 2	– Estimativas dos parâmetros com seus erros-padrão aproximados para a distribuição normal, t-Student e exponencial potência segundo o modelo multinível ajustado ao banco de dados dos filhotes de ratos.	41
Tabela 3	– Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhadas 6, 18, 6 e 18 juntas e sem as observações 66 e 227 juntas) para a distribuição normal, segundo os modelos ajustados ao banco de dados dos filhotes de ratos.	44
Tabela 4	– Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhada 6 e sem as observações 109 e 227), para a distribuição t-Student, segundo os modelos ajustados ao banco de dados dos filhotes de ratos.	47
Tabela 5	– Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhada 6 e sem as observações 66 e 275 juntas) para a distribuição exponencial potência, segundo os modelos ajustados ao banco de dados dos filhotes de ratos	48
Tabela 6	– Medidas descritivas para para a variável resposta $Lgsp$ relativa aos Dados do Produto Interno Bruto	52
Tabela 7	– Medidas descritivas para para a variável preditora $Lwater$ relativa aos Dados do Produto Interno Bruto	52
Tabela 8	– Estimativas dos parâmetros com seus erros-padrão aproximados para a distribuição normal, t-Student e exponencial potência, segundo os modelos ajustados ao banco de dados do produto interno bruto	53
Tabela 9	– Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto (sem a Região 8) para as distribuições normal e t-Student, segundo os modelos ajustados ao banco dados do produto interno bruto.	62
Tabela 10	– Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto (sem a observação 17 do estado 4 pertencente à Região 1) para as distribuições normal, t-Student e exponencial potência, segundo os modelos ajustados ao banco de dados do produto interno bruto.	63

Tabela 11 – Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto para a distribuições t-Student (sem a observação 1 do estado 1 pertencente à Região 8) e para a distribuição exponencial potência (sem a observação 17 do estado 2 pertencente à Região 6), segundo os modelos ajustados ao banco de dados do produto interno bruto. 64

SUMÁRIO

1	INTRODUÇÃO	15
2	MODELOS ELÍPTICOS MULTINÍVEIS	18
2.1	CLASSE DAS DISTRIBUIÇÕES ELÍPTICAS	18
2.2	MODELOS MULTINÍVEIS	19
2.3	ESTIMAÇÃO DOS PARÂMETROS E MATRIZ DE INFORMAÇÃO	24
3	INFLUÊNCIA LOCAL	28
3.1	CURVATURA NORMAL	29
3.2	ESQUEMAS DE PERTURBAÇÃO	30
3.2.1	Perturbação de caso	31
3.2.2	Perturbação na matriz de variância e covariância	32
3.2.3	Perturbação na variável resposta	34
3.2.4	Perturbação na variável preditora: Nível 1	35
4	EXEMPLOS DE APLICAÇÕES	38
4.1	FILHOTES DE RATOS	38
4.2	DADOS DO PRODUTO INTERNO BRUTO	50
5	CONSIDERAÇÕES FINAIS	65
	REFERÊNCIAS	66
	APÊNDICE A – INFORMAÇÃO DE FISHER	70
	APÊNDICE B – PERTURBAÇÃO NA MATRIZ V_k	71
	APÊNDICE C – PERTURBAÇÃO NA VARIÁVEL RESPOSTA	73

1 INTRODUÇÃO

Geralmente em dados coletados nas áreas de ciências humanas e biológicas verifica-se que há uma estrutura de hierarquia, ou seja, há um arranjo natural das unidades em níveis, geralmente estas estruturas são caracterizadas por serem organizadas em grupos. Filhotes de uma mesma ninhada tendem a ser semelhantes em relação aos aspectos físicos do que se comparados com indivíduos aleatórios de uma dada população; a herança de aspectos familiares por causa da descendência do grupo, são alguns dos exemplos em que há uma estrutura hierárquica (ver Goldstein, 2011), por exemplo.

Desta maneira, pode-se ver aplicações de bancos de dados que apresentam dois ou três níveis de hierarquia. Pinheiro (2005) apresentou um exemplo que utiliza um modelo multinível com dois níveis no qual avaliou o desempenho de alunos da 4ª e 8ª série do ensino fundamental e 3ª série do ensino médio, em Matemática e Português. Pires (2009) usa um modelo de dois níveis em que foram estudadas as habilidades em Língua Portuguesa e em Matemática dos alunos da terceira série do ensino médio de escolas públicas do estado de Pernambuco, segundo características dos alunos e das escolas que eles pertenciam. Soares (2005) considerou três níveis de hierarquia (alunos, turmas e colégios), no qual o objetivo era medir a proficiência dos alunos de quarta série em Língua Portuguesa, e avaliar a variabilidade da habilidade dos alunos devido a diferença entre eles, à turma e às escolas.

Embora a suposição de normalidade para variáveis seja adequada em diversas situações, ela pode não se mostrar adequada para dados com observações discrepantes. Desta forma, a classe elíptica surge como uma alternativa atraente devido a muitas variações possíveis de curtose e os aspectos robustos das estimativas de máxima verossimilhança para observações extremas, (ver Fang, Kotz e Ng, 1990).

É conhecido que modelos elípticos baseados em distribuições com caudas mais pesadas que a normal apresentam estimativas de parâmetros menos sensíveis na presença de observações extremas que podem ter uma influência desproporcional na estimativa dos parâmetros. Sabe-se que a classe das distribuições elípticas abrange todas as distribuições contínuas simétricas. A título de exemplo tem-se a Normal, t-Student, Exponencial Potência, Logística I, Logística II, entre outras, nas quais suas propriedades podem ser vistas em Fang, Kotz e Ng (1990). Por exemplo, a distribuição t-Student mostra-se mais flexível que a distribuição normal e pode lidar com várias situações envolvendo caudas pesadas. Neste sentido, podemos considerar os trabalhos de Galea, Paula e Bolfarine (1997), Díaz-García, Rojas e Leiva-Sánchez (2003), Cysneiros e

Vanegas (2008), Leiva *et al.* (2016), citando caso parecido.

Em modelos de regressão para dados correlacionados, os modelos lineares baseados na distribuição normal desempenham um papel essencial, no entanto, há conclusões em que a suposição de normalidade não é adequada em certas situações, desta forma pode-se considerar alternativas sob essas circunstâncias e uma escolha é a família de distribuição elíptica. Desta maneira, modelos multiníveis sob distribuição de probabilidade da classe elíptica, constituem uma alternativa ao uso do modelo normal, e assim, pode-se abranger aplicações para conjunto de dados com estruturas mais complexas.

Um outro aspecto que vem sendo estudado é o que tange à robustez das estimativas de máxima verossimilhança, quando se tem observações aberrantes e se usa modelos normais multiníveis, ou modelos mistos com distribuição de erros normais. Pois, dados mais extremos que os demais tendem a influenciar na estimativa dos parâmetros. Desta forma uma má especificação do modelo pode acarretar erros na inferência dos parâmetros, então, uma alternativa seria o uso de distribuições que apresentem caudas mais pesadas que a normal, como por exemplo a distribuição t-Student. Lange, Little e Taylor (1989) propuseram um modelo geral com erros seguindo a distribuição t-Student multivariada a fim de alcançar estimativas robustas. Já em Pinheiro, Liu e Wu (2001), algoritmos eficientes para estimação robusta em modelos lineares mistos usando a distribuição t-Student multivariada foram propostos.

Os coeficientes ajustados para o modelo normal muitas vezes são sensíveis à observações extremas, então na literatura tem sido sugerido que a distribuição dos erros permita caudas mais pesadas que a normal a fim de acomodar melhor as observações extremas e diminuir a influência desses pontos. O método de influência local tem sido bastante aplicado a modelos multivariados de regressão linear normal, por exemplo em Kim (1995), Fung e Tang (1997), além do mais esses estudos vem sendo amplamente aplicados também para modelos de regressão linear elípticos no caso univariado e multivariado, como feitos por Galea, Paula e Bolfarine (1997) e Liu (2002). Como exemplos de trabalhos recente neste sentido, podemos mencionar Leiva *et al.* (2016) que desenvolveram técnicas de diagnóstico em modelos de regressão elípticos com restrições estocásticas aplicadas à Econometria: Manghi (2011) desenvolve alguns aspectos dos modelos elípticos multiníveis, como o processo de estimação dos parâmetros via máxima verossimilhança, testes de hipóteses para os efeitos fixos e parâmetros de variância e covariância e análise de resíduos para verificação de características relacionadas aos ajustes e às suposições estabelecidas. No entanto, a análise de influência local para modelos elípticos multiníveis ainda não foi desenvolvida, sendo esta a proposta deste trabalho.

Diagnóstico de influência é uma importante etapa a ser considerada na modelagem de dados. Esse passo é geralmente apresentado depois da estimação dos parâmetros para avaliar o desempenho das estimativas. Métodos de diagnósticos têm sido utilizados em modelos de regressão linear normal nas distribuições de classes elípticas, como pode-se ver em Paula (2004), Liu (2002), Cysneiros e Vanegas (2008) e em Cysneiros e Paula (2005), são abordados pontos de alavanca, resíduos, medidas de influência, influência local e curvatura normal, entre outros procedimentos.

Toda a análise residual para os modelos multiníveis elípticos foram desenvolvidas e podem ser vistos em Manghi (2011) e Manghi, Paula e Cysneiros (2016). Nosso trabalho veio a preencher parcialmente a lacuna da técnica de diagnóstico, na análise de influência local via curvatura normal para os modelos elípticos multiníveis.

Esta dissertação está dividida da seguinte maneira, apresentamos no Capítulo 2 uma breve introdução à classe das distribuições elípticas, bem como a definição do modelo multinível com 3 níveis. No Capítulo 3 desenvolvemos medidas de diagnóstico sobre a influência local e a curvatura normal na classe dos modelos elípticos sob quatro abordagens de perturbações. Vale salientar que nesse capítulo, encontra-se a principal contribuição teórica do presente trabalho, uma vez que, não há na literatura trabalhos referentes a influência local nos modelos multiníveis elípticos. Logo, nosso trabalho veio a preencher parcialmente esta lacuna importante. No Capítulo 4 mostramos dois exemplos com as aplicações de influência local apresentadas no Capítulo 3, um para modelos de dois níveis e outro para três níveis. Por fim, no Capítulo 5 apresentamos as considerações finais.

A presente dissertação foi integralmente digitada utilizando o sistema de tipografia \LaTeX . Todas as rotinas computacionais foram construídas com o software R, este disponível gratuitamente em *www.r – project.org*.

2 MODELOS ELÍPTICOS MULTINÍVEIS

Manghi (2011) obteve uma generalização do modelo normal multinível, chamado modelo elíptico multinível, que sugere o uso de distribuições de probabilidade pertencentes à classe elíptica, pois esta pode apresentar caudas mais leves ou mais pesadas que a normal. No caso de observações extremas, distribuições com caudas mais pesadas podem acomodar melhor os dados. Neste trabalho, foram desenvolvidos processo de estimação de parâmetros via máxima verossimilhança, teste de hipóteses para efeitos fixos e parâmetros de variância e covariância e análise de resíduos.

Neste capítulo apresentaremos uma breve introdução aos modelos elípticos multiníveis. Na seção 2.1 introduziremos a classe das distribuições elípticas, na seção 2.2 a estrutura de modelos multiníveis de 3 níveis, na seção 2.3 a estimação dos parâmetros, a função escore e a matriz de informação observada.

2.1 CLASSE DAS DISTRIBUIÇÕES ELÍPTICAS

Fang, Kotz e Ng (1990) apresentaram algumas das propriedades das distribuições multivariadas contínuas e com um desenvolvimento substancial na teoria de distribuição multivariada, especialmente na área de distribuições multivariadas simétricas não normais.

Nesta seção apresentamos uma breve introdução à classe de distribuição elíptica, as notações a serem usadas, assim como, os estimadores de máxima verossimilhança. A classe elíptica de distribuições inclui todas as distribuições contínuas simétricas com suporte nos reais.

Dizemos que um vetor aleatório \mathbf{Y}_k segue uma distribuição elíptica n_k -dimensional com vetor de média $\boldsymbol{\mu}_k$ e matriz proporcional de variância-covariância \mathbf{V}_k assumida ser positiva definida, como visto em Jennrich e Schluchter (1986), se sua função de densidade é da forma

$$f(\mathbf{y}_k) = |\mathbf{V}_k|^{-\frac{1}{2}} g(u_k), \quad k = 1, \dots, n, \quad (2.1)$$

em que $u_k = (\mathbf{y}_k - \boldsymbol{\mu}_k)^\top \mathbf{V}_k^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_k)$ é a distância de Mahalanobis e a função geradora de densidade $g(\cdot) : \mathbb{R} \rightarrow [0, \infty)$ é de tal forma que $\int_0^\infty u^{n_k/2-1} g(u) du < \infty$, (ver Fang, Kotz e Ng (1990)). Para a função acima, valem as seguintes relações: $E(\mathbf{Y}_k) = \boldsymbol{\mu}_k$ e $Var(\mathbf{Y}_k) = \alpha_k \mathbf{V}_k$, em que α_k é obtido através da função característica de \mathbf{Y}_k , que é dado da forma $\varphi Y_k^{(t)} = E[\exp(it^\top \mathbf{Y}_k)] = \exp(it^\top \boldsymbol{\mu}_k) \phi(\mathbf{t}^\top \mathbf{V}_k \mathbf{t})$, para alguma função $\phi(\cdot)$. Desse modo, tem-se $\alpha_k = -2\phi'(0)$, com $\phi'(t) = \frac{\partial \phi(t)}{\partial t}$ e $\phi'(0) = \frac{\partial \phi(t)}{\partial t} |_{t=0}$. Na distribuição t-Student tem-se que para $\eta_k > 2$, então $\alpha_k = \frac{\eta_k}{(\eta_k - 2)}$, em que η_k são os graus de liberdade. Já para a distribuição Expo-

nencial Potência, visto em Fang, Kotz e Ng (1990), tem-se que $\alpha_k = 2 \left(\frac{1}{\eta_k}\right) \Gamma\{(\eta_k + 2) \setminus 2\eta_k\} \setminus (\eta_k \Gamma\{\eta_k \setminus 2\eta_k\})$, em que $\Gamma(\cdot)$ é função gama com $\eta_k > 0$ representando o parâmetro de forma. Este sendo definido de maneiras diferentes para cada distribuição elíptica.

Denota-se $\mathbf{Y}_k \sim El_{n_k}(\boldsymbol{\mu}_k, \mathbf{V}_k, \eta_k)$. Para uma notação mais simplificada escrevemos $\mathbf{Y}_k \sim El_{n_k}(\boldsymbol{\mu}_k, \mathbf{V}_k)$. No caso em que $\mathbf{Y}_k \sim El_{n_k}(\mathbf{0}, \mathbf{I}_{n_k})$, tem-se \mathbf{I}_{n_k} como matriz identidade de dimensão $n_k \times n_k$, α_k a variância das distribuições marginais univariadas obtidas a partir do vetor \mathbf{Y}_k . Para a distribuição Normal temos $\alpha_k = 1$. Veja, por exemplo, Fang, Kotz e Ng (1990), Anderson (1990), Kai-Tai e Yao-Ting (1990).

Dada a Equação (2.1), o logaritmo da função de verossimilhança (log-verossimilhança) para o vetor de parâmetros $\boldsymbol{\tau} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ dado por $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \ell(\boldsymbol{\tau})$, é dado por

$$\ell(\boldsymbol{\tau}) = -\frac{1}{2} \sum_{k=1}^K \log |\mathbf{V}_k| + \sum_{k=1}^K \log \mathbf{g}(\mathbf{u}_k),$$

sabendo que $\mathbf{g}(\cdot)$ é uma função contínua e diferenciável.

2.2 MODELOS MULTINÍVEIS

O modelo de regressão multinível permite que diferentes níveis sejam especificados em modelos separados e depois combinados em um único modelo, como mostrado por Hox (1995) e Goldstein (2011), por exemplo. Neste tipo de modelo há um conjunto de dados hierárquicos que tem uma variável resposta Y e as variáveis explicativas que podem estar no nível do indivíduo e/ou no nível de grupos.

O modelo linear hierárquico ou a modelagem hierárquica, tem sido utilizada amplamente em várias áreas do conhecimento onde se quer examinar efeitos que as variáveis do nível do grupo e do nível do indivíduo tem sobre a variável resposta, como visto em Anderson (1990).

Em Manghi, Paula e Cysneiros (2016) este modelo é apresentado em estrutura com três níveis, logo utilizando-se das notações apresentadas, considera-se P variáveis explicativas associadas ao nível 1, $\sum_{p=1}^P Q_p$ são variáveis preditoras associadas ao nível 2, em que Q_p é a variável explicativa associada ao p -ésimo parâmetro de nível 1 e $\sum_{p=1}^P \sum_{q=1}^{Q_p} S_{pq}$ variáveis preditoras associadas ao nível 3, em que S_{pq} é o número de variáveis preditoras associadas ao q -ésimo parâmetro do nível 2 associado ao p -ésimo parâmetro de nível 1. Sendo assim, temos as

equações gerais para o modelo multinível com três níveis dadas como:

$$\begin{aligned}
 \text{Nível 1: } Y_{ijk} &= \sum_{p=1}^P \gamma_{pj|k} G_{pijk} + \varepsilon_{ijk}, \\
 \text{Nível 2: } \gamma_{pj|k} &= \sum_{q=1}^{Q_q} \delta_{pqk} T_{pqjk} + \xi_{pj|k}, \\
 \text{Nível 3: } \delta_{pqk} &= \sum_{s=1}^{S_{pq}} \beta_{pqs} L_{pqsk} + \vartheta_{pqk},
 \end{aligned} \tag{2.2}$$

em que $i = 1, 2, \dots, n_{jk}$, $j = 1, 2, \dots, m_k$ e $k = 1, 2, \dots, K$. Tem-se que n_{jk} é o número de indivíduos pertencentes aos j -ésimo “cluster” do nível 2 pertencente ao k -ésimo “cluster” do nível 3; m_k é o número de “cluster” do nível 2 pertencente ao k -ésimo “cluster” do nível 3 e K é o total de “clusters” do nível 3.

Então a equação (2.2) é descrita como:

Y_{ijk} é o valor da variável resposta para o i -ésimo indivíduo pertencente ao j -ésimo “cluster” do nível 2 no qual pertence ao k -ésimo “cluster” do nível 3;

G_{pijk} é o valor da p variável preditora relacionada ao i -ésimo indivíduo do j -ésimo “cluster” do nível 2 pertencente ao k -ésimo “cluster” do nível 3;

$\gamma_{pj|k}$ é o parâmetro associado à variável G_p no j -ésimo “cluster” do nível 2 incluído no k -ésimo “cluster” do nível 3;

ε_{ijk} é o erro associado ao valor da variável resposta do nível 1;

T_{pqjk} é o valor da q -ésima variável preditora associada ao parâmetro $\gamma_{pj|k}$;

δ_{pqk} é o parâmetro associado à variável T_{pq} no k -ésimo “cluster”;

$\xi_{pj|k}$ é o erro associado ao parâmetro $\gamma_{pj|k}$;

L_{pqsk} é o valor da s -ésima variável explicativa associada ao parâmetro δ_{pqk} ;

β_{pqs} é o parâmetro associado à variável L_{pqsk} ;

ϑ_{pqk} é o erro associado ao parâmetro δ_{pqk} .

Para um notação matricial deste modelo, temos

$$\mathbf{Y}_k = \left(\mathbf{Y}_{1k}^\top, \mathbf{Y}_{2k}^\top, \dots, \mathbf{Y}_{m_k k}^\top \right)_{\sum_{j=1}^{m_k} n_{jk} \times 1}^\top, \quad \mathbf{Y}_{jk} = \left(\mathbf{Y}_{1jk}, \mathbf{Y}_{2jk}, \dots, \mathbf{Y}_{n_{jk}jk} \right)_{n_{jk} \times 1}^\top,$$

$$\mathbf{G}_k = \text{diag} \{ \mathbf{G}_{1k}, \mathbf{G}_{2k}, \dots, \mathbf{G}_{m_k k} \}_{\left(\sum_{j=1}^{m_k} n_{jk} \right) \times (P m_k)}, \quad \mathbf{G}_{jk} = \{ G_{pijk} \}_{n_{jk} \times P},$$

em que

$$\mathbf{G}_{\mathbf{jk}} = \begin{bmatrix} G_{11jk} & G_{21jk} & \cdots & G_{P1jk} \\ G_{12jk} & G_{22jk} & \cdots & G_{P2jk} \\ \vdots & \vdots & \ddots & \vdots \\ G_{1n_{jk}jk} & G_{2n_{jk}jk} & \cdots & G_{Pn_{jk}jk} \end{bmatrix}_{n_{jk} \times P},$$

$$\boldsymbol{\gamma}_{\mathbf{k}} = (\boldsymbol{\gamma}_{1\mathbf{k}}^\top, \boldsymbol{\gamma}_{2\mathbf{k}}^\top, \dots, \boldsymbol{\gamma}_{m_{\mathbf{k}}\mathbf{k}}^\top)^\top, \quad \boldsymbol{\gamma}_{\mathbf{jk}} = (\boldsymbol{\gamma}_{1|k}, \boldsymbol{\gamma}_{2|k}, \dots, \boldsymbol{\gamma}_{P|k})_{P \times 1}^\top,$$

$$\boldsymbol{\varepsilon}_{\mathbf{k}} = (\boldsymbol{\varepsilon}_{1\mathbf{k}}^\top, \boldsymbol{\varepsilon}_{2\mathbf{k}}^\top, \dots, \boldsymbol{\varepsilon}_{m_{\mathbf{k}}\mathbf{k}}^\top)^\top, \quad \boldsymbol{\varepsilon}_{\mathbf{jk}} = (\boldsymbol{\varepsilon}_{1jk}, \boldsymbol{\varepsilon}_{2jk}, \dots, \boldsymbol{\varepsilon}_{n_{jk}jk})_{n_{jk} \times 1}^\top,$$

$$\mathbf{T}_{\mathbf{k}} = [\mathbf{T}_{1\mathbf{k}}^\top | \mathbf{T}_{2\mathbf{k}}^\top | \cdots | \mathbf{T}_{m_{\mathbf{k}}\mathbf{k}}^\top]_{(Pm_{\mathbf{k}}) \times (\sum_{p=1}^P Q_p)}^\top,$$

$$\mathbf{T}_{\mathbf{jk}} = \text{diag} \{ \mathbf{T}_{1\mathbf{jk}}, \mathbf{T}_{2\mathbf{jk}}, \dots, \mathbf{T}_{P\mathbf{jk}} \}_{P \times (\sum_{p=1}^P Q_p)}, \quad \mathbf{T}_{P\mathbf{jk}} = (T_{P1jk}, T_{P2jk}, \dots, T_{PQ_{jk}})_{1 \times Q_p},$$

e tem-se que

$$\mathbf{T}_{\mathbf{jk}} = \begin{bmatrix} \mathbf{T}_{1\mathbf{jk}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2\mathbf{jk}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{P\mathbf{jk}} \end{bmatrix}_{P \times (\sum_{p=1}^P Q_p)},$$

$$\boldsymbol{\delta}_{\mathbf{k}} = (\delta_{11k}, \delta_{12k}, \dots, \delta_{1Q_1k}, \delta_{21k}, \delta_{22k}, \dots, \delta_{2Q_2k}, \dots, \delta_{P1k}, \delta_{P2k}, \dots, \delta_{PQ_{Pk}})_{(\sum_{p=1}^P Q_p) \times 1}^\top,$$

$$\mathbf{L}_{\mathbf{k}} = \text{diag} \{ \mathbf{L}_{11\mathbf{k}}, \mathbf{L}_{12\mathbf{k}}, \dots, \mathbf{L}_{1Q_1\mathbf{k}}, \dots, \mathbf{L}_{P1\mathbf{k}}, \mathbf{L}_{P2\mathbf{k}}, \dots, \mathbf{L}_{PQ_{Pk}} \}_{(\sum_{p=1}^P Q_p) \times (\sum_{p=1}^P \sum_{q=1}^{Q_p} S_{pq})},$$

$$\mathbf{L}_{P\mathbf{qk}} = (L_{Pq1k}, L_{Pq2k}, \dots, L_{PqS_{pqk}})_{1 \times S_{pq}},$$

sendo

$$\mathbf{L}_k = \begin{bmatrix} \mathbf{L}_{11k} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{12k} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \mathbf{L}_{1Q_1k} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} & \mathbf{L}_{PQpk} \end{bmatrix}_{(\sum_{p=1}^P Q_p) \times (\sum_{p=1}^P \sum_{q=1}^{Q_p} S_{pq})},$$

$$\boldsymbol{\beta} = (\beta_{111}, \beta_{112}, \dots, \beta_{11S_{11}}, \dots, \beta_{PQp1}, \beta_{PQp2}, \dots, \beta_{PQpS_{PQp}})^\top_{(\sum_{p=1}^P \sum_{q=1}^{Q_p} S_{pq}) \times 1},$$

$$\boldsymbol{\xi}_k = (\xi_{1k}^\top, \xi_{2k}^\top, \dots, \xi_{m_k k}^\top)^\top_{Pm_k \times 1}, \quad \boldsymbol{\xi}_{jk} = (\xi_{1j|k}, \xi_{2j|k}, \dots, \xi_{Pj|k})^\top_{P \times 1},$$

e

$$\boldsymbol{\vartheta} = (\vartheta_{11k}, \vartheta_{12k}, \dots, \vartheta_{1Q_1k}, \dots, \vartheta_{P1k}, \vartheta_{P2k}, \dots, \vartheta_{PQpk})^\top_{(\sum_{p=1}^P Q_p) \times 1}.$$

Logo as seguintes equações para o k -ésimo “cluster” podem ser reescritas como

$$\mathbf{Y}_k = \mathbf{G}_k \boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k, \quad (2.3)$$

$$\boldsymbol{\gamma}_k = \mathbf{T}_k \boldsymbol{\delta}_k + \boldsymbol{\xi}_k, \quad (2.4)$$

$$\boldsymbol{\delta}_k = \mathbf{L}_k \boldsymbol{\beta} + \boldsymbol{\vartheta}_k, \quad (2.5)$$

e substituindo a Equação 2.5 na Equação 2.4 e esta na Equação 2.3, é obtido o seguinte:

$$\mathbf{Y}_k = \mathbf{G}_k [\mathbf{T}_k (\mathbf{L}_k \boldsymbol{\beta} + \boldsymbol{\vartheta}_k) \boldsymbol{\xi}_k] + \boldsymbol{\varepsilon}_k = \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \boldsymbol{\beta} + \mathbf{G}_k \mathbf{T}_k \boldsymbol{\vartheta}_k + \mathbf{G}_k \boldsymbol{\xi}_k + \boldsymbol{\varepsilon}_k.$$

Desta forma, tem-se que $\mathbf{Y}_k = \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \mathbf{b}_k + \boldsymbol{\varepsilon}_k$, em que $\mathbf{X}_k = \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k$, $\mathbf{Z}_k = [\mathbf{G}_k \mathbf{T}_k | \mathbf{G}_k]$ e $\mathbf{b}_k = (\boldsymbol{\vartheta}^\top, \boldsymbol{\xi}^\top)^\top$. Considera-se também que alguns parâmetros em $\boldsymbol{\gamma}_k$ ou em $\boldsymbol{\delta}_k$ serem considerados fixos, o vetor de efeitos aleatórios \mathbf{b}_k deve ser organizado omitindo-se os erros relacionados aos respectivos parâmetros, e as colunas de \mathbf{Z}_k relativas aos erros omitidos são retiradas. Prosseguindo, considerando-se os “clusters” K para o nível 3, tem-se a seguinte notação:

$$\begin{aligned}
\mathbf{Y} &= \left(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top, \dots, \mathbf{Y}_K^\top \right)^\top_{\left(\sum_{k=1}^K \sum_{j=1}^{m_k} n_{jk} \right) \times 1}, \\
\mathbf{G} &= \text{diag} \{ \mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_K \}_{\left(\sum_{k=1}^K \sum_{j=1}^{m_k} n_{jk} \right) \times \left(P \sum_{k=1}^K m_k \right)}, \\
\boldsymbol{\gamma} &= \left(\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top, \dots, \boldsymbol{\gamma}_K^\top \right)^\top_{\left(P \sum_{k=1}^K m_k \right) \times 1}, \quad \boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}_1^\top, \boldsymbol{\varepsilon}_2^\top, \dots, \boldsymbol{\varepsilon}_K^\top \right)^\top_{\left(\sum_{k=1}^K \sum_{j=1}^{m_k} n_{jk} \right) \times 1}, \\
\mathbf{T} &= \text{diag} \{ \mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_K \}_{\left(P \sum_{k=1}^K m_k \right) \times \left(K \sum_{p=1}^P Q_p \right)}, \\
\boldsymbol{\delta} &= \left(\boldsymbol{\delta}_1^\top, \boldsymbol{\delta}_2^\top, \dots, \boldsymbol{\delta}_K^\top \right)^\top_{\left(K \sum_{p=1}^P Q_p \right) \times 1}, \quad \boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top, \dots, \boldsymbol{\xi}_K^\top \right)^\top_{\left(P \sum_{k=1}^K m_k \right) \times 1}, \\
\mathbf{L} &= \left[\mathbf{L}_1^\top | \mathbf{L}_2^\top | \dots | \mathbf{L}_K^\top \right]^\top_{\left(K \sum_{p=1}^P Q_p \right) \times \left(\sum_{p=1}^P \sum_{q=1}^{Q_p} S_{pq} \right)}, \\
\boldsymbol{\vartheta} &= \left(\boldsymbol{\vartheta}_1^\top, \boldsymbol{\vartheta}_2^\top, \dots, \boldsymbol{\vartheta}_K^\top \right)^\top_{\left(K \sum_{p=1}^P Q_p \right) \times 1}.
\end{aligned}$$

Por conseguinte, o modelo multinível para os K “clusters” pode ser reescrito como

$$\mathbf{Y} = \mathbf{G}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2.6)$$

$$\boldsymbol{\gamma} = \mathbf{T}\boldsymbol{\delta} + \boldsymbol{\xi}, \quad (2.7)$$

$$\boldsymbol{\delta} = \mathbf{L}\boldsymbol{\beta} + \boldsymbol{\vartheta}, \quad (2.8)$$

e substituindo a equação (2.8) na equação (2.7) e este na equação (2.6), tem-se

$$\mathbf{Y} = \mathbf{G}[\mathbf{T}(\mathbf{L}\boldsymbol{\beta} + \boldsymbol{\vartheta}) + \boldsymbol{\xi}] + \boldsymbol{\varepsilon} = \mathbf{GTL}\boldsymbol{\beta} + \mathbf{GT}\boldsymbol{\vartheta} + \mathbf{G}\boldsymbol{\xi} + \boldsymbol{\varepsilon},$$

em que $\mathbf{X} = \mathbf{GTL}$, $\mathbf{Z} = [\mathbf{GT}|\mathbf{G}]$ e $\mathbf{b} = (\boldsymbol{\vartheta}^\top, \boldsymbol{\xi}^\top)^\top$, sendo assim, obtém-se $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$.

Os modelos elípticos multiníveis podem ser definidos a partir da estrutura apresentada para modelos normais multiníveis, no qual pode ser visto em Manghi (2011). Para algumas distribuições elípticas pode não ser tão fácil obter a distribuição marginal de Y_k , sendo preciso a utilização de métodos de integração numérica. Desta forma, uma outra abordagem do modelo pode ser utilizada, considerando

$$\begin{bmatrix} \mathbf{Y}_k \\ \mathbf{b}_k \\ \boldsymbol{\varepsilon}_k \end{bmatrix} \sim El_{n_k+P+n_k} \begin{bmatrix} \mathbf{X}_k\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_k\Psi_k\mathbf{Z}_k^\top + \mathbf{R}_k & \mathbf{Z}_k\Psi_k & \mathbf{R}_j \\ \Psi_k\mathbf{Z}_k^\top & \Psi_k & \mathbf{0} \\ \mathbf{R}_k & \mathbf{0} & \mathbf{R}_k \end{bmatrix}, \quad (2.9)$$

em que $k = 1, 2, \dots, K$, e segundo Arellano-Valle (1994), a partir das propriedades das distribuições elípticas temos que:

$$\mathbf{Y}_k \sim El_{n_k}(\mathbf{X}_k\boldsymbol{\beta}, \mathbf{Z}_k\Psi_k\mathbf{Z}_k^\top + \mathbf{R}_k). \quad (2.10)$$

Temos que nesta abordagem \mathbf{b}_k e ε_k são não correlacionados mas não necessariamente independentes.

2.3 ESTIMAÇÃO DOS PARÂMETROS E MATRIZ DE INFORMAÇÃO

Nos modelos multiníveis têm-se três tipos de parâmetros a serem estimados: efeitos fixos, efeitos aleatórios e componentes de variância e covariância, como pode ser visto em Raudenbush e Bryk (2002). As suposições sobre as distribuições de probabilidades dos erros no contexto destes modelos pode ser visto em Manghi (2011). O logaritmo da função de verossimilhança (log-verossimilhança) para o vetor de parâmetros $\tau = (\beta^\top, \theta^\top)^\top$ dado por $\ell(\beta, \theta) = \ell(\tau)$, então,

$$\ell(\tau) = -\frac{1}{2} \sum_{k=1}^K \log |\mathbf{V}_k| + \sum_{k=1}^K \log g(\mathbf{u}_k), \quad (2.11)$$

sabendo que $u_k = (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta)$ conhecida como distância de Mahalanobis com $k = 1, 2, \dots, K$ e $g(\cdot)$ uma função contínua e diferenciável e $\mathbf{V}_k = \mathbf{Z}_k \Psi_k \mathbf{Z}_k^\top + \mathbf{R}_k$. Definimos o vetor $\theta = (\theta_1, \theta_2, \dots, \theta_D)^\top$ o vetor de parâmetros da matriz \mathbf{V}_k e β os parâmetros de efeitos fixos. Então fazendo as derivadas parciais da equação (2.11) temos a função escore para β e θ da seguinte forma:

$$\mathbf{S}(\beta) = \sum_{k=1}^K v_k(u_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta), \quad (2.12)$$

e

$$\mathbf{S}(\theta_d) = -\frac{1}{2} \sum_{k=1}^K \left\{ \text{tr} \left(\mathbf{V}_k^{-1} \dot{\mathbf{V}}_k(\theta_d) \right) - v_k(u_k) \mathbf{r}_k^\top \mathbf{V}_k^{-1} \dot{\mathbf{V}}_k(\theta_d) \mathbf{V}_k^{-1} \mathbf{r}_k \right\},$$

para $k = 1, 2, \dots, K$, $\dot{\mathbf{V}}_k(\theta_d) = \frac{\partial \mathbf{V}_k}{\partial \theta_d}$, $\mathbf{r}_k = (\mathbf{Y}_k - \mathbf{X}_k \beta)$ e $v_k(u_k) = -2W_g(u_k)$. Como não é possível isolar os parâmetros, então é utilizado um procedimento iterativo para se obter as estimativas de máxima verossimilhança de β e θ , que é dado por

$$\hat{\beta}^{(r+1)} = \left(\sum_{k=1}^K v_k(\hat{u}_k)^{(r)} \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1(r)} \mathbf{X}_k \right)^{-1} \left(\sum_{k=1}^K v_k(\hat{u}_k)^{(r)} \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1(r)} \mathbf{y}_k \right), \quad (2.13)$$

e

$$\hat{\theta}^{(r+1)} = \arg_{\theta} \max \left\{ \ell \left(\hat{\beta}^{(r+1)}, \theta \right) \right\}, \quad (2.14)$$

com $r = 0, 1, \dots$. Este método de maximização na equação (2.14) pode ser visto em Jr e Schnabel (1996). Inicia-se com $\beta^{(0)}$ e $\theta^{(0)}$, desta forma, dados valores de β e θ , pode-se utilizar rotinas

computacionais já implementadas para modelos normais multiníveis e assim utilizá-las para a atualização de β e métodos de otimização não linear para obter as atualizações de θ . Vê-se que também aparecem as quantidades $v_k(u_k)$ e são interpretadas como pesos, sendo uma forma diferente para diferentes distribuições e estas podem ser vistas em Osorio, Paula e Galea (2007), Cysneiros, Paula e Galea (2005), desde que $g(u_k)$ seja uma função decrescente positiva.

A partir das segundas derivadas do logaritmo da verossimilhança temos $-\ddot{\ell}_{\tau\tau}$ a matriz de informação observada de Fisher, na qual obtemos como segue

$$f(\mathbf{y}_k; \beta, \theta) = |\mathbf{V}_k|^{-\frac{1}{2}} g(\mathbf{u}_k),$$

na qual $u_k = (\mathbf{y}_k - \mathbf{X}_k\beta)^\top \mathbf{V}_k^{-1} (\mathbf{y}_k - \mathbf{X}_k\beta)$ e $\mathbf{V}_k = \mathbf{Z}_k \Psi \mathbf{Z}_k^\top + \mathbf{R}_k$ para $k = 1, \dots, K$. E o produto das K contribuições é a função de verossimilhança, dada por:

$$L(\beta, \theta) = \prod_{k=1}^K L_k(\beta, \theta) = \prod_{k=1}^K |\mathbf{V}_k|^{-\frac{1}{2}} g(u_k),$$

assim, o logaritmo da função de verossimilhança é como se segue

$$\begin{aligned} \ell(\beta, \theta) = \log(L(\beta, \theta)) &= -\frac{1}{2} \sum_{k=1}^K \log |\mathbf{V}_k| + \sum_{k=1}^K \log g(\mathbf{u}_k), \\ &= \sum_{k=1}^K \left[-\frac{1}{2} \log |\mathbf{V}_k| + \log g(\mathbf{u}_k) \right]. \end{aligned}$$

Derivando a função do logaritmo da verossimilhança em relação a β

$$\begin{aligned} \frac{\partial \ell(\beta, \theta)}{\partial \beta} &= \sum_{k=1}^K \left[\frac{1}{g(\mathbf{u}_k)} g'(\mathbf{u}_k) \frac{\partial \mathbf{u}_k}{\partial \beta} \right] \\ &= \sum_{k=1}^K \left[\frac{g'(\mathbf{u}_k)}{g(\mathbf{u}_k)} 2[-\mathbf{X}_k]^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k\beta) \right] \\ &= \sum_{k=1}^K v(u_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k\beta), \end{aligned}$$

tem-se que $W_g(u) = \frac{g'(u)}{g(u)}$, $W'_g(u) = \frac{d}{du} W_g(u)$ e $v(u_k) = -2W(u_k)$. As expressões de $W_g(u)$ e $W'_g(u)$ assim como características das distribuições normal, t-Student e exponencial potência, que serão utilizadas neste trabalho, podem ser vistas em Osorio, Paula e Galea (2007), por exemplo. Além do que, dentre as distribuições pertencentes à classe elíptica, pode-se citar também a normal contaminada, t-Student generalizada, logística dos tipos I e II, logística generalizada, Kotz, Kotz generalizada, entre outras.

Assumindo o vetor $\theta = (\theta_1, \theta_2, \dots, \theta_D)^\top$ e derivando a função do logaritmo da verossimilhança em relação a θ obtêm-se

$$\frac{\partial \ell(\beta, \theta)}{\partial \theta_d} = -\frac{1}{2} \sum_{k=1}^K \left[\text{tr} \left(\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \right) - \mathbf{v}(\mathbf{u}_k) (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right].$$

As segundas derivadas são dadas por

$$\ddot{\ell}_{\beta\beta} = \frac{\partial^2 \ell(\beta, \theta)}{\partial \beta \partial \beta^\top} = 2 \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k) \hat{\mathbf{V}}_k + 2 \mathbf{W}'_g(\hat{\mathbf{u}}_k) \hat{\mathbf{r}}_k \hat{\mathbf{r}}_k^\top \right\} \hat{\mathbf{V}}_k^{-1} \mathbf{X}_k,$$

$$\ddot{\ell}_{\beta\theta} = \frac{\partial^2 \ell(\beta, \theta)}{\partial \beta \partial \theta^\top} = 2 \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k) \hat{\mathbf{V}}_k + \mathbf{W}'_g(\hat{\mathbf{u}}_k) \hat{\mathbf{r}}_k \hat{\mathbf{r}}_k^\top \right\} \hat{\mathbf{V}}_k^{-1} \dot{\mathbf{V}}_k(r) \hat{\mathbf{V}}_k^{-1} \hat{\mathbf{r}}_k,$$

$$\begin{aligned} \ddot{\ell}_{\theta\theta} = \frac{\partial^2 \ell(\beta, \theta)}{\partial \theta \partial \theta^\top} &= \frac{1}{2} \text{tr} \hat{\mathbf{V}}_k^{-1} \left\{ \dot{\mathbf{V}}_k(\mathbf{r}) \hat{\mathbf{V}}_k^{-1} \dot{\mathbf{V}}_k(\mathbf{s}) - \ddot{\mathbf{V}}_k(\mathbf{r}, \mathbf{s}) \right\} \\ &+ \hat{\mathbf{r}}_k^\top \hat{\mathbf{V}}_k^{-1} \left\{ \mathbf{W}'_g(\hat{\mathbf{u}}_k) \dot{\mathbf{V}}_k(\mathbf{r}) \hat{\mathbf{V}}_k^{-1} \hat{\mathbf{r}}_k \hat{\mathbf{r}}_k^\top \hat{\mathbf{V}}_k^{-1} \dot{\mathbf{V}}_k(r) - \mathbf{W}_g(\hat{\mathbf{u}}_k) \ddot{\mathbf{V}}_k(r, s) \right. \\ &\left. + \mathbf{W}_g(\hat{\mathbf{u}}_k) \dot{\mathbf{V}}_k(\mathbf{r}) \hat{\mathbf{V}}_k^{-1} \dot{\mathbf{V}}_k(\mathbf{s}) + \mathbf{W}_g(\hat{\mathbf{u}}_k) \dot{\mathbf{V}}_k(\mathbf{s}) \hat{\mathbf{V}}_k^{-1} \dot{\mathbf{V}}_k(\mathbf{r}) \right\} \hat{\mathbf{V}}_k^{-1} \hat{\mathbf{r}}_k, \end{aligned}$$

para $r, s = 1, \dots, k$. Sendo $\hat{\mathbf{u}}_k = \hat{\mathbf{r}}_k^\top \hat{\mathbf{V}}_k^{-1} \hat{\mathbf{r}}_k$, $\hat{\mathbf{r}}_k = \mathbf{Y}_k - \mathbf{X}_k \hat{\beta}$, $\hat{\mathbf{V}}_k = V_k(\hat{\theta})$, para $j = 1, \dots, n_k$. Tem-se também que $\dot{V}_k(r) = \partial V_k / \partial \theta_r |_{\theta = \hat{\theta}}$, $\ddot{V}_k(r, s) = \partial^2 V_k / \partial \theta_r \partial \theta_s |_{\theta = \hat{\theta}}$ e $r, s = 1, \dots, k$. Veja, por exemplo, Osorio, Paula e Galea (2007). Desta forma, a matriz de informação observada de Fisher é dada da seguinte maneira

$$\ddot{\ell}_{\tau\tau} = \begin{bmatrix} \ddot{\ell}_{\beta\beta} & \ddot{\ell}_{\beta\theta} \\ \ddot{\ell}_{\theta\beta} & \ddot{\ell}_{\theta\theta} \end{bmatrix}. \quad (2.15)$$

A matriz de informação de Fisher para o vetor de parâmetros $(\beta^\top, \theta^\top)^\top$, assume a seguinte forma

$$\kappa_{\beta\beta} = \sum_{k=1}^K \frac{4d_{gk}}{n_k} \mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \quad \text{e} \quad \kappa_{\theta\theta} = \{ \kappa_{\theta_l \theta_z} \},$$

em que

$$\kappa_{\theta_l \theta_z} = \sum_{k=1}^K \left\{ -\frac{c_{klz}}{4} + \frac{f_{gk}}{n_k(n_k + 2)} \left[c_{klz} + 2 \text{tr} \left(\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_l} \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_z} \right) \right] \right\},$$

é o elemento pertencente à l -ésima linha e z -ésima coluna da matriz $\kappa_{\theta\theta}$, em que $d_{gk} = E(W_g^2(u_k))$, $f_{gk} = E(W_g^2(u_k) u_k^2)$ e $c_{j lz} = \text{tr} \left(\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_l} \right) \text{tr} \left(\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_z} \right)$, com $l, z = 1, 2, \dots, D$. É possível

obter expressões de forma fechada para d_{g_k} e f_{g_k} para algumas distribuições elípticas. Em particular, para a distribuição t-Student com η graus de liberdade, temos

$$d_{g_k} = \frac{n_k}{4} \left(\frac{\eta + n_k}{\eta + n_k + 2} \right) \quad \text{e} \quad f_{g_k} = \frac{n_k(n_k + 2)}{4} \left(\frac{\eta + n_k}{\eta + n_k + 2} \right),$$

e para a distribuição exponencial potência com parâmetro de forma η , temos

$$d_{g_k} = \frac{\eta^2}{2^{1/\eta}} \Gamma\left(\frac{n_k - 2}{2\eta} + 2\right) / \Gamma\left(\frac{n_k}{2\eta}\right) \quad \text{e} \quad f_{g_k} = \frac{n_k(n_k + e\eta)}{4},$$

em que $\Gamma(\cdot)$ denota a função gama. A distribuição assintótica da matriz de variância e covariância dos estimadores $\hat{\beta}$ e $\hat{\theta}$, é respectivamente $\kappa^{-1}(\beta)$ e $\kappa^{-1}(\theta)$. A matriz de informação de Fisher para os parâmetros são ortogonais, sendo assim, assume um bloco diagonal para o vetor de parâmetros $(\beta^\top, \theta^\top)^\top$ dado por

$$\kappa = \begin{bmatrix} \kappa_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \kappa_{\theta\theta} \end{bmatrix}.$$

3 INFLUÊNCIA LOCAL

Métodos de diagnóstico foram amplamente estudados na literatura estatística na qual a maioria das obras focaram no estudo do efeito de eliminar observações sobre os resultados do modelo ajustado, como por exemplo em Eubank (1985) e Silverman (1985), entre outros. No entanto, exclusão de casos não manifesta ou mede claramente o impacto de outras perturbações no modelo. De forma alternativa, Cook (1986) propôs uma abordagem de influência em que a ideia principal é investigar e avaliar a influência conjunta de pontos sob certas perturbações nos dados (ou modelo) sem a necessidade da retirada de uma observação ou um conjunto de observações. Logo, ferramentas de influência local são muito úteis para avaliar o quão sensível são os modelos propostos quando pequenas perturbações são introduzidas.

Desta forma, a influência local é um método de análise de sensibilidade em que avalia a influência de pequenas perturbações em um modelo estatístico geral. Temos visto que métodos de diagnóstico vem sendo bastante utilizado, como por exemplo, Lesaffre e Verbeke (1998) utilizaram a abordagem de influência local em modelos de efeitos mistos no contexto de dados longitudinais, Zhu e Lee (2003) propuseram seis modelos de esquemas de perturbação para detectar observações influentes, na qual usou esse procedimento em uma situação hipotética e dois casos reais, Nobre (2004) propôs e discutiu técnicas de diagnóstico em modelos lineares mistos e Nobre e Singer (2007) revisaram algumas técnicas de análise de resíduos que foram usadas no contexto de modelos lineares mistos e propuseram uma padronização do resíduo condicional para identificar possíveis observações e aglomerados discrepantes.

Liu (2002) aplicou esta metodologia a modelos de regressão linear elípticos multivariados, Paula, Medeiros e Vilca-Labra (2009) usou a classe de modelos lineares com erros elípticos auto-regressivos de primeira ordem na qual as curvaturas normais para alguns esquemas de perturbações são derivadas. Em Ibacache-Pulgar e Paula (2011) a abordagem de influência local é usada para comparar a sensibilidade das estimativas do modelo sob erros com distribuição t-Student. Podemos ver este tipo de procedimento em Russo *et al.* (2012) em que foi proposto diagnósticos de influência em modelos elípticos não-lineares para dados correlacionados com estruturas heteroscedásticas e/ou autorregressivas e Villegas *et al.* (2013) o diagnóstico de influência é feito em modelos simétricos lineares generalizados.

Nestes trabalhos as curvaturas normais da influência local são utilizadas para estudar a sensibilidade das estimativas de máxima verossimilhança sob alguns esquemas de perturbação usuais, como: perturbação de casos, variável resposta, variável explicativa contínua e na matriz de

variância e covariância. Também são utilizados gráficos de diagnóstico para avaliar a sensibilidade dessas estimativas, como visto também em Villegas *et al.* (2013).

3.1 CURVATURA NORMAL

Ferramentas de influência local são utilizadas para avaliar a sensibilidade dos modelos propostos quando diferentes perturbações são introduzidas no modelo (ou dados). Seja $\ell(\tau)$ o logaritmo da função de verossimilhança do modelo na ausência de perturbação para o vetor de parâmetros τ e seja ω um vetor de perturbação restrita a algum subconjunto Ω . Logo, tem-se que a log-verossimilhança perturbada é denotada por $\ell(\hat{\tau}_\omega)$. Admite-se que existe um $\omega_0 \in \Omega$, denominado vetor de não perturbação, de tal modo que $\ell(\hat{\tau}_{\omega_0}) = \ell(\hat{\tau})$, desta forma, será avaliado se há influência da menor perturbação na estimativa de máxima verossimilhança estimada. Assim, pode-se considerar o deslocamento da verossimilhança como:

$$LD(\omega) = 2[\ell(\hat{\tau}) - \ell(\hat{\tau}_\omega)],$$

que é usado para avaliar a influência da perturbação ω . Grandes valores observados em $LD(\omega)$ evidencia que $\hat{\tau}$ e $\hat{\tau}_\omega$ diferem-se consideravelmente, em que $\hat{\tau}_\omega$ é o vetor de estimativas dos parâmetros da log-verossimilhança perturbada e $\hat{\tau}$ o vetor de estimativas dos parâmetros da log-verossimilhança não perturbada.

Por conseguinte, a abordagem estuda o comportamento local do gráfico de influência, desta forma, Cook (1986) sugeriu investigar a direção da curvatura máxima de \mathbf{C}_d , em que $\mathbf{C}_d = 2|\mathbf{d}^\top \mathbf{F} \mathbf{d}|$ com \mathbf{F} sendo uma matriz $\sum_{k=1}^K n_k \times \sum_{k=1}^K n_k$ e \mathbf{d} o vetor de direção com comprimento unitário ($\|\mathbf{d}\| = 1$). Sendo assim, para encontrar a curvatura máxima \mathbf{C}_{\max} e o vetor de direção máxima \mathbf{d}_{\max} , precisa-se calcular

$$F = -\Delta(\hat{\tau}, \omega_0)^\top \ddot{\ell}(\hat{\tau})^{-1} \Delta(\hat{\tau}, \omega_0), \quad (3.1)$$

em que $\ddot{\ell}$ está definida na equação (2.15) e

$$\Delta(\tau, \omega) = \frac{\partial^2 \ell(\tau|\omega)}{\partial \tau \partial \omega^\top} = \begin{pmatrix} \frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega^\top} \\ \frac{\partial^2 \ell(\tau|\omega)}{\partial \theta \partial \omega^\top} \end{pmatrix} = \begin{pmatrix} \Delta_1(\beta, \omega) \\ \Delta_2(\theta, \omega) \end{pmatrix}. \quad (3.2)$$

Vale ressaltar que a equação (3.2) deve ser avaliada em $\tau = \hat{\tau}$ e $\omega = \omega_0$. Então, \mathbf{d}_{\max} é um vetor unitário associado ao maior autovalor absoluto \mathbf{C}_{\max} de F dado na equação

(3.1). Desta forma, tem-se que grandes valores absolutos dos elementos de \mathbf{C}_{\max} revelam que as observações são provavelmente influentes. Este método de detectar influencia pode ser visto em, Cook (1986), Verbeke e Molenberghs (2009) e Villegas *et al.* (2013), por exemplo.

A curvatura normal \mathbf{C}_d pode assumir qualquer valor não negativo e não é invariante sob transformações de escala. Então, pode-se usar como alternativa a curvatura normal conformal, para ter uma curvatura invariante sob uma mudança uniforme de escala, chamada \mathbf{B}_d na direção \mathbf{d} , esta medida varia de $0 \leq \mathbf{B}_d \leq 1$. Esta quantidade foi introduzida em Poon e Poon (1999), ela é obtida utilizando os cálculos de \mathbf{C}_d , e é definida por:

$$\mathbf{B}_d = \frac{|-\mathbf{d}^\top \Delta(\hat{\tau}, \boldsymbol{\omega}_0)^\top \ddot{\ell}(\hat{\tau})^{-1} \Delta(\hat{\tau}, \boldsymbol{\omega}_0) \mathbf{d}|}{\sqrt{\text{tr}(\Delta(\hat{\tau}, \boldsymbol{\omega}_0)^\top \ddot{\ell}(\hat{\tau})^{-1} \Delta(\hat{\tau}, \boldsymbol{\omega}_0))^2}}. \quad (3.3)$$

Na análise da influência local uma técnica usada para verificar se uma observação é possível influente ou não, foi proposta por Verbeke e Molenberghs (2009) e Poon e Poon (1999) definir o ponto de corte, desta forma é feito $2 \times \bar{\mathbf{B}}_d$ e/ou $3 \times \bar{\mathbf{B}}_d$, ou seja, é traçado no gráfico de perturbações a média da medida de curvatura normal conformal e valores de B_d muito acima dessa medida merece uma atenção especial. Veja também exemplos em Russo, Paula e Aoki (2009) e Russo *et al.* (2012). Logo após os gráficos de influência local são usados para selecionar modelos que se mostram menos sensíveis a alguns esquemas de perturbação.

Utilizando uma abordagem similar àquela utilizada na análise de alavancagem, vamos usar uma medida para analisar a influência que cada observação ou grupo de observações tem sobre o modelo. No caso em que se retira todo o grupo de observações temos $\frac{3x}{K}$, em que x é a quantidade de grupos analisados e K o total do grupos. Para analisar observações temos $\frac{3x}{N}$ em que x é a quantidade de observações consideradas e N o total de observações. Veja, por exemplo, Verbeke e Molenberghs (2009).

3.2 ESQUEMAS DE PERTURBAÇÃO

Em muitos trabalhos vê-se a abordagem de influência local para diferentes tipos de modelos em Lesaffre e Verbeke (1998), trata de influência local em modelos lineares mistos, em Zhu e Lee (2003) propõem um procedimento para detectar observações influentes em seis esquemas de perturbação para modelos lineares mistos generalizados. Em Ibacache-Pulgar, Paula e Galea (2012) a abordagem de influência local é usada para comparar a sensibilidade das estimativas de modelos lineares semi-paramétricos mistos com erros normais, a fim de permitir

distribuições com caudas mais pesadas do que as normais em que acomodam melhor valores extremos.

No trabalho de Leiva *et al.* (2016) foi proposto uma metodologia de diagnóstico de influência para modelos de regressão linear com restrições e erros estocásticos e estudado como uma perturbação pode impactar no procedimento de estimação de parâmetros no modelo misto. Nesta seção trabalha-se com a influência local inicialmente vista em Cook (1986) para alguns esquemas de perturbação usuais no modelo ou nos dados. Os esquemas de perturbação considerados são de caso, na matriz de variância e covariância (V_k), na variável resposta e na variável explicativa. Estes esquemas de perturbação foram explanados em diferentes abordagens de modelo como visto em Liu (2002), Zhu e Lee (2003), Nobre (2004), Nobre e Singer (2007), Paula, Medeiros e Vilca-Labra (2009), Ibacache-Pulgar, Paula e Galea (2014).

A perturbação de caso é considerada para detectar observações que podem exercer grande influência no processo de estimação e conseqüentemente afeta as estimativas dos parâmetros. A perturbação na matriz de variância e covariância é usada para avaliar a sensibilidade a pequenas modificações nas estimativas de máxima verossimilhança. Perturbação de variáveis explicativas contínuas é usada para detectar observações cujos valores de variáveis explicativas podem exercer grande influência sobre as estimativas de máxima verossimilhança, assim como na variável resposta. Estes esquemas também foram usados em Osorio, Paula e Galea (2007) para modelos elípticos com estrutura longitudinal e Ibacache-Pulgar, Paula e Galea (2012) em diagnóstico de influência para modelos mistos semi-paramétricos elípticos.

3.2.1 Perturbação de caso

Seja $\omega_0 = (1, \dots, 1)^\top$ o vetor de não perturbação, tal que $\ell(\tau|\omega_0) = \ell(\tau)$. Adicionalmente, tem-se que $\mathbf{W} = \text{diag}(\omega)$ é uma matriz $n_k \times n_k$ que contém as perturbações $\omega_{\mathbf{k}} = (\omega_{1jk}, \dots, \omega_{n_{jk}})^\top$ na sua diagonal, que satisfaz $0 < \omega_{\mathbf{k}} < 1$. Desta forma os ω_{ijk} s são positivos e são pesos correspondentes ao caso (ijk) . Sabe-se que o vetor de parâmetros é dado por: $\tau = (\beta^\top, \theta^\top)^\top$, então:

$$\begin{aligned} \ell(\tau|\omega) &= \sum_{k=1}^K \omega_k L_k(\tau) \\ &= \sum_{k=1}^K \omega_k \left\{ -\frac{1}{2} \log |\mathbf{V}_{\mathbf{k}}| + \log(\mathbf{u}_{\mathbf{k}}) \right\}, \end{aligned}$$

sabendo que $\mathbf{u}_{\mathbf{k}} = (\mathbf{Y}_{\mathbf{k}} - \mathbf{X}_{\mathbf{k}}\beta)^\top \mathbf{V}_{\mathbf{k}}^{-1} (\mathbf{Y}_{\mathbf{k}} - \mathbf{X}_{\mathbf{k}}\beta)$.

A primeira derivada da verossimilhança perturbada em relação a β é dada por:

$$\frac{\partial \ell(\tau|\omega)}{\partial \beta} = \sum_{k=1}^K \omega_k \left\{ v(u_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}. \quad (3.4)$$

Obtendo a derivada em relação a θ_d , temos:

$$\frac{\partial \ell(\tau|\omega)}{\partial \theta_d} = -\frac{1}{2} \sum_{k=1}^K \omega_k \left\{ \text{tr} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \right] - v(u_k) (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}. \quad (3.5)$$

Obtendo as segundas derivadas das equações (3.4) e (3.5) em relação a ω respectivamente, temos :

$$\frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega_k^\top} = \sum_{k=1}^K \hat{v}(u_k) \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}), \quad (3.6)$$

e

$$\frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k^\top} = -\frac{1}{2} \sum_{k=1}^K \left\{ \text{tr} \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \right] - \hat{v}(u_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top \hat{\mathbf{V}}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\}, \quad (3.7)$$

então estas foram as derivadas da verossimilhança que são necessárias para a perturbação de caso avaliando $\tau = \hat{\tau}$ e $\omega = \omega_0$.

Este esquema de perturbação permite identificar e avaliar se as observações exercem influência notável nas estimadores dos parâmetros de máxima verossimilhança, veja por exemplo em Rodrigues (2013) e SILVA (2016).

3.2.2 Perturbação na matriz de variância e covariância

Para perturbar $\mathbf{V}_k = \mathbf{Z}_k \Psi \mathbf{Z}_k^\top + \mathbf{R}_k$ os termos da matriz serão multiplicados por um escalar $\omega_k = (\omega_{1jk}, \dots, \omega_{n_{jk}})^\top$ na log-verossimilhança, sabendo que k é o “cluster” de nível 3 e tem-se $k = 1, \dots, K$. Neste caso, o vetor de não perturbação é dado por $\omega_0 = (1, \dots, 1)^\top$. Então a verossimilhança é dada por:

$$\ell(\tau) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log g(\mathbf{u}_k) \right\},$$

então com a verossimilhança perturbada, temos :

$$\ell(\tau|\omega) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\omega_k \mathbf{V}_k| + \log g \left[(\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right] \right\}.$$

Obtendo a log-verossimilhança perturbada faremos a primeira derivada em relação a β , segue como:

$$\frac{\partial \ell(\tau|\omega)}{\partial \beta} = \sum_{k=1}^K \left\{ v(u_k \omega_k) \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}, \quad (3.8)$$

e temos que:

$$\mathbf{u}_k \omega_k = (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta),$$

e sua derivada é dada da forma:

$$\frac{\partial \mathbf{u}_k \omega_k}{\partial \beta} = -2 \mathbf{X}_k^\top \omega_k \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta).$$

Então, fazendo a segunda derivada de β em relação a ω , obtemos:

$$\begin{aligned} \frac{\partial^2 \ell(\theta|\omega)}{\partial \beta \partial \omega_k^\top} &= -2 \sum_{k=1}^K \left\{ \mathbf{W}'_g(\mathbf{u}_k \omega_k) \frac{\partial \mathbf{u}_k \omega_k}{\partial \omega_k} \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right. \\ &\quad \left. + \mathbf{W}_g(\mathbf{u}_k \omega_k) \mathbf{X}_k^\top [-\omega_k]^{-2} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}, \end{aligned}$$

avaliando a derivada acima em $\tau = \hat{\tau}$; $\omega = \omega_0$, temos:

$$\begin{aligned} &= -2 \sum_{k=1}^K \left\{ \mathbf{W}'_g(\hat{\mathbf{u}}_k) \hat{\mathbf{u}}_k \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) + \mathbf{W}_g(\hat{\mathbf{u}}_k) \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\} \\ &= -2 \sum_{k=1}^K \left\{ [\mathbf{W}'_g(\hat{\mathbf{u}}_k) \hat{\mathbf{u}}_k + \mathbf{W}_g(\hat{\mathbf{u}}_k)] \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\}. \end{aligned} \quad (3.9)$$

Agora obteremos a primeira derivada da log-verossimilhança em relação a θ_d , temos que:

$$\begin{aligned} \frac{\ell(\tau|\omega)}{\partial \theta_d} &= \sum_{k=1}^K \left\{ -\frac{1}{2} \text{tr} \left[\omega_k \mathbf{V}_k^{-1} \frac{\partial \omega_k \mathbf{V}_k}{\partial \theta_d} \right] + \frac{\mathbf{g}'(\mathbf{u}_k \omega_k)}{\mathbf{g}(\mathbf{u}_k \omega_k)} \frac{\partial \mathbf{u}_k \omega_k}{\partial \theta_d} \right\} \\ &= - \sum_{k=1}^K \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top [\omega_k]^{-1} \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\}. \end{aligned} \quad (3.10)$$

A partir da equação anterior teremos a segunda derivada em relação a ω_k , que segue da seguinte maneira:

$$\begin{aligned} \frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k^\top} &= - \left\{ \mathbf{W}'_g(\mathbf{u}_k \omega_k) \frac{\partial \mathbf{u}_k \omega_k}{\partial \omega_k} (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [\omega_k]^{-1} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \beta) + \right. \\ &\quad \left. \mathbf{W}_g(\mathbf{u}_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [-\omega_k]^{-2} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}, \end{aligned}$$

avaliando a equação anterior com $\tau = \hat{\tau}$; $\omega = \omega_0$, temos

$$= - \left\{ \mathbf{W}'_g(\hat{\mathbf{u}}_k) \hat{\mathbf{u}}_k + \mathbf{W}_g(\hat{\mathbf{u}}_k) \right\} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}). \quad (3.11)$$

Este esquema de perturbação além de revelar quais indivíduos que são mais influentes, no sentido, do deslocamento da probabilidade na estrutura da escala e, conseqüentemente, na estimativa de θ , (ver Osorio, Paula e Galea (2007), por exemplo). Para mais detalhes nos cálculos ver Apêndice B.

3.2.3 Perturbação na variável resposta

Temos o modelo dado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$

em que $\mathbf{X} = \mathbf{GTL}$, $\mathbf{Z} = [\mathbf{GT}|\mathbf{G}]$ e $\mathbf{b} = (\mathbf{v}^\top, \boldsymbol{\xi}^\top)^\top$.

A perturbação na variável resposta $(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top, \dots, \mathbf{Y}_K^\top)^\top$ é introduzida substituindo \mathbf{Y}_k por $\mathbf{Y}_k + \boldsymbol{\omega}_k$. Aqui, considere uma perturbação aditiva sobre a i -ésima resposta, ou seja, a perturbação para cada indivíduo fica $(y_{i\omega} = y_{ijk} + \omega_{ijk})$. Neste caso o vetor de não perturbação é $\boldsymbol{\omega}_0 = (\mathbf{0}, \dots, \mathbf{0})^\top$.

Então temos a log-verossimilhança dada por:

$$\ell(\boldsymbol{\tau}) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log \mathbf{g} \left[(\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right] \right\},$$

e inserindo a perturbação na variável resposta, temos:

$$\ell(\boldsymbol{\tau}|\boldsymbol{\omega}) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log \left[((\mathbf{Y}_k + \boldsymbol{\omega}_k) - \mathbf{X}_k \boldsymbol{\beta})^\top \mathbf{V}_k^{-1} ((\mathbf{Y}_k - \boldsymbol{\omega}_k) - \mathbf{X}_k \boldsymbol{\beta}) \right] \right\}.$$

Neste esquema de perturbação por exemplo perturbaremos o nível 1, ou seja, cada indivíduo da resposta. Desta forma, temos que: $\mathbf{Y}_k = [\mathbf{Y}_{1k}^\top, \mathbf{Y}_{2k}^\top, \dots, \mathbf{Y}_{m_k k}^\top]_{\left(\sum_{j=1}^{m_k}\right) \times 1}$ e $\boldsymbol{\omega}_k = [\boldsymbol{\omega}_{1k}^\top, \boldsymbol{\omega}_{2k}^\top, \dots, \boldsymbol{\omega}_{m_k k}^\top]_{\left(\sum_{j=1}^{m_k}\right) \times 1}$ em que \mathbf{Y}_k é o k -ésimo “cluster” de nível três e m_k o total de indivíduos pertencente ao “cluster” k .

Temos ainda que: $\mathbf{Y}_{jk} = [\mathbf{Y}_{1jk}, \mathbf{Y}_{2jk}, \dots, \mathbf{Y}_{n_{jk}jk}]_{n_{jk} \times 1}^\top$ e para este j -ésimo “cluster” de nível dois pertencente ao k -ésimo “cluster” de nível três, temos o vetor de perturbação

$\boldsymbol{\omega}_{jk} = [\boldsymbol{\omega}_{1jk}, \boldsymbol{\omega}_{2jk}, \dots, \boldsymbol{\omega}_{n_{jk}jk}]_{n_{jk} \times 1}^\top$, em que n_{jk} é o número de indivíduos pertencente ao j -ésimo “cluster” no nível 2 pertencente ao k -ésimo “cluster” do nível 3. Desta forma a variável perturbada é dado da seguinte maneira:

$$\begin{aligned} (\mathbf{Y}_{jk} + \boldsymbol{\omega}_{jk})^\top &= [Y_{1jk}, Y_{2jk}, \dots, Y_{n_{jk}jk}]^\top + [\boldsymbol{\omega}_{1jk}, \boldsymbol{\omega}_{2jk}, \dots, \boldsymbol{\omega}_{n_{jk}jk}]^\top \\ &= [Y_{1jk} + \boldsymbol{\omega}_{1jk}, Y_{2jk} + \boldsymbol{\omega}_{2jk}, \dots, Y_{n_{jk}jk} + \boldsymbol{\omega}_{n_{jk}jk}]^\top. \end{aligned}$$

Então, a primeira derivada em relação a β é dada por:

$$\begin{aligned} \frac{\partial \ell(\tau|\omega)}{\partial \beta} &= \frac{\mathbf{g}'(\mathbf{u}_k \omega_k)}{\mathbf{g}(\mathbf{u}_k \omega_k)} \frac{\partial(\mathbf{u}_k \omega_k)}{\partial \beta} \\ &= -2\mathbf{W}_g(\mathbf{u}_k \omega_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} [\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta]. \end{aligned} \quad (3.12)$$

Com a segunda derivada em relação a ω_k e avaliando $\tau = \hat{\tau}$; $\omega = \omega_0$, temos

$$\frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega_k^\top} = -2\mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k) \hat{\mathbf{V}}_k^{-1} + 2\mathbf{W}'_g(\hat{\mathbf{u}}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top \right\} \mathbf{V}_k^{-1}. \quad (3.13)$$

A primeira derivada da log-verossimilhança em relação θ_d fica da forma:

$$\frac{\partial \ell(\tau|\omega)}{\partial \theta_d} = \mathbf{W}_g(\mathbf{u}_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[-\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta). \quad (3.14)$$

Realizando a segunda derivada da equação (3.14) em relação a ω_k , obtemos a equação:

$$\begin{aligned} \frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k^\top} &= - \left\{ 2\mathbf{W}'_g(\mathbf{u}_k \omega_k) \mathbf{V}_k^{-1} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \times \right. \\ &\quad \times \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \\ &\quad + \mathbf{W}_g(\mathbf{u}_k \omega_k) \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \\ &\quad \left. + \mathbf{W}_g(\mathbf{u}_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] \right\}, \end{aligned}$$

e avaliamos a equação anterior em $\tau = \hat{\tau}$; $\omega = \omega_0$ e temos:

$$= -2(\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k) \hat{\mathbf{V}}_k + \mathbf{W}'_g(\hat{\mathbf{u}}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top \right\} \hat{\mathbf{V}}_k^{-1}. \quad (3.15)$$

Tem-se para esta perturbação, o gráfico de índices de B_d pode revelar observações com alta influência em seus próprios valores ajustados. Essas considerações podem ser vistas em Wei, Hu e Fung (1998) para alavancagens generalizadas. Para mais detalhes ver Apêndice C.

3.2.4 Perturbação na variável preditora: Nível 1

Considerando que temos um modelo de três níveis, a log-verossimilhança é dada por:

$$\ell(\tau) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log g \left[(\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \beta)^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \beta) \right] \right\}.$$

Então neste caso, queremos fazer a perturbação em uma variável preditora de nível 1, $(G_k + \omega_k s_k)$, ou seja, $(G_{ijk} + \omega_{ijk} s_k)$, perturbar cada observação da variável explicativa $(G_{ijk}$

associada a um indivíduo i , e s_k um fator de escala. Aqui é considerado o vetor de não perturbação como $\boldsymbol{\omega}_0 = (\mathbf{0}, \dots, \mathbf{0})^\top$. Logo, esse esquema de perturbação é dado por:

$$\ell(\boldsymbol{\tau}|\boldsymbol{\omega}) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log g \left[(\mathbf{Y}_k - (\mathbf{G}_k + \boldsymbol{\omega}_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \boldsymbol{\beta})^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - (\mathbf{G}_k + \boldsymbol{\omega}_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \boldsymbol{\beta}) \right] \right\},$$

então: $\mathbf{G}_k = \text{diag} \{ \mathbf{G}_{1k}, \mathbf{G}_{2k}, \dots, \mathbf{G}_{m_k k} \}$ e $\boldsymbol{\omega}_k = \text{diag} \{ \omega_{1k} \mathbf{s}_{1k}, \omega_{2k} \mathbf{s}_{2k}, \dots, \omega_{m_k k} \mathbf{s}_{m_k k} \}$.

Portanto, $\mathbf{G}_k \boldsymbol{\omega}_k = [\mathbf{G}_k + \boldsymbol{\omega}_k \mathbf{s}_k]$, e segue que:

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{G}_{1k} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{2k} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_{m_k k} \end{bmatrix} \text{ e } \boldsymbol{\omega}_k = \begin{bmatrix} \omega_{1k} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \omega_{2k} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \omega_{m_k k} \end{bmatrix},$$

sabendo que $k = 1, \dots, K$ e K é o total de “cluster” no nível 3.

Temos que G_{jk} é a variável preditora do j -ésimo “cluster” pertencente ao nível 2 pertencente ao k -ésimo “cluster” do nível 3. Vamos escolher apenas uma das P variáveis da matriz G_{jk} para somar a perturbação, então temos

$$\mathbf{G}_{jk} = \begin{bmatrix} G_{11jk} & G_{21jk} & \dots & G_{s1jk} & \dots & G_{P1jk} \\ G_{12jk} & G_{22jk} & \dots & G_{s2jk} & \dots & G_{P2jk} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{1n_{jk}jk} & G_{2n_{jk}jk} & \dots & G_{sn_{jk}jk} & \dots & G_{Pn_{jk}jk} \end{bmatrix} \text{ e } \boldsymbol{\omega}_{jk} = \begin{bmatrix} 0 & 0 & \dots & \omega_{s1jk} & \dots & 0 \\ 0 & 0 & \dots & \omega_{s2jk} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_{sn_{jk}jk} & \dots & 0 \end{bmatrix},$$

logo, temos que a matriz $\mathbf{G}_{jk} \boldsymbol{\omega}_k$ é dada da seguinte maneira:

$$\mathbf{G}_{jk} \boldsymbol{\omega}_k = \begin{bmatrix} G_{11jk} & G_{21jk} & \dots & G_{s1jk} + \omega_{s1jk} & \dots & G_{P1jk} \\ G_{12jk} & G_{22jk} & \dots & G_{s2jk} + \omega_{s2jk} & \dots & G_{P2jk} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{1n_{jk}jk} & G_{2n_{jk}jk} & \dots & G_{sn_{jk}jk} + \omega_{sn_{jk}jk} & \dots & G_{Pn_{jk}jk} \end{bmatrix}.$$

A primeira derivada da log-verossimilhança perturbada em relação e $\boldsymbol{\beta}$:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\tau}|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\mathbf{g}'(\mathbf{u}_k \boldsymbol{\omega}_k)}{\mathbf{g}(\mathbf{u}_k \boldsymbol{\omega}_k)} \frac{\partial \mathbf{u}_k \boldsymbol{\omega}_k}{\partial \boldsymbol{\beta}} \\ &= -2\mathbf{W}_g(\mathbf{u}_k \boldsymbol{\omega}_k) \mathbf{X}_{k\boldsymbol{\omega}_k}^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{k\boldsymbol{\omega}_k} \boldsymbol{\beta}) \\ &= -2\mathbf{W}_g(\mathbf{u}_k \boldsymbol{\omega}_k) [(\mathbf{G}_k + \boldsymbol{\omega}_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k]^\top \mathbf{V}_k^{-1} [\mathbf{Y}_k - (\mathbf{G}_k + \boldsymbol{\omega}_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \boldsymbol{\beta}]. \end{aligned} \quad (3.16)$$

Obtendo a segunda derivada da equação anterior em relação a ω_k , temos que:

$$\frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega_k^\top} = \left\{ \left[\mathbf{W}'_g(\mathbf{u}_k \omega_k) ((-\mathbf{2s}_k) \beta \mathbf{V}_k^{-1} (\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \beta)) ((\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k)^\top + \mathbf{W}_g(\mathbf{u}_k \omega_k) \mathbf{s}_k \right] \mathbf{V}_k^{-1} (\mathbf{Y}_k - (\mathbf{G}_k + \omega_k) \mathbf{T}_k \mathbf{L}_k) \beta + \mathbf{W}_g(\mathbf{u}_k \omega_k) ((\mathbf{G}_k + \omega_k) \mathbf{T}_k \mathbf{L}_k)^\top \mathbf{V}_k^{-1} (-\mathbf{s}_k) \beta \right\},$$

agora, avaliando $\tau = \hat{\tau}$; $\omega = \omega_0$ obtemos:

$$\begin{aligned} &= 4\mathbf{W}'_g(\hat{\mathbf{u}}_k) \mathbf{s}_k \hat{\beta} (\mathbf{G}_k \mathbf{T}_k \mathbf{L}_k)^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta}) (\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta})^\top \hat{\mathbf{V}}_k^{-1} \\ &- 2\mathbf{W}_g(\hat{\mathbf{u}}_k) \mathbf{s}_k \left[\mathbf{c}(\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k)^\top - \hat{\beta} (\mathbf{G}_k \mathbf{T}_k \mathbf{L}_k)^\top \right] \hat{\mathbf{V}}_k^{-1}. \end{aligned} \quad (3.17)$$

Então, temos que a primeira derivada da log-verossimilhança em relação a θ_d :

$$\begin{aligned} \frac{\partial \ell(\tau|\omega)}{\partial \theta_d} &= -\mathbf{W}_g(\mathbf{u}_k \omega_k) [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta]^\top \times \\ &\times \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta], \end{aligned} \quad (3.18)$$

desta maneira pode-se fazer a segunda derivada da log-verossimilhança da equação acima em relação a ω :

$$\begin{aligned} \frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k^\top} &= - \left\{ \mathbf{W}'_g(\mathbf{u}_k \omega_k) (-\mathbf{2s}_k \mathbf{T}_k \mathbf{L}_k \beta) \mathbf{V}_k^{-1} [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta] \times \right. \\ &\times [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta]^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta] \\ &+ \mathbf{W}_g(\mathbf{u}_k \omega_k) (-\mathbf{s}_k \mathbf{T}_k \mathbf{L}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta] \\ &\left. + \mathbf{W}_g(\mathbf{u}_k \omega_k) [\mathbf{Y}_k - (\mathbf{G}_k + \omega_k \mathbf{s}_k) \mathbf{T}_k \mathbf{L}_k \beta]^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (-\mathbf{s}_k \mathbf{T}_k \mathbf{L}_k \beta) \right\}, \end{aligned}$$

e por conseguinte, avaliando em $\tau = \hat{\tau}$; $\omega = \omega_0$

$$\begin{aligned} &= 2\mathbf{s}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta} [\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta}]^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] \times \\ &\times \left\{ \mathbf{W}_g(\hat{\mathbf{u}}_k) \hat{\mathbf{V}}_k + \mathbf{W}'_g(\hat{\mathbf{u}}_k) [\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta}] [\mathbf{Y}_k - \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \hat{\beta}]^\top \right\} \hat{\mathbf{V}}_k^{-1}. \end{aligned} \quad (3.19)$$

Esta perturbação é realizada na variável contínua explicativa do modelo e pode ser aplicado a qualquer nível. Este esquema tem por objetivo verificar se valores de variáveis explicativas contínuas que são muito sensíveis no sentido do deslocamento da verossimilhança, veja, por exemplo Osorio, Paula e Galea (2007).

4 EXEMPLOS DE APLICAÇÕES

Nesta seção, consideramos aplicações de dois conjuntos de dados utilizando a abordagem de influência local apresentada no Capítulo 3 para modelos multiníveis. Uma das aplicações envolve um conjunto de dados para modelos multiníveis com dois níveis e o segundo exemplo um modelo multinível com três níveis. Para o nível dois tem-se unidades de nível 1 que são agrupadas em unidades de nível 2. Para o nível três tem-se unidades de nível 1 que são agrupadas em unidades de nível 2 que por sua vez são selecionadas em unidades de nível 3.

4.1 FILHOTES DE RATOS

Este primeiro conjunto de dados, pode ser encontrado em Pinheiro e Bates (2006) e retrata o estudo de um experimento com 30 ratos fêmeas, que consiste em utilizar três tipos de dosagens de tratamento nas mães de um composto experimental e comparar os pesos dos recém nascidos nas ninhadas que receberam tratamentos diferentes, a saber, dosagem baixa e alta com o tratamento controle. Foram selecionadas fêmeas para cada tipo de tratamento, onde 3 das quais que receberam dose alta morreram, tornando o estudo assim desbalanceado. No estudo obteve-se um total de 322 filhotes, em que o mínimo por ninhada foi 2 e o máximo 18.

Toda a análise descritiva deste conjunto de dados foi realizada por Manghi (2011) e foi verificado que o menor peso observado foi de 3,68g para o sexo feminino na qual recebeu dosagem controle, e o maior peso, 8,33g, referente ao sexo masculino que também recebeu dosagem controle de tratamento, sendo essas duas observações pertencentes à ninhada 6. A mediana dos pesos foram maiores para o sexo masculino para as dosagens controle e baixa, e maior para o sexo feminino para a dosagem alta. Verificou por meio do coeficiente de correlação intraclasse que ratos de uma mesma ninhada apresentam correlação, devido ao parentesco destes com a fêmea que os gerou. Portanto, considerou-se um efeito aleatório para cada ninhada, assim como efeitos fixos para analisar variações entre tratamentos. West, Welch e Galecki (2014) testaram a hipóteses de que as variâncias dos pesos dos filhotes que receberam dosagens alta e baixa são iguais. Para uma melhor apresentação das variáveis Manghi (2011) elaborou a Tabela 1 que mostra a descrição de cada uma das variáveis de nível 1 (ratos) e de nível 2 (ninhadas).

Tabela 1 – Descrição das variáveis pertencentes aos experimentos dos ratos

Variáveis do nível 1 (ratos)	
Pup Id:	número de identificação dos ratos recém nascidos;
Weight:	peso (em gramas) dos ratos recém nascidos;
Sex:	sexo dos ratos recém nascidos.
Variáveis do nível 2 (ninhadas)	
Litter:	número de identificação da ninhada;
Treatment:	dosagem do tratamento (controle, baixa e alta);
Litsize:	número de filhotes de ratos observados por ninhada.

Fonte: Adaptado de Manghi (2011).

Os passos para a obtenção do modelo é dado por West, Welch e Galecki (2014) que segue uma estrutura dada como:

Equação de nível 1:

$$Weight_{jk} = \gamma_{1k} + \gamma_{2k}Sex_{jk} + \varepsilon_{jk}$$

Equação de nível 2:

(4.1)

$$\gamma_{1k} = \beta_{11} + \beta_{12}Treat1_k + \beta_{13}Treat2_k + \beta_{14}Litsize_k + b_{1k};$$

$$\gamma_{2k} = \beta_{21},$$

para $j = 1, \dots, n_k$ e $k = 1, \dots, 27$ em que, n_k é o número de filhotes que fazem parte da k -ésima. Tem-se que $Weight_{jk}$ é o peso em gramas do j -ésimo filhote pertencente à k -ésima ninhada, Sex_{jk} (1, se o j -ésimo filhote da k -ésima ninhada é fêmea e 0, caso contrário), $Treat1_k$ (1, se a k -ésima ninhada receber a dosagem baixa ou 0, em outros casos) e $Treat2_k$ (1, se a k -ésima ninhada recebeu dosagem alta e 0, caso contrário). Para $Litsize_k$ tem-se o número de filhotes referentes à k -ésima ninhada. Com relação à equação do nível 1, γ_{1k} é o intercepto para a k -ésima ninhada, γ_{2k} é o parâmetro associado com o sexo dos filhotes para a k -ésima ninhada e ε_{jk} é o erro associado com o peso, e é assumido que:

$$\varepsilon_{jk} \stackrel{iid}{\sim} \begin{cases} N(0, \sigma_{CT}^2), & \text{se a } k\text{-ésima ninhada recebeu dosagem controle} \\ N(0, \sigma_{HL}^2), & \text{se a } k\text{-ésima ninhada recebeu dosagem alta ou baixa} \end{cases}$$

O erros relacionados ao nível 2, $b_{1k} \stackrel{iid}{\sim} N(0, \sigma_L^2)$, são independentes entre si e entre ε_{jk} . Para a equação de nível dois, tem-se o β_{11} como média geral para os interceptos e efeito em γ_{1k} para ninhadas sem filhotes e que recebeu tratamento controle. Já β_{12} representa o efeito no intercepto γ_{1k} para a ninhada que recebeu dosagem alta de tratamento, β_{13} o efeito em γ_{1k}

para ninhadas com dosagem de tratamento baixa, β_{14} é o parâmetro associado ao tamanho da ninhada. Tem-se também que b_{1k} é um efeito aleatório associado a γ_{1k} . O parâmetro β_{12} na segunda equação de nível 2 indica que o efeito da variável *Sex* na média do peso dos filhotes é o mesmo para todas as ninhadas. Agora, colocando a equação de nível 2 dentro da equação de nível 1, pode-se obter o seguinte modelo:

$$Weight_{jk} = \beta_{11} + \beta_{12}Treat1_k + \beta_{13}Treat2_k + \beta_{14}Litsize_k + \beta_{21}Sex_{jk} + b_{1k} + \varepsilon_{jk},$$

com $j = 1, \dots, n_k$, ou seja, n_k é o número de filhotes da ninhada k -ésima ninhada, e $k = 1, \dots, 27$. Com este modelo completo β_{11} é o intercepto do modelo, β_{12} é o efeito fixo no valor esperado dos pesos dos filhotes nas ninhadas que receberam dose alta no tratamento, enquanto que β_{13} é o efeito fixo do peso médio dos filhotes que receberam dose baixa no tratamento, ao passo que β_{14} representa o efeito fixo do tamanho da ninhada, e, β_{21} é o efeito fixo nos pesos esperados dos filhotes de ratos fêmeas. Ainda assim, tem-se que os b_{1k} 's são efeitos aleatórios das ninhadas nos pesos dos ratos recém-nascidos enquanto que ε_{jk} 's são erros aleatórios associados aos valores dos respectivos pesos considerados neste estudo.

Em West, Welch e Galecki (2014) testa-se a hipótese em que as variâncias dos pesos dos filhotes cujas ninhadas receberam tratamento alto e baixo são iguais e, para que as que receberam tratamento controle, são diferentes. Na Tabela 2, temos as estimativas dos parâmetros para efeitos fixos obtidas pelo método de máxima verossimilhança para o modelo normal, t-Student e Exponencial Potência, estas também fornecidas por Manghi (2011), vemos que todas as estimativas são próximas para os três modelos e foram estatisticamente significantes ao nível de 5% de significância. Também foi calculado o grau de liberdade para o modelo t-Student que foi selecionado pelo critério de Akaike (AIC) com $\hat{\eta} = 5$, ao passo que para o modelo exponencial potência o parâmetro de forma foi selecionado pelo mesmo critério e obtém-se $\hat{\eta} = 0.6$, foi realizado a distância de Mahalanobis estimada para a distribuição normal em que as ninhadas 5, 6 e 18 apresentaram valores maiores, na distância de Mahalanobis modificada para a distribuição t-Student e exponencial potência, a ninhada com maior ênfase foi a 6, então o autor concluiu que o conjunto de valores de pesos para essas ninhadas podem ser consideradas aberrantes.

Logo, para esta aplicação vamos utilizar o mesmo modelo dado na Equação 4.1, que segue a distribuição conjunta dada por:

$$\begin{bmatrix} \mathbf{Y}_k \\ \mathbf{b}_k \end{bmatrix} \sim El_{k_1} \left(\begin{bmatrix} \mathbf{X}_k \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_k \boldsymbol{\Psi}_k \mathbf{Z}_k^\top + R_k & \mathbf{Z}_k \boldsymbol{\Psi}_k \\ \boldsymbol{\Psi}_k \mathbf{Z}_k^\top & \boldsymbol{\Psi}_k \end{bmatrix} \right), \quad (4.2)$$

este modelo foi considerado por Manghi, Paula e Cysneiros (2016).

Para prosseguir com o estudo nessa base de dados vamos aplicar os esquemas de perturbação aqui propostos e a análise de influência local proposta no Capítulo 3.

Tabela 2 – Estimativas dos parâmetros com seus erros-padrão aproximados para a distribuição normal, t-Student e exponencial potência segundo o modelo multinível ajustado ao banco de dados dos filhotes de ratos.

Parâmetro	Normal	t-Student	Exponencial Potência
β_{11}	8,3280 (0,2593)	8,1613 (0,2508)	8,2365 (0,1914)
β_{12}	-0,8648 (0,1715)	-0,9050 (0,1652)	-0,8496 (0,1524)
β_{13}	-0,4344 (0,1424)	-0,4444 (0,1359)	-0,4316 (0,1378)
β_{14}	-0,1307 (0,0175)	-0,1200 (0,0171)	-0,1258 (0,0131)
β_{21}	-0,3419 (0,0422)	-0,3004 (0,0375)	-0,3154 (0,0408)
σ_N^2	0,0832 (0,0265)	0,0732 (0,0256)	0,0068 (0,0022)
σ_C^2	0,2639 (0,0339)	0,1166 (0,0289)	0,0157 (0,0026)
σ_{HL}^2	0,0913 (0,0098)	0,0782 (0,0157)	0,0073 (0,0010)

Fonte: Adaptada de Manghi, Paula e Cysneiros (2016).

Com a realização da perturbação de caso na distribuição normal, como visto no Gráfico 1 observa-se que a ninhada possivelmente influente é a 6, no Gráfico 2 as observações que se destacaram foram as ninhadas 6 e 18, na perturbação na resposta, o Gráfico 3 mostra que os valores discrepantes estão na observação 66 da ninhada 6 e 227 da ninhada 18.

Na Tabela 3 mostra a variação relativa percentual em valor absoluto das novas estimativas sem as ninhadas 6, 18, 6 e 18 e sem as observações 66 e 227 juntas. Sem a ninhada 6, todos as estimativas para o vetor de efeitos fixos β foram significativas ao nível de $\alpha = 5\%$, mas, nota-se que houve uma diferença notável na variância do tratamento controle. Quando se retirou a ninhada 18, todos os coeficientes também foram significativos ao nível de 5%. Nota-se que há uma maior variação em σ_C^2 , 40,74%, quando se retira a ninhada 6, que é justamente os filhotes que receberam tratamento controle, com a retirada da ninhada 18 o maior percentual de variação foi em σ_{HL}^2 com 12,68%, que não é considerado alto. Quando são retiradas as ninhadas 16 e 18 juntas, o percentual de variação para a estimativa de σ_C^2 fica em torno de 40,60%. Com a retirada das observações 66 e 227 juntas, o percentual de variação para σ_C^2 é cerca de 32,03%, o que nos leva a pensar que o que está influenciando na estimativa de σ_C^2 é em grande parte, devido à observação 66 que pertence à ninhada 6, esta observação é do sexo feminino, que recebeu tratamento controle e entre todas as observações é a que apresenta menor peso. No entanto, apesar de apresentar essas variações, nas estimativas dos parâmetros não houve troca de sinal e ainda se mostraram estatisticamente significantes ao nível de 5%.

Analisando a influência que cada grupo tem sobre o modelo, levando em conta o critério de $\frac{3x}{K}$, para a perturbação de caso e matriz de variância e covariância V_k esse valor é de 11,11%. Valores percentuais relativos dos parâmetros maiores que este valor pode levar à interpretação de que as observações avaliadas podem influenciar nas estimativas dos parâmetros. Desta forma, temos que quando retiramos a ninhada 6 houve variação notável para σ_C^2 . Sem a ninhada 18 indica que houve influência para a estimação de β_{13} e para σ_{HL}^2 . Com a retirada das ninhadas 6 e 18 juntas há variação relativa percentual considerável para as estimativas de β_{13} , β_{21} , σ_C^2 e σ_{HL}^2 . Entretanto, quando retiradas as observações 66 e 227 juntas, temos uma variação percentual relativa maior que $\frac{3x}{N} = 1,86\%$, para todas as estimativas dos parâmetros, exceto para β_{11} .

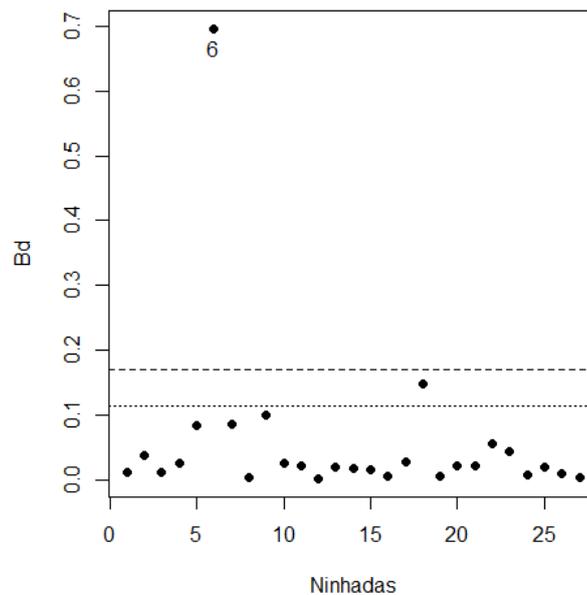


Gráfico 1 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.

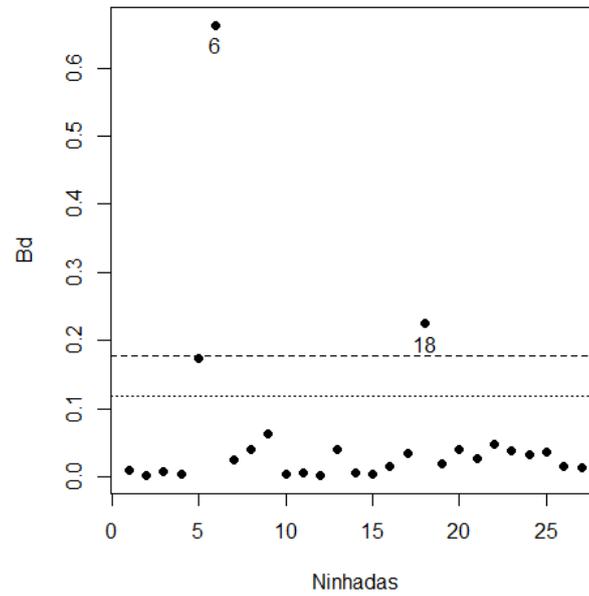


Gráfico 2 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.

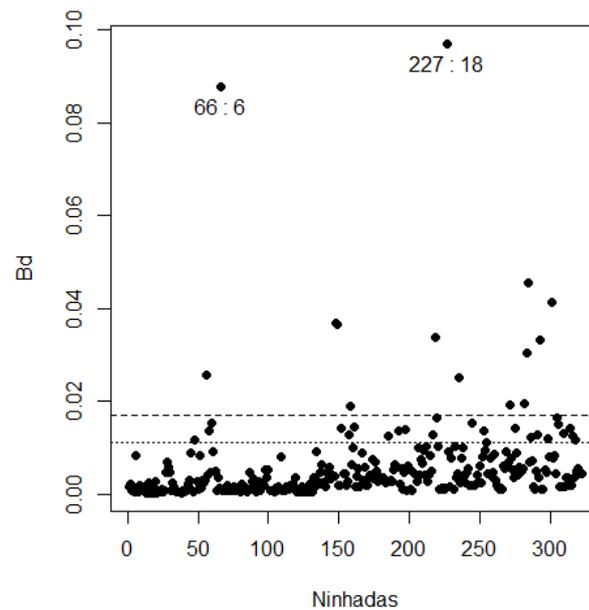


Gráfico 3 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição normal aplicado ao banco de dados dos filhotes de ratos.

Tabela 3 – Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhadas 6, 18, 6 e 18 juntas e sem as observações 66 e 227 juntas) para a distribuição normal, segundo os modelos ajustados ao banco de dados dos filhotes de ratos.

Parâmetro	Sem ninhada 6		Sem ninhada 18	
	Estimativa	Variação %	Estimativa	Variação %
β_{11}	8,3213	0,08	8,3635	0,43
β_{12}	-0,8696	0,56	-0,8820	1,99
β_{13}	-0,4415	1,63	-0,4971	14,43
β_{14}	-0,1304	0,23	-0,1342	2,68
β_{21}	-0,3248	5,00	-0,3123	8,66
σ_N^2	0,0832	5,04	0,0750	9,79
σ_C^2	0,2639	40,74	0,2656	0,64
σ_{HL}^2	0,0913	0,02	0,0797	12,68
Parâmetro	Sem ninhada 6 e 18		Sem obs. 66 e 227	
	Estimativa	Variação %	Estimativa	Variação %
β_{11}	8,3588	0,37	8,3967	0,82
β_{12}	-0,8870	2,57	-0,9112	5,36
β_{13}	-0,5040	16,01	-0,4633	6,65
β_{14}	-0,1338	2,39	-0,1338	2,37
β_{21}	-0,2990	12,53	-0,3247	5,02
σ_N^2	0,0799	3,92	0,0906	8,90
σ_C^2	0,1568	40,60	0,1794	32,03
σ_{HL}^2	0,0796	12,79	0,0809	11,40

Fonte: O autor (2019).

O Gráfico 4 trata sobre perturbação de caso para a distribuição t-Student, nenhuma ninhada se mostra em destaque. No Gráfico 5 para a perturbação na matriz de variância vemos uma notoriedade maior para a ninhada 6, e por último, no Gráfico 6 para a perturbação na variável resposta, as observações de mais realce foram a 109 da ninhada 9 e 227 da ninhada 18.

Assim, com relação às perturbações feitas sob a distribuição t-Student, a Tabela 4 apresenta as novas estimativas, primeiro sem a ninhada 6, na qual foi identificada como discrepante em relação às demais quando se realizou a perturbação na matriz de variância e covariância. Na mesma tabela, seguem as estimativas dos parâmetros após a retirada das observações 109 e 227, em que se destacaram diante das demais quando foi feita a perturbação na variável resposta. São fornecidas as estimativas dos parâmetros e variação relativa percentual em valor absoluto. Tanto para a retirada da ninhada 6 quanto para as observações 109 e 227, não houve variação relativa absoluta de grande destaque, sendo a maior em torno de 11,6%. As estimativas de efeito fixo não mudaram de sinal e foram significativas ao nível de 5%.

Analisando a influência quando retirada as ninhadas e observações no modelo relativas à Tabela 4 e comparando com as variações relativas da mesma, temos que sem a ninhada 6, não há nenhuma variação que seja maior que a medida de 11,11%. Sem as observações 109 e 227 juntas temos que as variações para as estimativas de β_{12} , β_{13} , σ_N^2 , σ_C^2 e σ_{HL}^2 são maiores que 1,86%, o que evidencia que essas observações exerceram alguma influência nas estimativas, mesmo que não tenha havido mudanças inferenciais.

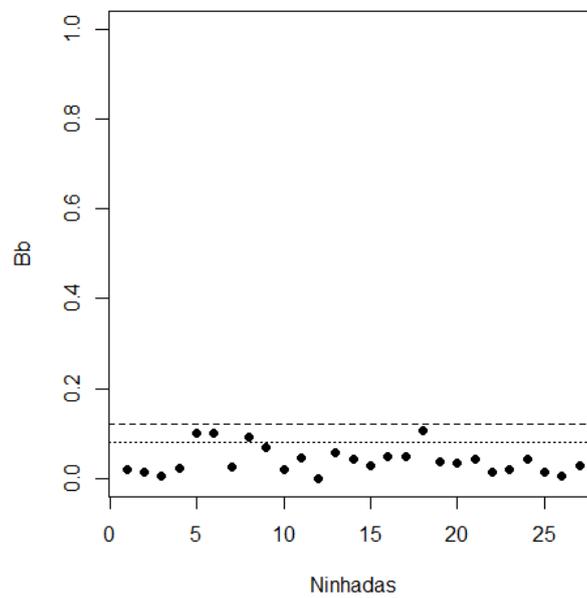


Gráfico 4 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição t-Student com o banco de dados dos filhotes de ratos

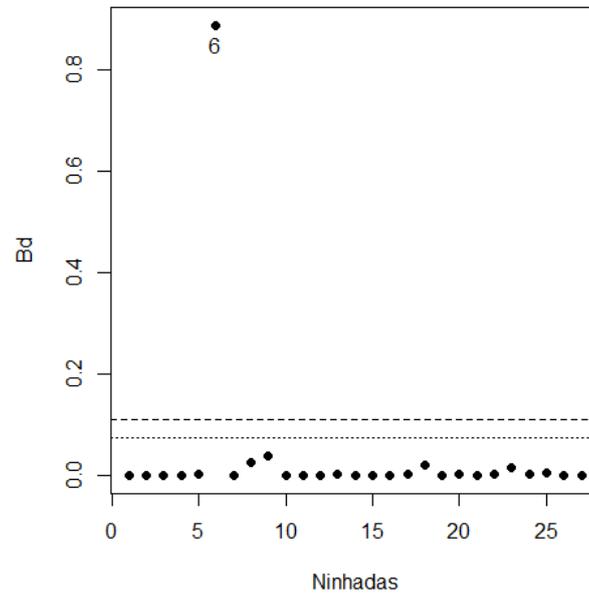


Gráfico 5 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição t-Student aplicado ao banco de dados dos filhotes de ratos.

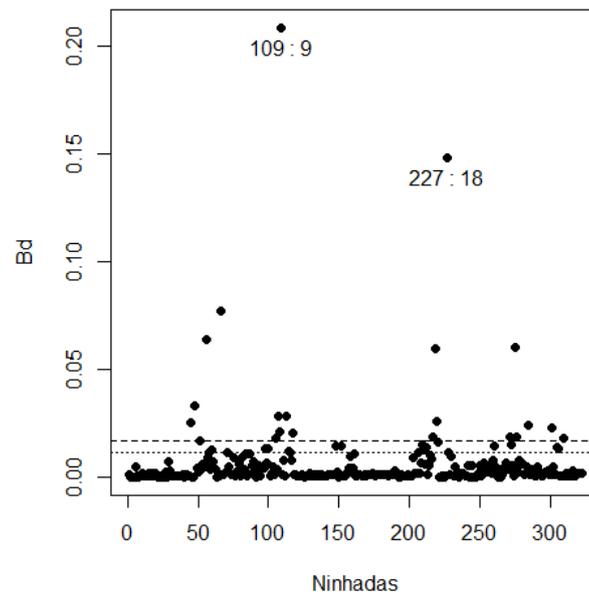


Gráfico 6 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição t-Student aplicado ao banco de dados dos filhotes de ratos.

Tabela 4 – Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhada 6 e sem as observações 109 e 227), para a distribuição t-Student, segundo os modelos ajustados ao banco de dados dos filhotes de ratos.

Parâmetro	Sem ninhada 6		Sem obs. 109 e 227	
	Estimativa	Variação %	Estimativa	Variação %
β_{11}	8,1840	0,28	8,1587	0,03
β_{12}	-0,8929	1,33	-0,8820	2,54
β_{13}	-0,4415	0,65	-0,3928	11,60
β_{14}	-0,1220	1,69	-0,1218	1,54
β_{21}	-0,2991	0,41	-0,3046	1,40
σ_N^2	0,0761	3,90	0,0817	11,63
σ_C^2	0,1037	11,04	0,1069	8,31
σ_{HL}^2	0,0795	1,76	0,0751	3,95

Fonte: O autor (2019).

No Gráfico 7 temos a perturbação de caso e no Gráfico 8 a perturbação na matriz de variância e covariância, para esses dois casos a ninhada que ficou mais evidente foi a 6, portando essa foi retirada para se realizar novas estimativas. Na perturbação na variável resposta, como mostra o Gráfico 9, as observações que mais se destacaram foram a 66 da ninhada 6, que apresentou menor peso e esta sendo tratamento controle, e a observação 275 da ninhada 22, na qual era o menor peso com dosagem de tratamento alta.

Na Tabela 5 são apresentadas as estimativas dos parâmetros e variação relativa percentual em valor absoluto quando foram retiradas a ninhada 6 e a observação 66 e 75, estas evidenciadas quando feitos os esquemas de perturbação sob a distribuição Exponencial Potência. Para as novas estimativas dos parâmetros de efeito fixo não houve grande variação, mas nas estimativas para as variâncias os valores das variações mostraram-se muito altos, mas que apesar disso não afetou as estimativas dos parâmetros de efeito fixo, ou seja, não houve mudança de sinal e todas as estimativas continuaram significativas ao nível de 5%.

Analisando as influências das ninhadas e observações sobre as estimativas, temos na Tabela 5 que sem a ninhada 6 as variações relativas percentuais para as variâncias são bem maiores de 11,11%, desta forma há uma variação considerável nestas estimativas. Sem as observações 66 e 275 juntas, as variações relativas percentuais para as estimativas de β_{12} , β_{13} , β_{14} , σ_N^2 , σ_C^2 e σ_{HL}^2 foram maiores que 1,86%, o que nos indica que as observações 66 e 275 juntas exercem influência significativa nas estimativas dos parâmetros.

Logo, notou-se que para este conjunto de dados a distribuição que sofreu menos impacto foi a t-Student, pois diante dos esquemas de perturbação, apenas na matriz V_k a ninhada

6 se mostrou mais aberrante, e para a perturbação na resposta poucos foram os dados que se mostraram potencialmente influentes, com destaque apenas para as observações 109 e 227. Com as novas estimativas dos parâmetros sem estas observações influentes não foi suficiente para mudar a inferência das estimativas. Desta forma, temos que para este conjunto de dados a distribuição t-Student mostrou-se superior à distribuição Normal.

Tabela 5 – Estimativas dos parâmetros e variação relativa percentual em valor absoluto (sem ninhada 6 e sem as observações 66 e 275 juntas) para a distribuição exponencial potência, segundo os modelos ajustados ao banco de dados dos filhotes de ratos

Parâmetro	Sem ninhada 6		Sem obs. 66 e 275	
	Estimativa	Variação %	Estimativa	Variação %
β_{11}	8,2496	0,16	8,3450	1,32
β_{12}	-0,8519	0,26	-0,9097	7,07
β_{13}	-0,4338	0,49	-0,4666	8,09
β_{14}	-0,1268	0,79	-0,1313	4,36
β_{21}	-0,3096	1,86	-0,3146	0,27
σ_N^2	0,0174	157,98	0,0357	428,14
σ_C^2	0,0273	73,10	0,0666	322,97
σ_{HL}^2	0,0186	154,08	0,0363	394,74

Fonte: O autor (2019).

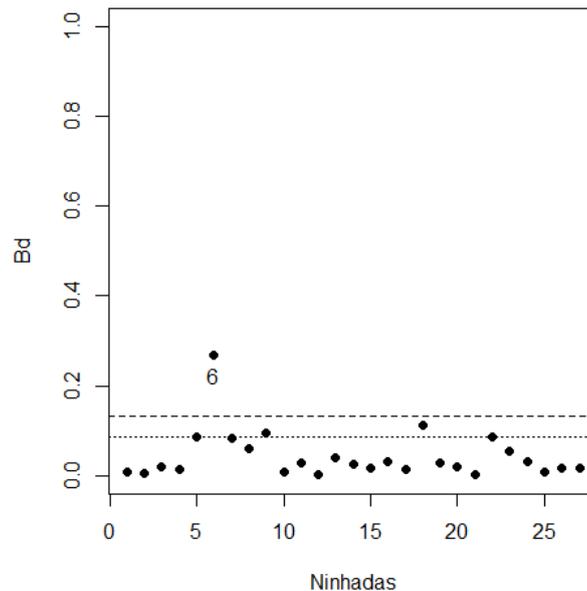


Gráfico 7 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos

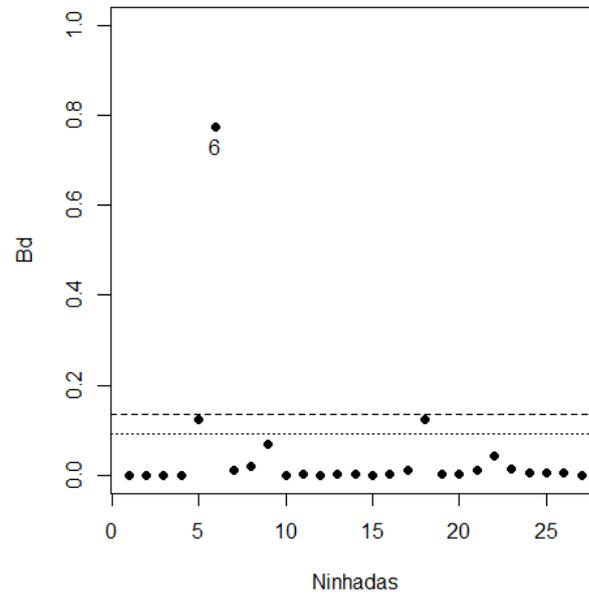


Gráfico 8 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos.

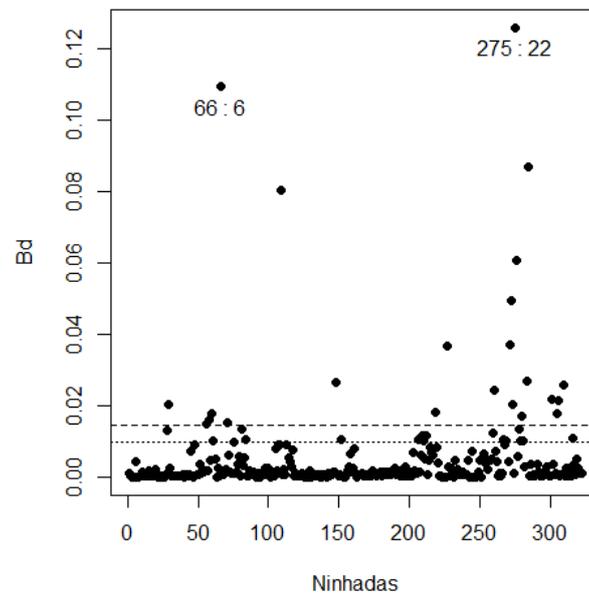


Gráfico 9 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta para a distribuição exponencial potência aplicado ao banco de dados dos filhotes de ratos.

4.2 DADOS DO PRODUTO INTERNO BRUTO

Neste segundo exemplo vamos usar dados do produto interno bruto que foram obtidos de Munnell (1990), caracterizando dados hierárquicos de três níveis. Este são compostos pelo produto interno bruto de 48 estados divididos por 9 regiões dos Estados Unidos da América, em que foram analisados durante os anos de 1970 a 1986, o que leva a uma dependência temporal, totalizando 816 observações. Neste exemplo foi investigado a produtividade do capital público que foi usado como variáveis o capital do serviço público de água e o ano de registro de dados, na qual para cada um dos 48 estados tem-se 17 anos de registro, que tem como resposta o logaritmo natural do produto interno bruto $Lgsp$. Desta forma, este estudo são dados longitudinais com três níveis de hierarquia: ano, estado e Região, ver Greene (2003) para mais detalhes.

Este banco de dados também foi trabalhado por Manghi, Paula e Cysneiros (2016) onde foi realizada estimativa de máxima verossimilhança dos parâmetros, correlação intraclasse entre regiões e entre os estados. Obtiveram também a distância de Mahalanobis para a distribuição t-Student e exponencial potência em que não apresentou nenhuma Região de destaque sob a distribuição Student-t, enquanto que a Região 8 foi apontada como possível observação aberrante sob a distribuição exponencial de potência. Foi estimado também os graus de liberdade para a distribuição t-Student sendo $\hat{\eta} = 12$ e parâmetro de forma para a exponencial potência de $\hat{\eta} = 0.9$, sendo utilizado o critério de Akaike (AIC) para a seleção. Pesos foram estimados de acordo com a Região sob os modelos multiníveis t-Student e exponencial de potência com os dados do produto do produto interno bruto. Apresentou-se peso menor para a Região 8 sob o modelo exponencial de potência, aplicando ao nível 3, ou seja, no nível Região, muitas das vezes esse peso pequeno significa que a observação mostra-se aberrante, porém pode não ser potencialmente influente.

Então, o modelo multinível considerado por esses autores foi:

Equação de nível 1:

$$Lgsp_{ijk} = \gamma_{1j|k} + \gamma_{2j|k}Lwater_{ijk} + \gamma_{3j|k}Year_{ijk}\epsilon_{ijk}$$

Equação de nível 2:

$$\gamma_{k|j} = \delta_{11k} + \xi_{1j|k}; \gamma_{2j|k} = \delta_{21k}; \gamma_{3j|k} = \delta_{31k};$$

(4.3)

Equação de nível 3:

$$\delta_{11k} = \beta_{111} + \vartheta_{11k}; \delta_{21k} = \beta_{211}; \delta_{31k} = \beta_{311},$$

em que $i = 1, \dots, 17$, $j = 1, 2, \dots, n_k$ e $k = 1, \dots, 9$, n_k é o número de estados da k -ésima Região.

Temos que $Lgsp_{ijk}$ é o logaritmo natural do produto interno bruto do estado para o i – étimo ano do j –ésimo estado da k –ésima Região, enquanto que $Lwater_{ijk}$ é o logaritmo natural do capital da utilidade da água no i –ésimo ano do j –ésimo estado da k –ésima Região. Tem-se que $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$, $\xi_{1j|k} \stackrel{iid}{\sim} N(0, \sigma_S^2)$, $\vartheta_{11k} \stackrel{iid}{\sim} N(0, \sigma_R^2)$, e cada um dos ε_{ijk} 's, $\xi_{1j|k}$'s e ϑ_{11k} 's são independentes entre si.

Na equação de nível 1 do modelo temos $\gamma_{1j|k}$ como o intercepto do j –ésimo estado da k –ésima Região, $\gamma_{2j|k}$ é o coeficiente relacionado à variável $Lwater$ que pertence ao j –ésimo estado da k –ésima Região, $\gamma_{3j|k}$ é o coeficiente associado a variável ano para o j –ésimo estado da k –ésima Região e o erro aleatório ε_{ijk} que está associado com a variável resposta. Adicionalmente, para a equação de nível 2 tem-se δ_{11k} que é a média de $\gamma_{1j|k}$ na k –ésima Região, δ_{21k} indica que o efeito de $Lwater$ em $Lgsp$ é o mesmo para todos estado na k –ésima Região e δ_{31k} tem a mesma interpretação, ou seja, que o ano considerado tem o mesmo efeito para todos os estado em $Lgsp$. Para o nível 3 temos que β_{111} é o valor esperado do efeito aleatório δ_{11k} e β_{211} aponta que o efeito de $Lwater$ é o mesmo para todos os estados e para todas as regiões e com mesma interpretação temos β_{311} mas em relação ao ano considerado. Então podemos reescrever o modelo de nível 3 definido na Equação 4.3 da seguinte maneira

$$Lgsp_{ijk} = \beta_{111} + \beta_{211}Lwater_{ijk} + \beta_{311}Year_{ijk} + \vartheta_{11k} + \xi_{1j|k} + \varepsilon_{ijk} \quad (4.4)$$

em que β_{111} é o intercepto do modelo, ou seja, o valor esperado de $Lgsp$ quando as variáveis $Lwater$ e $Year$ foram nulas. Já β_{211} compreende o efeito fixo do logaritmo da água $Lwater$ em $Lgsp$ e β_{311} é o efeito fixo do ano $Year$ em $Lgsp$. Temos também os efeitos aleatórios com ϑ_{11k} sendo o efeito aleatório para a k –ésima Região, enquanto que $\xi_{1j|k}$ é o efeito aleatório para o j –ésimo estado pertencente à k –ésima Região e por ultimo o erro aleatório ε_{ijk} que se refere à variável resposta $Lgsp_{ijk}$.

Vamos usar o mesmo modelo usado por Manghi, Paula e Cysneiros (2016) para as análises, em que a distribuição conjunta é dada como segue:

$$\begin{bmatrix} \mathbf{Y}_k \\ \mathbf{b}_k \end{bmatrix} \sim El_{k_2} \left(\begin{bmatrix} \mathbf{X}_k \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_k \boldsymbol{\Psi}_k \mathbf{Z}_k^\top + R_k & \mathbf{Z}_k \boldsymbol{\Psi}_k \\ \boldsymbol{\Psi}_k \mathbf{Z}_k^\top & \boldsymbol{\Psi}_k \end{bmatrix} \right), \quad (4.5)$$

com $k = 1, 2, \dots, 9$, e $k_2 = 17n_k + 1 + n_k$. Temos ainda que $\mathbf{Y}_k = (\mathbf{Y}_{1k}^\top, \mathbf{Y}_{2k}^\top, \dots, \mathbf{Y}_{n_k k}^\top)^\top$ e $\mathbf{Y}_{jk} = (Lgsp_{1jk}, Lgsp_{2jk}, \dots, Lgsp_{17jk})^\top$, para $j = 1, 2, \dots, n_k$. Ainda, \mathbf{X}_k é uma matriz $17n_k \times 3$ com a primeira coluna formada por 1 e as demais colunas são as variáveis explanatórias $Lwater$ e $Year$ para a k –ésima Região, \mathbf{Z}_k é uma matriz $17n_k \times (1 + n_k)$ com a primeira coluna formada por $17n_k$ número 1 e $17(n_k - 1)$ de zero. Temos ainda o vetor de parâmetros $\boldsymbol{\beta} = (\beta_{111}, \beta_{211}, \beta_{311})^\top$

e $\mathbf{b}_k = (\vartheta_{11k}, \xi_{11|k}, \xi_{12|k}, \dots, \xi_{1n_k|k})^\top$ vetor de efeitos aleatórios. Ainda, $\Psi_k = \text{diag}\{\sigma_R^2, \sigma_S \mathbf{I}_{n_k}\}$ e $\mathbf{R}_k = \sigma^2 \mathbf{I}_{17n_k}$, em que \mathbf{I}_{n_k} é uma matriz identidade de ordem n_k .

Temos as medidas descritivas para a variável resposta, *Lgsp*, dadas na Tabela 6. Observa-se que o valor mínimo encontra-se na Região 1, e máximo na Região 8 que também obteve maior desvio padrão. Apresenta a menor mediana na Região 1 e menor média na Região 8. Para a variável resposta contínua *Lwater*, temos as medidas descritivas apresentadas na Tabela 7. Os menores valores do mínimo, mediana e média estão na Região 8, o valor máximo obtido encontra-se na Região 9 assim como maior desvio padrão.

Tabela 6 – Medidas descritivas para para a variável resposta *Lgsp* relativa aos Dados do Produto Interno Bruto

Região	Min.	Mediana	Média	Máx	Desvio Padrão
1	8,38	9,34	9,73	11,53	0,95
2	11,37	11,88	11,96	12,64	0,40
3	10,72	11,69	11,52	12,12	0,43
4	8,68	10,33	10,08	11,19	0,78
5	8,83	10,86	10,60	11,93	0,78
6	9,76	10,58	10,50	11,07	0,31
7	9,64	10,92	11,01	12,56	0,87
8	8,88	9,518	9,66	10,86	0,54
9	10,01	10,95	11,31	13,05	1,07

Fonte: O autor (2019).

Tabela 7 – Medidas descritivas para para a variável preditora *Lwater* relativa aos Dados do Produto Interno Bruto

Região	Min.	Mediana	Média	Máx	Desvio Padrão
1	5,54	6,53	6,84	8,43	0,96
2	8,28	9,00	9,09	9,84	0,52
3	7,94	8,99	8,75	9,30	0,47
4	5,56	7,33	7,19	8,50	0,92
5	5,44	7,97	7,67	9,12	0,95
6	6,72	7,65	7,60	8,42	0,48
7	6,47	7,85	7,87	9,60	0,96
8	5,43	6,45	6,64	8,38	0,80
9	7,44	8,47	8,69	10,11	0,96

Fonte: O autor (2019).

Para complementar as análises feitas a esse banco de dados, vamos aplicar a análise de influência local, que é a proposta deste trabalho. As estimativas para os parâmetros do modelo multinível da equação 4.4 e seus respectivos erros-padrão estão descritos na Tabela 8. Todas as estimativas foram significativas ao nível de 5% para os modelos normal, t-Student e

exponencial potência, estes resultados assim como os resultados da distância de Mahalanobis para as distribuições Normal, t-Student e Exponencial Potência foram realizados em Manghi, Paula e Cysneiros (2016).

Para a distribuição normal, temos os quatro esquemas de perturbação propostos: perturbação de caso, perturbação na matriz de variância e covariância, perturbação na variável explanatória contínua de nível 1 e perturbação na variável resposta. Como visto nos Gráficos 10 e 11 (perturbação de caso e na matriz V_k) respectivamente, a Região 8 foi a mais notória. Já para a perturbação na variável preditora contínua de nível 1 *Lwater* e para a variável resposta, (Gráficos 12 e 13), a observação 17 do estado 4 da Região 1 foi a que se mostrou mais evidente para ambos esquemas de perturbação.

Tabela 8 – Estimativas dos parâmetros com seus erros-padrão aproximados para a distribuição normal, t-Student e exponencial potência, segundo os modelos ajustados ao banco de dados do produto interno bruto

Parâmetro	Normal	t-Student	Exponencial Potência
β_{111}	-29,6200 (1,9399)	-29,7721 (1,7810)	-29,4274 (1,9279)
β_{211}	0,2476 (0,0288)	0,2391 (0,0265)	0,2474 (0,0288)
β_{311}	0,0194 (0,0011)	0,0195 (0,0010)	0,0193 (0,0011)
σ_R^2	0,2392 (0,1449)	0,1645 (0,1128)	0,1278 (0,0779)
σ_S^2	0,3244 (0,0735)	0,3237 (0,0866)	0,1768 (0,0402)
σ^2	0,0052 (0,0003)	0,0043 (0,0006)	0,0027 (0,0004)

Fonte: Adaptada de Manghi, Paula e Cysneiros (2016).

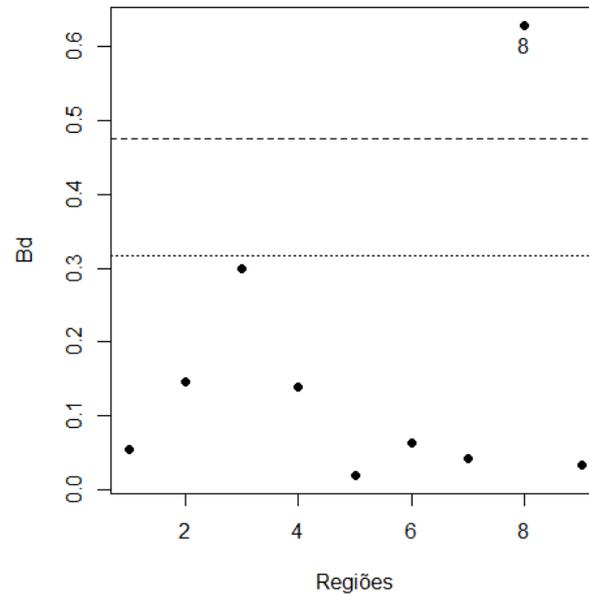


Gráfico 10 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição normal aplicado ao banco de dados do produto interno bruto.

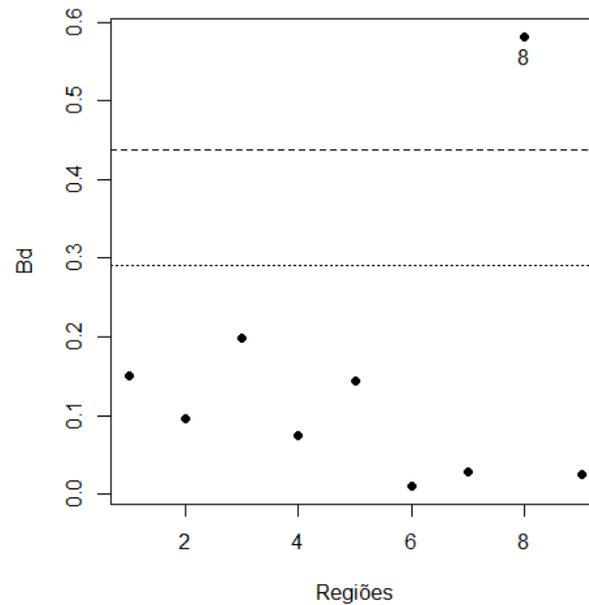


Gráfico 11 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição normal aplicado ao banco de dados do produto interno bruto.

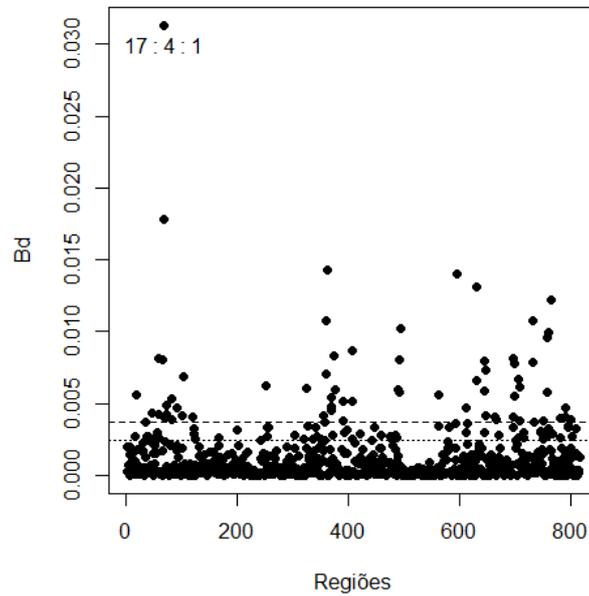


Gráfico 12 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 L_{water} , para a distribuição normal aplicado ao banco de dados do produto interno bruto.

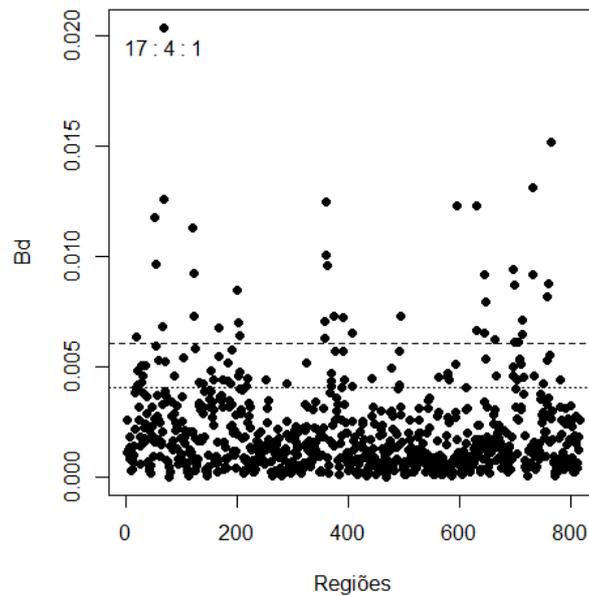


Gráfico 13 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição normal aplicado ao banco de dados do produto interno bruto.

Para a distribuição t-Student os quatro esquemas de perturbação são aplicados aos

dados do produto interno bruto, e temos que para a perturbação de caso, como indicado no Gráfico 14, nenhuma Região se destacou, já na perturbação na matriz de variância e covariância V_k , visto no Gráfico 15, a Região 8 se destacou de maneira bem evidente. Com a perturbação na variável preditora contínua de nível 1 $Lwater$ a observação que mais se sobressaiu foi a 17 do estado 4 da Região 1, (como mostra no Gráfico 16). Na perturbação da variável resposta a observação de mais destaque foi a observação do estado 1 da Região 8, visto no Gráfico 17.

Temos os esquemas de perturbação aplicados na distribuição exponencial potência e não houve regiões que se destacaram tanto na perturbação de caso como na matriz de variância e covariância, visto nos Gráficos 18 e 19 respectivamente. Na perturbação para a variável explicativa contínua de nível 1 $Lwater$, a observação mais evidente foi a 17 do estado 4 da Região 1, ver Gráfico 20. Para a perturbação na variável resposta, Gráfico 21, a observação de mais realce foi a 17 do estado 2 da Região 8.

Deste modo, prosseguimos com as análises retirando as regiões e observações que mais se destacaram quando colocadas sobre os esquemas de perturbação propostos.

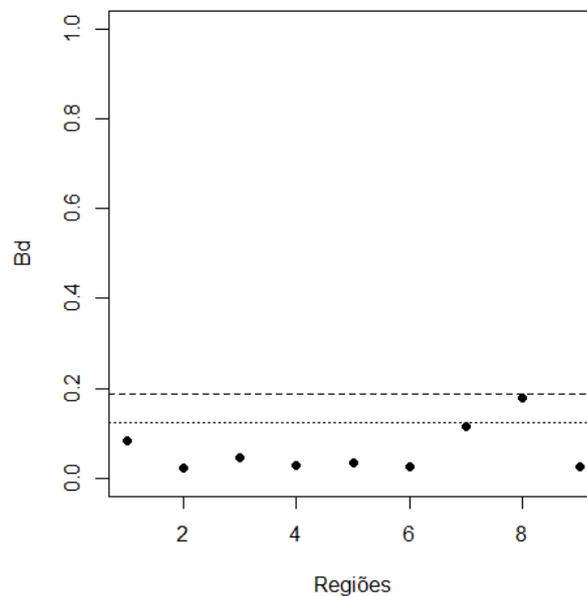


Gráfico 14 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.

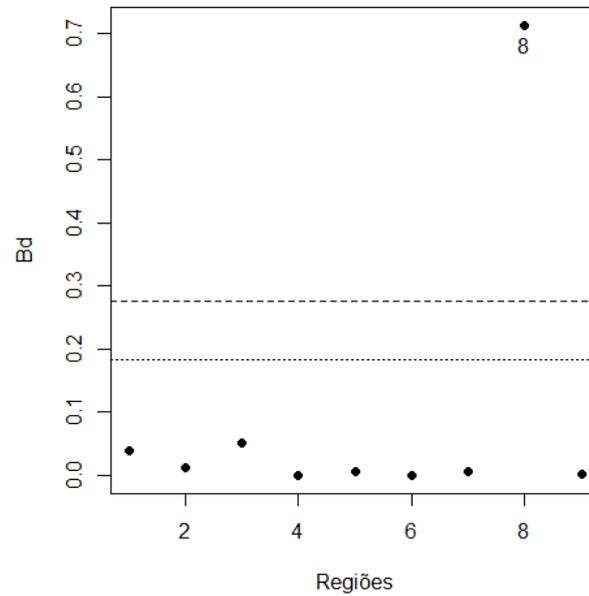


Gráfico 15 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.

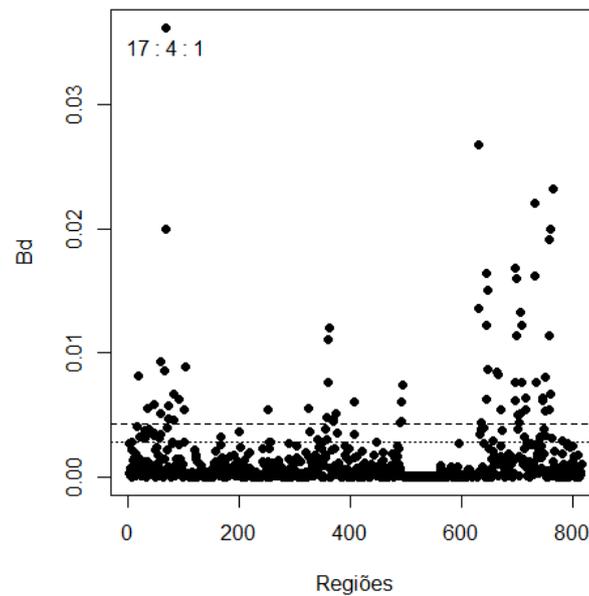


Gráfico 16 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 L_{water} , para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.

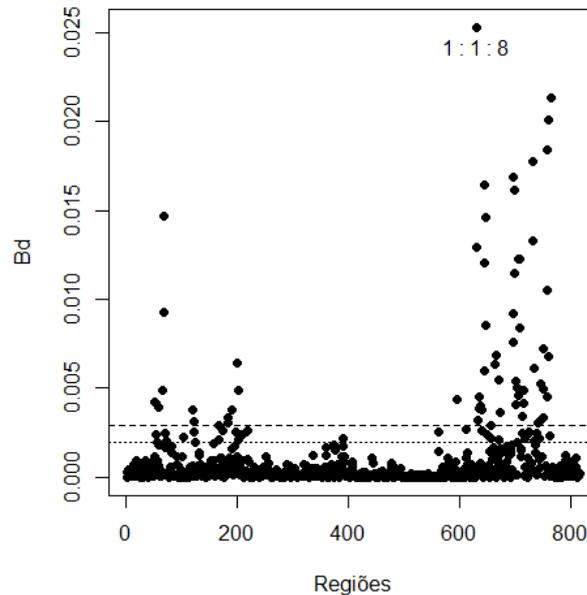


Gráfico 17 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição t-Student aplicado ao banco de dados do produto interno bruto.

A Tabela 9 mostra as novas estimativas para as distribuições normal e t-Student e variação relativa percentual em valor absoluto sem a Região 8, já que nestas distribuições esta Região se destacou. Com a retirada desta Região e realizado novas estimativas não houve um percentual de variação muito substancial, temos a maior variação em torno de 19% para $\hat{\sigma}_S^2$, que é o efeito aleatório para as estados. Também não houve mudança nas estimativas para os parâmetros de efeito fixo, sendo estes significativos para um nível de 5%. Avaliando a influência que cada Região teria nas estimativas usamos a medida $\frac{3x}{K}$, em que x é a quantidade de Regiões avaliadas, ou seja, neste caso queremos saber o quanto uma região pode afetar nas estimativas, e então obtemos 33,33%, que é o que se espera de contribuição de cada grupo. Desta forma, vemos que sem a Região 8 nas variações relativas percentuais da Tabela 9 nenhuma foi maior que este valor, então não obteve-se uma mudança significativa nas novas estimativas quando retirada esta Região.

A observação 17 da Região 4 pertencente à Região 1 foi evidenciada sob os esquemas de perturbação na variável explicativa para as três distribuições, logo, esta foi retirada do banco de dados e após feito novas estimativas. Como pode ser visto na Tabela 10. Não houve grande variação percentual, a maior foi de 4,78% para σ_R^2 , que é a variância correspondente à Região, esta sob a distribuição t-Student. Também não houve mudanças significativas para as estimativas dos

parâmetros β ao nível de 5%. Avaliando o impacto que esta observação tem sobre as estimativas, obtemos um valor de $\frac{3x}{N} = 0,36\%$, aqui x é quantidade de observações que serão retiradas e N o total de observações. Então as variações relativas percentuais para o modelo normal foi significativa para as estimativas de β_{211} , σ_R^2 , σ_S^2 e σ^2 . Para o modelo t-Student todas as variações relativas percentuais foram maiores que 0,36% e para o modelo Exponencial Potência variações mais significativas foram para as estimativas de β_{111} , β_{211} , β_{311} , σ_R^2 e σ_S^2 . Lembrando que esta observação mesmo tendo certa influência para a estimação dos parâmetros, com a sua ausência não houve mudanças inferenciais nas novas estimativas.

Na Tabela 11 temos os resultados para as novas estimativas sem a observação 1 do estado 18 da Região 8, já que esta se destacou sob a perturbação na variável resposta aplicado na distribuição t-Student, a maior variação foi de 1,09%, logo não foi potencialmente suficiente para influenciar nas estimativas. Na mesma Tabela, temos as novas estimativas para a distribuição exponencial potência quando retirada a observação 17 do estado 2 da Região 6. Observou-se 41,73% de variação para σ_R^2 , que é a variância para as regiões. Apesar de esta estimativa ter um percentual de variação considerado alto, todas as estimativas para os parâmetros foram estatisticamente significativos ao nível 5%.

Analisando a influência que se espera que cada observação tenha sobre as estimativas, de acordo com a Tabela 11 temos que para as novas estimativas sob o modelo t-Student apenas as variações relativas percentuais de β_{311} e σ_R^2 são maiores que 0,36%, e para o modelo Exponencial Potência todas as variações percentuais foram maiores que 0,36%, exceto para σ^2 . Esta análise nos indica que estas observações causa uma pequena mudança nas novas estimativas, no entanto, não provocam mudanças inferenciais.

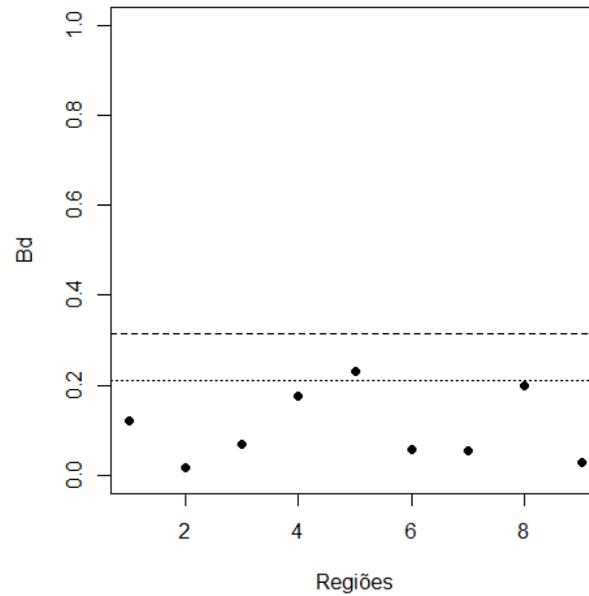


Gráfico 18 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação de caso para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.

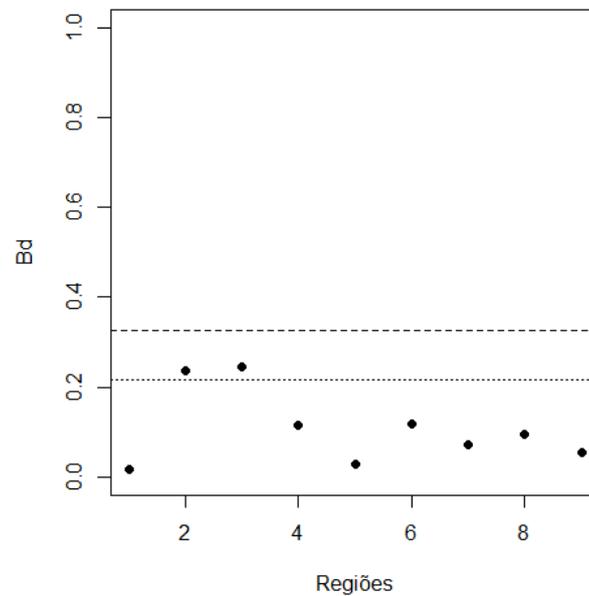


Gráfico 19 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na matriz de variância e covariância V_k para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.

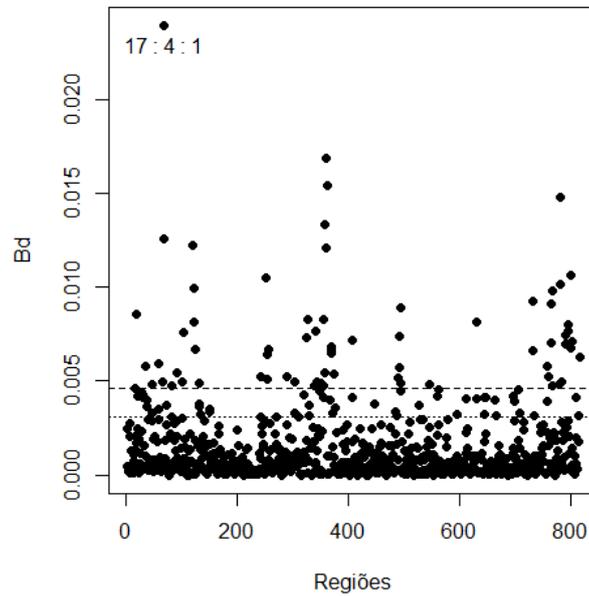


Gráfico 20 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável explicativa contínua de nível 1 L_{water} , para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.

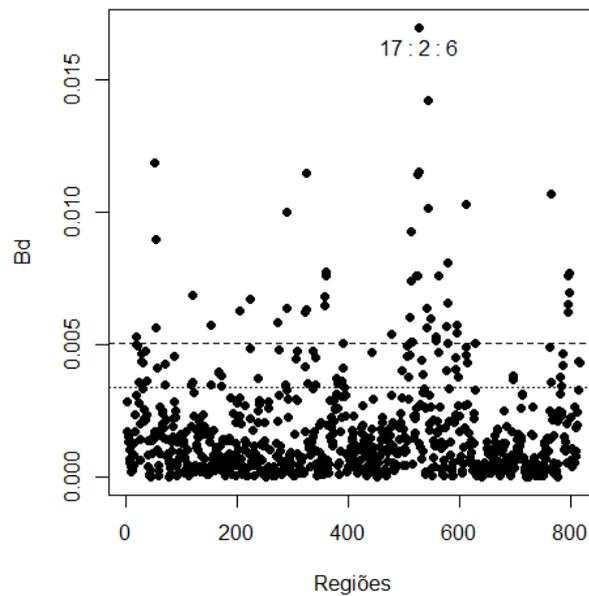


Gráfico 21 – Gráfico da medida de curvatura normal conformal B_d para avaliar a influência local sob esquema de perturbação na variável resposta, para a distribuição exponencial potência aplicado ao banco de dados do produto interno bruto.

Tabela 9 – Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto (sem a Região 8) para as distribuições normal e t-Student, segundo os modelos ajustados ao banco dados do produto interno bruto.

Sem Região 8				
Parâmetro	Normal		t-Student	
	Estimativa	Variação %	Estimativa	Variação %
β_{111}	-26,6823 (1,9487)	9,92	-27,8235 (1,8719)	6,55
β_{211}	0,2309 (0,0301)	6,74	0,2248 (0,0289)	5,98
β_{311}	0,0180 (0,0011)	7,22	0,0186 (0,0010)	4,62
σ_R^2	0,1943 (0,1401)	18,77	0,1519 (0,1202)	7,66
σ_S^2	0,388 (0,0969)	19,61	0,379 (0,1054)	17,08
σ^2	0,0044 (0,0040)	15,38	0,004 (0,0005)	6,98

Fonte: O autor (2019).

Tabela 10 – Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto (sem a observação 17 do estado 4 pertencente à Região 1) para as distribuições normal, t-Student e exponencial potência, segundo os modelos ajustados ao banco de dados do produto interno bruto.

Sem observação 17:4:1		
Normal		
Parâmetro	Estimativa	Variação %
β_{111}	-29,6827 (1,9199)	0,21
β_{211}	0,2420 (0,0285)	2,27
β_{311}	0,0194 (0,0011)	0,00
σ_R^2	0,2435 (0,1475)	1,79
σ_S^2	0,3301 (0,0748)	1,76
σ^2	0,0051 (0,0003)	1,92
t-Student		
Parâmetro	Estimativa	Variação %
β_{111}	-30,0281 (1,7705)	0,86
β_{211}	0,2306 (0,0264)	3,56
β_{311}	0,0197 (0,0010)	1,03
σ_R^2	0,1724 (0,1179)	4,78
σ_S^2	0,3381 (0,0895)	4,45
σ^2	0,0042 (0,0006)	2,33
Exp. Potência		
Parâmetro	Estimativa	Variação %
β_{111}	-29,5356 (1,9286)	0,37
β_{211}	0,2410 (0,0288)	2,57
β_{311}	0,0194 (0,0011)	0,52
σ_R^2	0,1337 (0,0815)	4,62
σ_S^2	0,1852 (0,0421)	4,75
σ^2	0,0027 (0,0001)	0,00

Fonte: O autor (2019).

Tabela 11 – Estimativas dos parâmetros com seus erros-padrão aproximados e variação relativa percentual em valor absoluto para a distribuições t-Student (sem a observação 1 do estado 1 pertencente à Região 8) e para a distribuição exponencial potência (sem a observação 17 do estado 2 pertencente à Região 6), segundo os modelos ajustados ao banco de dados do produto interno bruto.

Parâmetro	Sem observação 1:1:8		Sem Observação 17:2:6	
	t-Student		Exponencial Potência	
	Estimativa	Variação %	Estimativa	Variação %
β_{111}	-29,6777 (1,7818)	0,32	-29,7994 (1,9380)	1,26
β_{211}	0,2397 (0,0265)	0,24	0,2411 (0,0289)	2,55
β_{311}	0,0194 (0,0010)	0,51	0,0195 (0,0011)	1,04
σ_R^2	0,1663 (0,1135)	1,09	0,1812 (0,1042)	41,73
σ_S^2	0,3237 (0,0857)	0,00	0,1872 (0,0426)	5,85
σ^2	0,0043 (0,0006)	0,00	0,0027 (0,0001)	0,00

Fonte: O autor (2019).

Podemos ver que, apesar de a Região 8 se destacar em determinados esquemas de perturbação, quando esta foi retirada para ver se haveria uma mudança significativa nas novas estimativas dos parâmetros, não foi potencialmente suficiente para que estes sofressem grandes mudanças. Logo a Região 8 não pode ser considerada influente, apesar de ser um outlier. O mesmo ocorre para as outras observações que foram destaques nos esquemas de perturbação na variável preditora e na resposta de nível 1. Podemos ver também que para este conjunto de dados a distribuição exponencial foi ligeiramente melhor que a t-Student para comportar dados discrepantes, uma vez que para os esquemas de perturbação de caso e na matriz de variância e covariância nenhuma Região se mostrou notória.

Foi realizado o cálculo de pesos sob modelos de erro multinível no sentido da distância de Mahalanobis para a distribuição exponencial potência e t-Student em Manghi, Paula e Cysneiros (2016), em que o menor peso foi atribuído para a Região 8 sob o modelo exponencial potência, o que indica que mesmo esta Região se destacado como um outlier ela não é potencialmente suficiente para afetar a estimativa dos parâmetros. Isso corrobora para os resultados obtidos nas análises, na qual mostrou que as variações percentuais das estimativas quando retirada a Região 8 não houve eminente influência para alterar a inferência das estimativas dos parâmetros.

5 CONSIDERAÇÕES FINAIS

Nesta dissertação foram apresentadas técnicas de diagnóstico, mais especificamente a análise de influência local via curvatura normal para os modelos elípticos multiníveis, sendo esta a maior contribuição deste trabalho, por não haver na literatura estudos na área para a referida classe de modelos. Assim, este trabalho preencheu uma lacuna em que o Capítulo 3 constitui a principal contribuição teórica desta dissertação.

Assim, nota-se que os métodos diagnóstico para os modelos elípticos multiníveis desenvolvidos neste trabalho mostraram-se eficientes para a detecção de pontos com alta alavancagem, assim como observações aberrantes e influentes entre os conjuntos de dados com estrutura de hierarquia.

Com isso, conclui-se que as técnicas de diagnóstico na classe dos modelos elípticos multiníveis desenvolvidas aqui contribuem amplamente na área de modelagem por ser bastante útil em situações práticas em análise de regressão.

Nota-se que as técnicas desenvolvidas nesta dissertação detectaram pontos discrepantes como no banco de dados dos filhotes de rato as ninhadas 5, 6 e 18 que foi exemplo de nível dois. Para exemplo de nível 3 aplicado ao banco de dados do produto interno bruto, a Região 8 foi a que mais se destacou. Mas, uma análise mais profunda, verificou que, apesar dos pontos se destacarem na análise, eles não afetam as estimativas dos parâmetros.

Este trabalho veio preencher parcialmente uma lacuna nos métodos de diagnósticos para modelos de regressão elípticos multiníveis com a análise de influência local. Este, veio a reforçar o trabalho de Manghi (2011), de que os modelos elípticos multiníveis se mostram muito eficientes para a modelagem de dados com estrutura de hierarquia e com observações discrepantes. Nos exemplos utilizados no trabalho, temos que a distribuição t-Student se adequou melhor ao modelo de dois níveis e a distribuição exponencial potência, com o exemplo de três níveis conseguiu adequar-se de forma satisfatória ao conjunto de dados do produto interno bruto.

Toda a análise de estimações e métodos gráficos para a verificação de observações potencialmente influentes foram realizados por meio do software livre R em sua versão 3.6.0, este disponível gratuitamente em www.r-project.org.

REFERÊNCIAS

- ANDERSON, T. W. **Statistical inference in elliptically contoured and related distributions**. [S.l.]: Allerton Pr, 1990. 19
- ARELLANO-VALLE, R. **Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão**. Tese (Doutorado) — Ph. D. thesis, Instituto de Matemática e Estatística, Universidade de São Paulo, 1994. 23
- COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 48, n. 2, p. 133–155, 1986. 28, 29, 30, 31
- CYSNEIROS, F. J. A.; PAULA, G. A. Restricted methods in symmetrical linear regression models. **Computational statistics & data analysis**, Elsevier, v. 49, n. 3, p. 689–708, 2005. 17
- CYSNEIROS, F. J. A.; PAULA, G. A.; GALEA, M. Modelos simétricos aplicados. **Escola de Modelos de Regressao**, v. 9, 2005. 25
- CYSNEIROS, F. J. A.; VANEGAS, L. H. Residuals and their statistical properties in symmetrical nonlinear models. **Statistics & Probability Letters**, Elsevier, v. 78, n. 18, p. 3269–3273, 2008. 15, 16, 17
- DÍAZ-GARCÍA, J. A.; ROJAS, M. G.; LEIVA-SÁNCHEZ, V. Influence diagnostics for elliptical multivariate linear regression models. **Communications in statistics-Theory and Methods**, Taylor & Francis, v. 32, n. 3, p. 625–641, 2003. 15, 16
- EUBANK, R. Diagnostics for smoothing splines. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 332–341, 1985. 28
- FANG, K.-T.; KOTZ, S.; NG, K. W. **Symmetric Multivariate and Related Distributions**. [S.l.]: Springer US, 1990. (Monographs on Statistics and Applied Probability). 15, 18, 19
- FUNG, W. K.; TANG, M. Assessment of local influence in multivariate regression analysis. **Communications in statistics-theory and methods**, Taylor & Francis, v. 26, n. 4, p. 821–837, 1997. 16
- GALEA, M.; PAULA, G. A.; BOLFARINE, H. Local influence in elliptical linear regression models. **Journal of the Royal Statistical Society: Series D (The Statistician)**, Wiley Online Library, v. 46, n. 1, p. 71–79, 1997. 15, 16
- GOLDSTEIN, H. **Multilevel statistical models**. [S.l.]: John Wiley & Sons, 2011. v. 922. 15, 19
- GREENE, W. H. **Econometric analysis**. 5. ed. [S.l.]: Prentice Hall, 2003. 50
- HOX, J. J. **Applied multilevel analysis**. [S.l.]: TT-publikaties, 1995. 19
- HUFF, D. **Como mentir com estatística**. [S.l.]: Editora Intrínseca, 2016. 6
- IBACACHE-PULGAR, G.; PAULA, G. A. Local influence for student-t partially linear models. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 3, p. 1462–1478, 2011. 28
- IBACACHE-PULGAR, G.; PAULA, G. A.; GALEA, M. Influence diagnostics for elliptical semiparametric mixed models. **Statistical Modelling**, Sage Publications Sage India: New Delhi, India, v. 12, n. 2, p. 165–193, 2012. 30, 31

- IBACACHE-PULGAR, G.; PAULA, G. A.; GALEA, M. On influence diagnostics in elliptical multivariate regression models with equicorrelated random errors. **Statistical Methodology**, Elsevier, v. 16, p. 14–31, 2014. 31
- JENNRICH, R. I.; SCHLUCHTER, M. D. Unbalanced repeated-measures models with structured covariance matrices. **Biometrics**, v. 42, n. 4, p. 805–820, 1986. 18
- JR, J. E. D.; SCHNABEL, R. B. **Numerical methods for unconstrained optimization and nonlinear equations**. [S.l.]: Siam, 1996. v. 16. 24
- KAI-TAI, F.; YAO-TING, Z. **Generalized multivariate analysis**. [S.l.]: Science Press Beijing and Springer-Verlag, Berlin, 1990. v. 19. 19
- KIM, M. G. Local influence in multivariate regression. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 24, n. 5, p. 1271–1278, 1995. 16
- LANGE, K. L.; LITTLE, R. J.; TAYLOR, J. M. Robust statistical modeling using the t distribution. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 84, n. 408, p. 881–896, 1989. 16
- LEIVA, V.; LIU, S.; SHI, L.; CYSNEIROS, F. J. A. Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics. **Journal of Applied Statistics**, Taylor & Francis, v. 43, n. 4, p. 627–642, 2016. 15, 16, 31
- LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, JSTOR, p. 570–582, 1998. 28, 30
- LIU, S. Local influence in multivariate elliptical linear regression models. **Linear Algebra and Its Applications**, Elsevier, v. 354, n. 1-3, p. 159–174, 2002. 16, 17, 28, 31
- MANGHI, R. F. **Modelos elípticos multiníveis**. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2011. 16, 17, 18, 23, 24, 38, 39, 40, 65
- MANGHI, R. F.; PAULA, G. A.; CYSNEIROS, F. J. A. On elliptical multilevel models. **Journal of Applied Statistics**, Taylor & Francis, v. 43, n. 12, p. 2150–2171, 2016. 17, 19, 41, 50, 51, 53, 64
- MUNNEL, A. H. Why has productivity declined? productivity and public investment. **New England Economic Review**, January Feb, 1990. 50
- NOBRE, J. S. **Métodos de diagnóstico para modelos lineares mistos**. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2004. 28, 31
- NOBRE, J. S.; SINGER, J. da M. Residual analysis for linear mixed models. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 49, n. 6, p. 863–875, 2007. 28, 31
- OSORIO, F.; PAULA, G. A.; GALEA, M. Assessment of local influence in elliptical linear models with longitudinal structure. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4354–4368, 2007. 25, 26, 31, 34, 37
- PAULA, G. A. **Modelos de regressão: com apoio computacional**. [S.l.]: IME-USP São Paulo, 2004. 17

- PAULA, G. A.; MEDEIROS, M.; VILCA-LABRA, F. E. Influence diagnostics for linear models with first-order autoregressive elliptical errors. **Statistics & probability letters**, Elsevier, v. 79, n. 3, p. 339–346, 2009. 28, 31
- PINHEIRO, J.; BATES, D. **Mixed-effects models in S and S-PLUS**. [S.l.]: Springer Science & Business Media, 2006. 38
- PINHEIRO, J. C.; LIU, C.; WU, Y. N. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 10, n. 2, p. 249–276, 2001. 16
- PINHEIRO, S. **Modelo linear hierárquico: Um método alternativo para análise de desempenho escolar**. Tese (Doutorado) — Master's thesis, Universidade Federal de Pernambuco, 2005. 15
- PIRES, J. F. **Influência local através da curvatura normal em modelos multiníveis**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, Recife, 2009. 15
- POON, W.-Y.; POON, Y. S. Conformal normal curvature and assessment of local influence. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 61, n. 1, p. 51–61, 1999. 30
- RAUDENBUSH, S. W.; BRYK, A. S. **Hierarchical linear models: Applications and data analysis methods**. [S.l.]: Sage, 2002. v. 1. 24
- RODRIGUES, H. d. M. **Técnicas de diagnóstico nos modelos lineares generalizados com superdispersão**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2013. 32
- RUSSO, C. M.; PAULA, G. A.; AOKI, R. Influence diagnostics in nonlinear mixed-effects elliptical models. **Computational Statistics & Data Analysis**, Elsevier, v. 53, n. 12, p. 4143–4156, 2009. 30
- RUSSO, C. M.; PAULA, G. A.; CYSNEIROS, F. J. A.; AOKI, R. Influence diagnostics in heteroscedastic and/or autoregressive nonlinear elliptical models for correlated data. **Journal of Applied Statistics**, Taylor & Francis, v. 39, n. 5, p. 1049–1067, 2012. 28, 30
- SILVA, P. G. d. **Inferência e diagnóstico em modelos não lineares Log-Gama generalizados**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2016. 32
- SILVERMAN, B. W. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 47, n. 1, p. 1–21, 1985. 28
- SOARES, T. M. Modelo de três níveis hierárquicos para a proficiência dos alunos de 4a série avaliados no teste de língua portuguesa do simave/proeb-2002. **Revista brasileira de educação**, Associação Nacional de Pós-Graduação e Pesquisa em Educação, v. 29, p. 73–87, 2005. 15
- VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. [S.l.]: Springer Science & Business Media, 2009. 30
- VILLEGAS, C.; PAULA, G. A.; CYSNEIROS, F. J. A.; GALEA, M. Influence diagnostics in generalized symmetric linear models. **Computational Statistics & Data Analysis**, Elsevier, v. 59, p. 161–170, 2013. 28, 29, 30

WEI, B.-C.; HU, Y.-Q.; FUNG, W.-K. Generalized leverage and its applications. **Scandinavian Journal of statistics**, Wiley Online Library, v. 25, n. 1, p. 25–37, 1998. 35

WEST, B. T.; WELCH, K. B.; GALECKI, A. T. **Linear mixed models: a practical guide using statistical software**. [S.l.]: Chapman and Hall/CRC, 2014. 38, 39, 40

ZHU, H.-T.; LEE, S.-Y. Local influence for generalized linear mixed models. **Canadian Journal of Statistics**, Wiley Online Library, v. 31, n. 3, p. 293–309, 2003. 28, 30, 31

APÊNDICE A – INFORMAÇÃO DE FISHER

Essas quantidades são úteis para o cálculo da curvatura normal conformal. Para obter a matriz de informação observada $\ddot{\ell}_{\tau\tau}$ a matriz de informação observada de Fisher segue como:

$$\ddot{\ell}_{\tau\tau} = \begin{bmatrix} \ddot{\ell}_{\beta\beta} & \ddot{\ell}_{\beta\theta} \\ \ddot{\ell}_{\theta\beta} & \ddot{\ell}_{\theta\theta} \end{bmatrix},$$

segue que

$$\begin{aligned} \ddot{\ell}_{\beta\beta} &= \frac{\partial^2 \ell(\beta, \theta)}{\partial \beta \partial \beta^\top} = -2 \sum_{k=1}^K \left[\left(\frac{g''(u_k) \frac{\partial u_k}{\partial \beta} g(u_k) - g'(u_k) g'(u_k) \frac{\partial u_k}{\partial \beta}}{g^2(u_k)} \right) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right. \\ &\quad \left. + \frac{g'(u_k)}{g(u_k)} \left(-\mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \right) \right] \\ &= -2 \sum_{k=1}^K \left[\left(\frac{g''(u_k) \frac{\partial u_k}{\partial \beta}}{g(u_k)} - \left(\frac{g'(u_k)}{g(u_k)} \right)^2 \frac{\partial u_k}{\partial \beta} \right) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) - \frac{g'(u_k)}{g(u_k)} \mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \right], \\ \ddot{\ell}_{\beta\theta} &= \frac{\partial^2 \ell(\beta, \theta)}{\partial \beta \partial \theta^\top} = -2 \sum_{k=1}^K \left[\left(\frac{g''(u_k) (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \left[-\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \beta)}{g(u_k)} \right. \right. \\ &\quad \left. \left. - \left(\frac{g'(u_k)}{g(u_k)} \right)^2 (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \left[-\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \beta) \right) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right. \\ &\quad \left. + \frac{g'(u_k)}{g(u_k)} \mathbf{X}_k^\top \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right], \\ \ddot{\ell}_{\theta\theta} &= \frac{\partial^2 \ell(\beta, \theta)}{\partial \theta \partial \theta^\top} = - \sum_{k=1}^K \left[\left(\frac{g''(u_k) \frac{\partial u_k}{\partial \theta_d}}{g(u_k)} - \left(\frac{g'(u_k)}{g(u_k)} \right)^2 \frac{\partial u_k}{\partial \theta_d} \right) \times \right. \\ &\quad \left. (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \beta) \right] \\ &\quad + \frac{g'(u_k)}{g(u_k)} (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \\ &\quad + \frac{g'(u_k)}{g(u_k)} (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \mathbf{V}_k^{-1} \left(\frac{\partial \dot{\mathbf{V}}_k}{\partial \theta_d} \right) \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \\ &\quad \left. + \frac{g'(u_k)}{g(u_k)} (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right]. \end{aligned}$$

APÊNDICE B – PERTURBAÇÃO NA MATRIZ V_k

Sabendo que $\mathbf{V}_k = \mathbf{Z}_k \Psi \mathbf{Z}_k^\top + \mathbf{R}_k$. Para perturbar a matriz \mathbf{V}_k é feita a esta a escalar de perturbação ω_k , isto é, $\omega_k \mathbf{V}_k$ na log-verossimilhança. Então a verossimilhança é dada por:

$$\ell(\tau) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log g(u_k) \right\},$$

e a log-verossimilhança perturbada:

$$\ell(\tau|\omega) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\omega_k \mathbf{V}_k| + \log g \left[(\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right] \right\}.$$

Da log-verossimilhança perturbada vamos fazer a primeira derivada em relação a β :

$$\begin{aligned} \frac{\partial \ell(\tau|\omega)}{\partial \beta} &= \sum_{k=1}^K \left\{ \frac{g'(u_k \omega_k)}{g(u_k \omega_k)} \left[-2 \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right] \right\} \\ &= \sum_{k=1}^K \left\{ -2 W_g(u_k \omega_k) \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\} \\ &= \sum_{k=1}^K \left\{ v(u_k \omega_k) \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}, \end{aligned}$$

sabemos que:

$$\begin{aligned} u_k \omega_k &= (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \quad e \\ \frac{\partial u_k \omega_k}{\partial \beta} &= -2 \mathbf{X}_k^\top \omega_k \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta). \end{aligned}$$

Então, realizando a segunda derivada da equação anterior em relação a ω_k temos:

$$\begin{aligned} \frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega_k} &= -2 \sum_{k=1}^K \left\{ W_g'(u_k \omega_k) \frac{\partial u_k \omega_k}{\partial \omega_k} \mathbf{X}_k^\top [\omega_k \mathbf{V}_k]^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right. \\ &\quad \left. + W_g(u_k \omega_k) \mathbf{X}_k^\top [-\omega_k]^{-2} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\} \\ &= -2 \sum_{k=1}^K \left\{ W_g'(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \beta)^\top [-\omega_k]^{-2} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right. \\ &\quad \left. + W_g(u_k \omega_k) \mathbf{X}_k^\top [-\omega_k]^{-2} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \beta) \right\}, \end{aligned}$$

avaliando a derivada acima em $\tau = \hat{\tau}; \omega = \omega_0$, temos:

$$\begin{aligned} &= -2 \sum_{k=1}^K \left\{ W_g'(\hat{u}_k) \hat{u}_k \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) + W_g(\hat{u}_k) \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\} \\ &= -2 \sum_{k=1}^K \left\{ [W_g'(\hat{u}_k) \hat{u}_k + W_g(\hat{u}_k)] \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) \right\}. \end{aligned} \tag{B.1}$$

A primeira derivada da log-verossimilhança em relação a θ_d :

$$\begin{aligned}
\frac{\partial \ell(\tau|\omega)}{\partial \theta_d} &= \sum_{k=1}^K \left\{ -\frac{1}{2} \text{tr} \left[\omega_k \mathbf{V}_k^{-1} \frac{\partial \omega_k \mathbf{V}_k}{\partial \theta_d} \right] + \frac{g'(u_k \omega_k)}{g(u_k \omega_k)} \frac{\partial u_k \omega_k}{\partial \theta_d} \right\} \\
&= \sum_{k=1}^K \left\{ -\frac{1}{2} \text{tr} \left[\omega_k \mathbf{V}_k^{-1} \frac{\partial \omega_k \mathbf{V}_k}{\partial \theta_d} \right] + W_g(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [\omega_k]^{-1} \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right\} \\
&= \sum_{k=1}^K \left\{ -\frac{1}{2} \text{tr} \left[\omega_k \mathbf{V}_k^{-1} \frac{\partial \omega_k \mathbf{V}_k}{\partial \theta_d} \right] \right. \\
&\quad \left. + W_g(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [\omega_k]^{-1} \left[-\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right\} \\
&= -\frac{1}{2} \sum_{k=1}^K \left\{ \text{tr} \left[\omega_k \mathbf{V}_k^{-1} \frac{\partial \omega_k \mathbf{V}_k}{\partial \theta_d} \right] \right. \\
&\quad \left. - \frac{1}{2} v(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [\omega_k]^{-1} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right\} \\
&= -\frac{1}{2} \sum_{k=1}^K \left\{ \underbrace{\text{tr} \left[\omega_k \hat{\mathbf{V}}_k^{-1} \frac{\partial \omega_k \hat{\mathbf{V}}_k}{\partial \theta_d} \right]}_0 \right. \\
&\quad \left. - \frac{1}{2} v(\hat{u}_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top [\omega_k]^{-1} \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right\} \\
&= -\sum_{k=1}^K \left\{ W_g(\hat{u}_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top [\omega_k]^{-1} \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right\}.
\end{aligned}$$

A segunda derivada da equação anterior em relação a ω_k , é dada por:

$$\begin{aligned}
\frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k} &= - \left\{ W'_g(u_k \omega_k) \frac{\partial u_k \omega_k}{\partial \omega_k} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [\omega_k]^{-1} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) + \right. \\
&\quad \left. W_g(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [-\omega_k]^{-2} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right\} \\
&= - \left\{ W'_g(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [-\omega_k]^{-2} \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [\omega_k]^{-1} \right. \\
&\quad \times \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) + \\
&\quad \left. W_g(u_k \omega_k) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top [-\omega_k]^{-2} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right\},
\end{aligned}$$

avaliando $\tau = \hat{\tau}$; $\omega = \omega_0$, neste caso $\omega_k = (1, \dots, 1)^\top$

$$\begin{aligned}
&= - \left\{ W'_g(\hat{u}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right. \\
&\quad \left. + W_g(\hat{u}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right\} \quad (\text{B.2}) \\
&= - \{ W'_g(\hat{u}_k) \hat{u}_k + W_g(\hat{u}_k) \} (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}).
\end{aligned}$$

APÊNDICE C – PERTURBAÇÃO NA VARIÁVEL RESPOSTA

Temos o modelo dado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$

em que $\mathbf{X} = \mathbf{G}\mathbf{T}\mathbf{L}$, $\mathbf{Z} = [\mathbf{G}\mathbf{T}|\mathbf{G}]$ e $\mathbf{b} = (\mathbf{v}^\top, \boldsymbol{\xi}^\top)^\top$.

A perturbação na variável resposta $(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top, \dots, \mathbf{Y}_K^\top)^\top$ é introduzida substituindo \mathbf{Y}_k por $\mathbf{Y}_k + \boldsymbol{\omega}_k$. Aqui, considere uma perturbação aditiva sobre a i -ésima resposta, ou seja, a perturbação para cada indivíduo fica $(y_{i\omega} = y_{ijk} + \omega_{ijk})$. Neste caso o vetor de não perturbação é $\boldsymbol{\omega}_0 = (\mathbf{0}, \dots, \mathbf{0})^\top$.

Então temos a log-verossimilhança dada por:

$$\ell(\boldsymbol{\tau}) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log \mathbf{g} \left[(\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}) \right] \right\},$$

inserindo a perturbação na variável resposta, temos:

$$\ell(\boldsymbol{\tau}|\boldsymbol{\omega}) = \sum_{k=1}^K \left\{ -\frac{1}{2} \log |\mathbf{V}_k| + \log \left[((\mathbf{Y}_k + \boldsymbol{\omega}_k) - \mathbf{X}_k \boldsymbol{\beta})^\top \mathbf{V}_k^{-1} ((\mathbf{Y}_k + \boldsymbol{\omega}_k) - \mathbf{X}_k \boldsymbol{\beta}) \right] \right\}.$$

Neste esquema de perturbação por exemplo perturbaremos o nível 1, ou seja, cada indivíduo da resposta. Desta forma, temos que: $\mathbf{Y}_k = [\mathbf{Y}_{1k}^\top, \mathbf{Y}_{2k}^\top, \dots, \mathbf{Y}_{m_k k}^\top]_{\left(\sum_{j=1}^{m_k}\right) \times 1}$ e $\boldsymbol{\omega}_k = [\boldsymbol{\omega}_{1k}^\top, \boldsymbol{\omega}_{2k}^\top, \dots, \boldsymbol{\omega}_{m_k k}^\top]_{\left(\sum_{j=1}^{m_k}\right) \times 1}$ em que \mathbf{Y}_k é o k -ésimo “cluster” de nível três e m_k o total de indivíduos pertencente ao “cluster” k .

Temos ainda que: $\mathbf{Y}_{jk} = [\mathbf{Y}_{1jk}, \mathbf{Y}_{2jk}, \dots, \mathbf{Y}_{n_{jk}jk}^\top]_{n_{jk} \times 1}$ e para este j -ésimo “cluster” de nível dois pertencente ao k -ésimo “cluster” de nível três, temos o vetor de perturbação

$\boldsymbol{\omega}_{jk} = [\boldsymbol{\omega}_{1jk}, \boldsymbol{\omega}_{2jk}, \dots, \boldsymbol{\omega}_{n_{jk}jk}^\top]_{n_{jk} \times 1}$, em que n_{jk} é o número de indivíduos pertencente ao j -ésimo “cluster” no nível 2 pertencente ao k -ésimo “cluster” do nível 3. Desta forma a variável perturbada é dado da seguinte maneira:

$$\begin{aligned} (\mathbf{Y}_{jk} + \boldsymbol{\omega}_{jk})^\top &= [Y_{1jk}, Y_{2jk}, \dots, Y_{n_{jk}jk}]^\top + [\boldsymbol{\omega}_{1jk}, \boldsymbol{\omega}_{2jk}, \dots, \boldsymbol{\omega}_{n_{jk}jk}]^\top \\ &= [Y_{1jk} + \boldsymbol{\omega}_{1jk}, Y_{2jk} + \boldsymbol{\omega}_{2jk}, \dots, Y_{n_{jk}jk} + \boldsymbol{\omega}_{n_{jk}jk}]^\top. \end{aligned}$$

Então, a primeira derivada em relação a $\boldsymbol{\beta}$ é dada por:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\tau}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta}} &= \frac{g'(u_k \omega_k)}{g(u_k \omega_k)} \frac{\partial (u_k \omega_k)}{\partial \boldsymbol{\beta}} \\ &= -2W_g(u_k \omega_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} [\mathbf{Y}_k + \boldsymbol{\omega}_k - \mathbf{X}_k \boldsymbol{\beta}]. \end{aligned}$$

Segunda derivada em relação a ω_k

$$\begin{aligned}
\frac{\partial^2 \ell(\tau|\omega)}{\partial \beta \partial \omega_k^\top} &= -2 \left\{ W'_g(u_k \omega_k) \frac{\partial(u_k \omega_k)}{\partial \omega_k} \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) + W_g(u_k \omega_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} \right\} \\
&= -2 \left\{ 2W'_g(u_k \omega_k) \mathbf{V}_k^{-1} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \mathbf{X}_k^\top \mathbf{V}_k^{-1} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) + \right. \\
&\quad \left. W_g(u_k \omega_k) \mathbf{X}_k^\top \mathbf{V}_k^{-1} \right\} \\
&= -2 \mathbf{X}_k^\top \mathbf{V}_k^{-1} \left\{ W_g(u_k \omega_k) \mathbf{V}_k \right. \\
&\quad \left. + 2W'_g(u_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \right\} \mathbf{V}_k^{-1},
\end{aligned}$$

avaliando em $\tau = \hat{\tau}$; $\omega = \omega_0$

$$= -2 \mathbf{X}_k^\top \hat{\mathbf{V}}_k^{-1} \left\{ W_g(\hat{u}_k) \hat{\mathbf{V}}_k^{-1} + 2W'_g(\hat{u}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\beta})^\top \right\} \mathbf{V}_k^{-1}. \quad (\text{C.1})$$

Primeira derivada da log-verossimilhança em relação θ_d :

$$\begin{aligned}
\frac{\partial \ell(\tau|\omega)}{\partial \theta_d} &= -\frac{1}{2} \text{tr} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \right] + \frac{g'(u_k \omega_k)}{g(u_k \omega_k)} \frac{\partial(u_k \omega_k)}{\partial \theta_d} \\
&= W_g(u_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \frac{\partial \mathbf{V}_k^{-1}}{\partial \theta_d} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \\
&= W_g(u_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[-\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta).
\end{aligned}$$

Na segunda derivada temos:

$$\begin{aligned}
\frac{\partial^2 \ell(\tau|\omega)}{\partial \theta_d \partial \omega_k^\top} &= - \left\{ W'_g(u_k \omega_k) \frac{\partial u_k \omega_k}{\partial \omega_k} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \right. \\
&\quad + W_g(u_k \omega_k) \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \\
&\quad \left. + W_g(u_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] \right\} \\
&= - \left\{ 2W'_g(u_k \omega_k) \mathbf{V}_k^{-1} (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \right. \\
&\quad \times \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta) \\
&\quad + W_g(u_k \omega_k) \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \\
&\quad \left. + W_g(u_k \omega_k) (\mathbf{Y}_k + \omega_k - \mathbf{X}_k \beta)^\top \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_d} \mathbf{V}_k^{-1} \right] \right\},
\end{aligned}$$

avaliando $\tau = \hat{\tau}$; $\omega = \omega_0$ temos:

$$\begin{aligned}
&= - \left\{ 2W'_g(\hat{u}_k) \hat{\mathbf{V}}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right. \\
&\quad \left. + 2W_g(\hat{u}_k) \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) \right\} \\
&= -2(\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \left[\hat{\mathbf{V}}_k^{-1} \frac{\partial \hat{\mathbf{V}}_k}{\partial \theta_d} \hat{\mathbf{V}}_k^{-1} \right] \left\{ W_g(\hat{u}_k) \mathbf{V}_k + W'_g(\hat{u}_k) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}}) (\mathbf{Y}_k - \mathbf{X}_k \hat{\boldsymbol{\beta}})^\top \right\} \hat{\mathbf{V}}_k^{-1}.
\end{aligned} \tag{C.2}$$