



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E
CONTABILIDADE
CURSO DE GRADUAÇÃO EM CIÊNCIAS ATUARIAIS

JAIME PHASQUINEL LOPES CAVALCANTE

MODELOS DE REGRESSÃO BETA COM MEDIDAS REPETIDAS VIA
TEORIA DE EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS

FORTALEZA

2019

JAIME PHASQUINEL LOPES CAVALCANTE

MODELOS DE REGRESSÃO BETA COM MEDIDAS REPETIDAS VIA TEORIA DE
EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciências Atuariais da Faculdade de Economia, Administração, Atuária e Contabilidade da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciências Atuariais.

Orientador: Prof. Dr. Juvêncio Santos Nobre

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C364m Cavalcante, Jaime Phasquinel Lopes.
Modelos de Regressão Beta com Medidas Repetidas Via Teoria de Equações de Estimação Generalizadas /
Jaime Phasquinel Lopes Cavalcante. – 2019.
98 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Faculdade de Economia,
Administração, Atuária e Contabilidade, Curso de Ciências Atuariais, Fortaleza, 2019.
Orientação: Prof. Dr. Juvêncio Santos Nobre.

1. Modelagem. 2. Medidas Repetidas. 3. Equações de Estimação Generalizadas. 4. Modelo de Regressão
Beta. I. Título.

CDD 368.01

JAIME PHASQUINEL LOPES CAVALCANTE

MODELOS DE REGRESSÃO BETA COM MEDIDAS REPETIDAS VIA TEORIA DE
EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciências Atuariais da Faculdade de Economia, Administração, Atuária e Contabilidade da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciências Atuariais.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Juvêncio Santos Nobre (Orientador)
Universidade Federal do Ceará (UFC)

Prof^ª. Me Luciana Moura Reinaldo
Universidade Federal do Ceará - UFC

Prof^ª. Dra Iana Bezerra Jucá
Universidade Federal do Ceará - UFC

À minha família, por acreditar e investir em
minha educação. Aos amigos pelo suporte.
Ao Prof. Juvêncio pela confiança e orientação.
À Prof^a. Luciana pela oportunidade de dar
vazão às minhas potências.

AGRADECIMENTOS

Tomando por base de que nada é por acaso e que tudo tem um motivo, gostaria de deixar claro que dentro da construção desse trabalho, bem como ao longo de minha vida pessoas escolhidas a dedo pelo Universo surgiram para somar e me mostrar que estou no caminho certo em busca da realização dos meus planos.

Diante disso, sou grato, mesmo que não cabendo em palavras, a meus pais por todo amor e preocupação em todas as esferas da minha vida. Mãe e Pai, tudo o que já sou e o que estou por ser é graças a vocês. A meu irmão e cunhada, sou grato pelas nossas idas à praia, risadas e companheirismo. Vocês, minha família, são um pilar primordial para mim, mesmo que o isolamento em meu quarto tenha demonstrado o contrário, mas uma monografia não se faz sozinha.

À Ana Glads - minha eterna amiga, irmã de alma e afilhada de curso - gratidão pelas descobertas, pelos risos e lágrimas! Sua presença trouxe a cor em uma fase da minha vida em que eu só conseguia enxergar preto e branco. Novamente muito obrigado pelos gritos, conselhos, por me fazer despertar o prazer da companhia de Platão, Sócrates e Nietzsche, mas principalmente por acreditar em mim e embarcar nas minhas ideias, afinal na “Matrix” somos um só.

À Victória Viana, outra grande amiga, sou extremamente grato por desde a disciplina de álgebra linear se fazer presente em minha vida, tornando-a ainda mais feliz e suave, como nossos passeios no Campus do Pici.

Gostaria deixar expresso aqui meu agradecimento ao meu amigo Pedro, talvez sem você, meu oposto complementar, eu não tivesse conseguido atravessar o roseiral, posto que nem tudo são flores!

Também agradeço a Adriano Rennan, que mesmo tendo chegado recentemente tem feito um bem danado. À Suelem Ribeiro, Thais Castro, Luiz Paulo, Mainara Cardoso, Gisele Alves e July Saraiva fica uma imensa gratidão! Parafraseando Aristóteles, sem amigos ninguém escolheria viver, mesmo que tivesse todos os outros bens.

Dentro desse processo, gostaria de deixar minha eterna gratidão à professora Luciana Moura Reinaldo e aquele estimador de mínimos quadrados com intercepto nulo. Digo isso, pois foi através do meu erro (acreditar previamente que o estimador era o mesmo independente da presença ou não intercepto) que eu amadureci para o fato de que eu precisava estudar ainda mais e que sempre há algo para se aprender! Por consequência do

espertar para o conhecimento que você me causou, dando vazão às minhas potências, vieram minhas apresentações nos Encontros Universitário, nosso artigo PRIMORDIALMENTE apresentado na UFPB, a prazerosa habilidade de escrever em L^AT_EX e principalmente a sensação de ter alguém dizendo “Vai tartaruguinha, respira pelo umbigo e vai, pois você é capaz!”. Para além desses fatos, fica a felicidade por tantos outros momentos construídos desde o dia em que a conheci.

Também sou imensamente grato ao meu orientador, professor Juvêncio Santos Nobre. Sem sua orientação esse trabalho não seria possível! Além disso, agradeço pela confiança a mim concedida, isso foi fundamental para a construção do trabalho apresentado. Como nada é por acaso, conhecê-lo e ter a sorte de ser seu orientando despertou ainda mais o meu amor pelo conhecimento, além de me confirmar que eu ainda tenho muito a aprender, pois mesmo que eu ache que já sei ou que já tenha feito tudo, sempre há algo a ser descoberto e melhorado. Gratidão por essa experiência que me amadureceu, me dando um propósito, e que me tornou ainda mais pontual para as reuniões!

À professora Iana Jucá agradeço pelas disciplinas ministradas, pois foi em suas aulas que aprendi e vi as contribuições da Estatística em nossa Ciência Atuarial, um exemplo nítido apresentado em sala é a teoria da credibilidade. Ademais, agradeço por acreditar em mim e me estimular a focar nos estudos.

Também gostaria de agradecer a professora Maria Kelly Venezuela, pela tese que serviu de base e inspiração para este trabalho e por de forma solícita ajudar em minhas dúvidas.

Agradeço a todos os professores (desde o jardim de infância até a faculdade) por me proporcionar não apenas “o conhecimento dos livros”, mas pelo desenvolvimento do pensamento crítico em mim gerado. Quando Paulo Freire disse que o educador se eterniza em cada ser que educa, ele estava coberto de razão! Os levo no peito.

Também deixo meus agradecimentos ao eterno conhecedor da coordenação do curso, o excelentíssimo Senhor Secretário Emilson.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) agradeço pelo auxílio financeiro através de uma bolsa de iniciação científica.

Finalmente, gratidão ao Universo pela benevolência com que tem me tratado.

“Quem sabe faz a hora, não espera acontecer.”

(Geraldo Vandré)

RESUMO

Nas mais diversas áreas do conhecimento, tais como Atuária, Biologia, Economia, Engenharia, Medicina entre outras é comum a observação de estudos com medidas repetidas, ou seja, cada parcela amostral é observada em pelo menos duas vezes. Ademais, deve-se compreender que a presença de medidas repetidas pressupõe a existência de dependência entre as medidas das intra-unidades amostrais. No que tange a modelagem, ajuste e avaliação da qualidade do modelo para dados com medidas repetidas, o presente estudo destaca o uso das Equações de Estimação Generalizadas (EEG's) proposta por Liang e Zeger (1986) sob a ótica de funções de estimação (Godambe,1991). A escolha se deu devido a vantagem de que tal proposta permite a flexibilização do modelo probabilístico adotado para a variável resposta. Dessa forma, não se faz necessário sequer ter o conhecimento sobre a distribuição conjunta associada, mas somente as distribuições marginais, no qual inicialmente foi assumido serem membros da família exponencial linear, permitindo assim uma alta flexibilidade de modelagem. Com base no exposto, o presente estudo, tomando por base a proposta de Venezuela (2008), apresenta a extensão do uso das EEG's na presença de medidas repetidas quando a distribuição marginal é Beta, generalizando assim o modelo de regressão Beta de Ferarri e Cribari-Neto (2004).

Palavras-chave: Modelagem. Medidas Repetidas. Equações de Estimação Generalizadas. Modelo de Regressão Beta.

ABSTRACT

In the most diverse areas of knowledge, such as Actuarial Science, Biology, Economics, Engineering, Medicine, among others, it is common to observe studies with repeated measurements, that is, each sample plot is observed at least twice. Moreover, it should be understood that the presence of repeated measurements imply the existence of dependence between the measurements of the sample intraunits. Regarding the modeling, adjustment and evaluation of the quality of the model for data with repeated measurements, this study highlights the use of Generalized Estimation Equations (GEE'S) proposed by Liang and Zeger (1986) from the perspective of estimation functions (Godambe, 1991). The choice was due to the advantage that this proposal allows the flexibility of the probabilistic model adopted for the response variable. Thus, it is not even necessary to have the knowledge about the associated joint distribution, but only the marginal distributions, in which it was initially assumed to be members of the exponential linear family, thus allowing a high flexibility of modeling. Based on the above, this study, based on the proposal of Venezuela (2008), presents the extent of the use of EEGs in the presence of repeated measures when the marginal distribution is Beta, thus generalizing the Beta de Ferarri and Cribari-Neto regression model (2004).

Keywords: Modeling. Repeated measurements. Generalized Estimation Equations. Beta Regression Model.

LISTA DE FIGURAS

Figura 1 – Densidades beta para combinações diferentes de (μ, ϕ)	26
Figura 2 – Gráficos de diagnóstico do modelo de regressão beta ajustado para os dados.	35
Figura 3 – Gráficos de dispersão do percentual de escore versus a idade corrigida, para os grupos “A termo” e “Pré-Termo”.	84
Figura 4 – Curvas ajustadas para o modelo de regressão beta sob homogeneidade da dispersão.	85
Figura 5 – Gráficos de diagnóstico do modelo de regressão beta sob homogeneidade da dispersão.	86
Figura 6 – Gráficos de diagnóstico do modelo de regressão beta sob homogeneidade da dispersão.	87
Figura 7 – Curva ajustadas para o modelo de regressão beta sob heterogeneidade da dispersão.	88
Figura 8 – Gráficos de diagnóstico do ajuste de β para o modelo de regressão beta sob heterogeneidade da dispersão.	89
Figura 9 – Gráficos de diagnóstico do ajuste de γ para o modelo de regressão beta sob heterogeneidade da dispersão.	90

LISTA DE TABELAS

Tabela 1 – Algumas distribuições da família exponencial linear	18
Tabela 2 – Funções de ligação canônicas para alguns membros da Família Exponencial Linear	20
Tabela 3 – Valores estimados do modelo de regressão beta	34
Tabela 4 – Valores estimados do modelo de regressão beta (sem a observação 4) .	36
Tabela 5 – Estimativas e erros padrões para os parâmetros da modelagem sob homogeneidade da dispersão	84
Tabela 6 – Estimativas e erros padrões para os parâmetros da modelagem sob heterogeneidade da dispersão	88

SUMÁRIO

1	INTRODUÇÃO	15
2	O MODELO LINEAR GENERALIZADO	16
2.1	Família exponencial linear	17
2.1.1	<i>Média e variância</i>	17
2.1.2	<i>Alguns casos particulares da família exponencial linear</i>	18
2.2	Especificação do Modelo Linear Generalizado	19
2.3	Função desvio	21
2.4	Testes de hipóteses	22
2.5	Técnicas de diagnóstico	22
2.5.1	<i>Pontos de alavanca</i>	22
2.5.2	<i>Resíduos</i>	23
2.5.3	<i>Envelope Simulado</i>	23
3	MODELO DE REGRESSÃO BETA	25
3.1	Distribuição Beta	26
3.2	Definição do modelo de regressão beta	27
3.3	Resíduos	28
3.3.1	<i>Resíduo ordinário</i>	28
3.3.2	<i>Resíduo ponderado padronizado 1</i>	29
3.3.3	<i>Resíduo ponderado padronizado 2</i>	29
3.4	Métodos de diagnóstico	30
3.4.1	<i>Ponto de alavanca</i>	30
3.4.2	<i>Distância de Cook</i>	31
3.4.3	<i>Gráfico de probabilidade meio-normal com envelope simulado</i>	32
3.5	Aplicação dos modelos de regressão beta através do uso do software R	33
3.5.1	<i>Aplicação</i>	34
4	EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS	37
4.1	Funções de estimação	38
4.2	Modelagem da média	43
4.3	Equações de Estimação Generalizadas	43

4.4	Testes de hipóteses	46
4.4.1	<i>Teste de Wald</i>	46
4.4.2	<i>Teste Escore</i>	47
4.5	Técnicas de diagnóstico para EEG's	48
4.5.1	<i>Alavancagem</i>	48
4.5.2	<i>Análise de resíduos</i>	49
4.5.3	<i>Envelope simulado</i>	50
5	MODELOS DE REGRESSÃO BETA COM MEDIDAS RE- PETIDAS	52
5.1	EEG's para modelos de regressão beta com presen- ça de medidas repetidas	52
5.1.1	<i>Modelagem do parâmetro de posição</i>	53
5.1.2	<i>Estimação de β, ϕ e α</i>	56
5.1.3	<i>Modelagem conjunta dos parâmetros de posição e dispersão</i>	59
5.1.4	<i>Estimação de β, γ e α</i>	61
5.2	Técnicas de diagnósticos para dados com medidas repetidas .	63
5.2.1	<i>Pontos influentes, aberrantes e alavanca</i>	64
5.2.2	<i>Ponto influente associado à distância de Cook</i>	65
5.2.3	<i>Ponto aberrante associado ao resíduo padronizado</i>	65
5.2.4	<i>Ponto de alavanca associado à matriz de projeção</i>	66
5.3	Critérios de seleção de modelos e de matriz de correlação de trabalho	67
5.4	Influência local para dados com medidas repetidas	68
5.5	Influência local generalizada	70
5.6	Influência local para equações de estimação	71
5.7	Esquemas de perturbação sob homogeneidade da dispersão .	75
5.7.1	<i>Ponderação de casos</i>	75
5.7.2	<i>Perturbação da variável resposta</i>	76
5.7.3	<i>Perturbação individual das covariáveis</i>	77
5.7.4	<i>Perturbação do parâmetro de precisão</i>	78
5.7.5	<i>Perturbação na matriz de correlação de trabalho $R(\alpha)$</i>	79
5.8	Esquemas de perturbação sob heterogeneidade da dispersão . .	81

5.8.1	<i>Ponderação de casos</i>	81
5.8.2	<i>Perturbação da variável resposta</i>	82
6	APLICAÇÃO	83
7	CONSIDERAÇÕES FINAIS	91
	REFERÊNCIAS	92

1 INTRODUÇÃO

Comumente, as mais diversificadas áreas do conhecimento apresentam estudos em que cada parcela amostral é verificada em pelo menos duas vezes, de modo que os mesmos passam a ser caracterizados como análises de medidas repetidas.

Outro aspecto referente à utilização de estudos com medidas repetidas é o pressuposto de que há existência de dependência entre as medidas das intra-unidades amostrais.

Além disso, no que tange o ajuste, modelagem e avaliação da qualidade dos modelos para dados com medidas repetidas, destaca-se o uso das Equações de Estimação Generalizadas (EEG's) proposta por Liang e Zeger (1986) sob a ótica de funções de estimação (Godambe,1991).

A utilização de tal metodologia é justificada fato de que não se faz necessário sequer ter o conhecimento sobre a distribuição conjunta associada, mas somente as distribuições marginais, no qual inicialmente foi assumido serem membros da família exponencial linear, proporcionando alta flexibilidade de modelagem.

Com base no exposto, o presente estudo, tomando por base a proposta de Venezuela (2008), apresenta como objetivo principal o estudo da extensão do uso das EEG's na presença de medidas repetidas quando a distribuição marginal é Beta, generalizando assim o modelo de regressão Beta de Ferarri e Cribari-Neto (2004). Já os objetivos específicos são: estudar a teoria dos modelos de regressão beta, estudar a teoria de funções de estimação e de equações de estimação generalizadas, compreender os métodos de diagnósticos para o referido tema e aplicar o tema em estudo em um conjunto de dados reais.

Finalmente, o presente trabalho está estruturado de acordo com a seguinte ordem: no Capítulo 02 são apresentados os aspectos fundamentais para a compreensão do Modelo Linear Generalizado; no Capítulo 03 é apresentada uma discussão sobre o Modelo de Regressão Beta, onde é apresentado uma aplicação por meio do software R; no Capítulo 04 é destacada a fundamentação teórica das Equações de Estimação Generalizadas; no Capítulo 05 seguem as teorias sobre os Modelos de Regressão Beta com Medidas Repetidas, além da apresentação das propostas de diagnósticos para os modelos; no Capítulo 06 se encontra uma aplicação que busca a compreensão das performances motoras de recém-nascidos; no Capítulo 07 são dispostas as considerações finais do estudo.

2 O MODELO LINEAR GENERALIZADO

Comumente em diversas situações de interesse, sejam elas da ordem experimental ou observacional, pesquisadores são confrontados com problemas cujo o objetivo principal é o de estudar o comportamento (relação) entre variáveis. Nesse sentido, especificamente, busca-se compreender a influência exercida por uma ou mais variáveis (*explicativas*) sobre uma determinada variável de interesse denominada *variável resposta*. Assim, a metodologia, em geral, utilizada em tais situações é através da análise de regressão.

Em linha com Turkman e Silva (2000), o modelo linear, “desenvolvido” no século XIX por Legendre e Gauss, foi o principal método utilizado para a modelagem estatística até meados do século XX. Contudo, ao longo de tal período, uma vasta gama de modelos baseados em distribuições diferentes da normal foram desenvolvidos para atuar em situações em que o modelo baseado na distribuição normal não era adequado. Pode-se citar, como exemplos, o modelo *complemento log-log* (Fisher, 1922), o modelo *Probit* (Bliss, 1935), os modelos *log-lineares* para dados de contagem (Birch, 1963), os modelos de regressão para análise de sobrevivência (Feigl and Zelen, 1965; Zippin and Armitage, 1966; Glasser, 1967).

Tais modelos apresentados, entre outros, ditos alternativos ao modelo linear normal, compartilham as seguintes características: geralmente apresentam uma estrutura de regressão linear e a variável resposta do modelo segue uma distribuição pertencente a *família exponencial linear*, em que são chamados de Modelos Lineares Generalizados (MLG's).

Segundo Paula (2013), os modelos lineares generalizados expandem a gama de opções no que tange a distribuição da variável resposta, fazendo com que a mesma faça parte da família exponencial linear. Desta forma, geram uma maior flexibilidade para a relação funcional entre o valor médio da variável resposta e o preditor linear η .

As contribuições presentes em Nelder e Wedderburn (1972), além de responsáveis pela proposta do modelo, apresentam um processo iterativo objetivando a estimação dos parâmetros e introduziram o conceito de função desvio (amplamente utilizado para a mensuração da qualidade do ajuste dos MLG's).

2.1 Família exponencial linear

Conceitua-se que uma variável aleatória Y apenas é pertencente à família exponencial linear se a sua função densidade de probabilidade (f.d.p) ou sua função de massa de probabilidade (f.m.p) possa ser escrita na forma a seguir:

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\theta)} + c(\theta, \phi) \right\}, \quad (2.1)$$

em que, $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções reais específicas. θ é chamado de parâmetro de localização e ϕ é o parâmetro de dispersão. Ademais, ressalta-se que o suporte não pode depender funcionalmente de θ .

No modelo apresentado em (2.1), θ é denominado parâmetro canônico e ϕ é um parâmetro, em geral, conhecido. Dessa forma, diz-se que a distribuição descrita faz parte da família exponencial linear uniparamétrica.

2.1.1 Média e variância

Ainda no que tange a família exponencial linear, seja $l(\theta; \phi, y) = \ln(f(y|\theta, \phi))$. Define-se a função *score*:

$$S(\theta) = \frac{\partial l(\theta; \phi, Y)}{\partial \theta}$$

Considerando válida as condições de regularidade, tem-se que $\forall \theta \in \Theta$:

$$E(S(\theta)) = 0,$$

e

$$E(S^2(\theta)) = E \left[\left(\frac{\partial l(\theta; \phi, Y)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 l(\theta; \phi, Y)}{\partial \theta^2} \right].$$

Lembrando (2.1), obtêm-se:

$$l(\theta; \phi, y) = \frac{y\theta - b(\theta)}{a(\phi)},$$

de forma que:

$$E(Y) = \mu = b'(\theta)$$

e

$$\text{Var}(Y) = a^2(\phi) \frac{b'(\theta)}{a(\phi)} = a(\phi)b''(\theta),$$

de modo que b' representa a derivada da função $b(\cdot)$ e b'' a derivada segunda da mesma função. Assim, verifica-se que a variância de Y é o produto de duas funções: **função de variância** $V(\mu)$, que depende apenas do parâmetro canônico θ (valor médio μ) e outra, $a(\phi)$, dependendo apenas do parâmetro de dispersão ϕ .

É possível, ainda, que função $a(\phi)$ possa ser reescrita da seguinte forma:

$$a(\phi) = \frac{\phi}{\omega},$$

em que ω se apresenta como uma constante conhecida. Logo, é possível obter a variância de Y como o produto da métrica de dispersão por uma função apenas do valor médio. Nesse caso, (2.1) pode ser reescrita escrita como:

$$f(y|\theta, \phi, \omega) = \exp \left\{ \frac{\omega}{\phi} (y\theta - b(\theta)) + c(y, \phi, \omega) \right\}. \quad (2.2)$$

2.1.2 Alguns casos particulares da família exponencial linear

Na tabela a seguir é apresentada uma listagem das principais distribuições pertencentes à família exponencial e suas características.

Tabela 1 – Algumas distribuições da família exponencial linear

Distribuição	Normal	Binomial	Poisson	Gama
Notação	$N(\mu, \sigma)$	$B(m, \pi)/m$	$P(\lambda)$	$Ga(v, \frac{m\mu}{v})$
Suporte	$\mathbb{R}(-\infty, +\infty)$	$\{0, \frac{1}{m}, \dots, 1\}$	$\mathbb{N}(0, 1, \dots)$	$\mathbb{R}(0, +\infty)$
θ	μ	$\ln\left(\frac{\pi}{1-\pi}\right)$	$\ln(\lambda)$	$-\frac{1}{\mu}$
$a(\phi)$	σ^2	$\frac{1}{m}$	1	$\frac{1}{v}$
ϕ	σ^2	1	1	$\frac{1}{v}$
ω	1	m	1	1
$c(y, \phi)$	$-\frac{1}{2}\left(\frac{y^2}{\phi} + \ln(2\pi\phi)\right)$	$\ln\left(\frac{m}{my}\right)$	$-\ln y!$	$v \ln v - \ln \Gamma(v) + (v-1) \ln y$
$b(\theta)$	$\frac{\theta^2}{2}$	$\ln(1 + e^\theta)$	e^θ	$-\ln(-\theta)$
$b'(\theta) = E(Y)$	θ	$\pi = \frac{e^\theta}{1+e^\theta}$	$\lambda = e^\theta$	$\mu = -\frac{1}{\theta}$
$b''(\theta) = V(\mu)$	1	$\pi(1-\pi)$	λ	μ^2
$\text{Var}(Y)$	σ^2	$\frac{\pi(1-\pi)}{m}$	λ	$\frac{\mu^2}{v}$

Fonte: elaborado pelo autor.

2.2 Especificação do Modelo Linear Generalizado

O modelo linear generalizado pode ser compreendido como um extensão do modelo linear, definido como:

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

assim, X pode ser compreendido como uma matriz de dimensão $n \times p$, relacionada a um vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ de parâmetros e $\boldsymbol{\varepsilon}$ é a componente (vetor) de erros aleatoriamente distribuídos com distribuição $N_n(0, \sigma^2 \mathbf{I})$.

Tais suposições definem que o valor esperado da variável resposta (\mathbf{Y}) é uma função linear de suas covariáveis, ou seja, $E(\mathbf{Y}|X) = \boldsymbol{\mu}$ com $\boldsymbol{\mu} = X\boldsymbol{\beta}$.

A extensão do modelo linear é realizada em dois passos. No primeiro, com relação a distribuição considerada, não é necessário que a mesma seja normal, podendo ser qualquer distribuição pertencente a família exponencial linear. Para o segundo, embora seja mantida a estrutura de linearidade, a função responsável por relacionar o valor esperado e o vetor de covariáveis pode ser qualquer função ao menos duplamente diferenciável (TURKMAN e SILVA, 2000).

Diante disso, os modelos lineares generalizados (MLG's) podem ser definidos de acordo com os seguintes componentes:

1. *Componente aleatório*

Conhecendo o vetor de covariáveis \mathbf{x}_i as variáveis Y_i são condicionalmente independentes com distribuição pertencente a *família exponencial linear* de acordo com a (2.1) ou a (2.2).

2. *Componente Sistemático*

O valor esperado μ_i está relacionado com com o *preditor linear* $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ de acordo com as relações:

$$g(\mu_i) = \eta_i = \mathbf{z}_i^\top \boldsymbol{\beta},$$

em que :

- $g(\cdot)$ é a *função de ligação*, considerada monótona e ao menos duplamente diferenciável;
- $\boldsymbol{\beta}$ é um vetor de parâmetros com dimensão $p \times 1$;
- \mathbf{z}_i é um vetor de covariáveis com dimensão $p \times 1$.

Além disso, a escolha da função de ligação depende do tipo de resposta e da condução do estudo. Dessa forma, na [Tabela 2](#) são apresentadas as principais *funções de ligação canônicas*, i.e., nos casos em que tais funções o preditor linear coincide com o parâmetro canônico. Dessa forma, $\theta_i = \eta_i$ e implica $\theta_i = \mathbf{z}_i^\top \boldsymbol{\beta}$.

Tabela 2 – Funções de ligação canônicas para alguns membros da Família Exponencial Linear

Distribuição	Normal	Binomial	Poisson	Gama
Ligação	$\mu = \eta$	$\ln(\frac{\mu}{1-\mu}) = \eta$	$\ln(\mu) = \eta$	μ^{-1}

Fonte: elaborado pelo autor.

Segundo Paula (2013), além das ligações canônicas, destacam-se as seguintes ligações:

- Ligação proibito;
- Ligação log-log;
- Ligação de Box-Cox;
- Ligação Aranda- Ordaz.

Dessa forma, de acordo com o teorema da fatoração, quando a ligação é canônica, isto é $\theta_i = \eta_i = \mathbf{z}_i^\top \boldsymbol{\beta}$ e a sua função de log-verossimilhança $L(\boldsymbol{\beta})$ pode ser escrita da seguinte forma:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \phi(y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c(y_i, \phi).$$

Nos casos em que o parâmetro canônico ($\boldsymbol{\theta}$) for correspondente ao preditor linear, ou seja, quando $\theta_i = \eta_i = \sum_{j=1}^p x_{i,j} \beta_j$ e com estatística $S_j = \phi \sum_{i=1}^n Y_i x_{i,j}$ $L(\boldsymbol{\beta})$ pode ser reescrita como:

$$L(\boldsymbol{\beta}) = \sum_{j=1}^p s_j \beta_j - \phi \sum_{i=1}^n b\left(\sum_{j=1}^p x_{i,j} \beta_j\right) + \sum_{i=1}^n c(y_i, \phi).$$

Com relação às ligações canônicas, as vantagens de sua utilização estão no fato de que sendo o parâmetro de escala (ϕ) conhecido, o vetor de parâmetros desconhecido da estrutura linear admite uma estatística suficiente minimal de dimensão fixa e que as mesmas garantem a estrutura côncava de $L(\boldsymbol{\beta})$, facilitando a obtenção de resultados assintóticos (PAULA, 2013).

O processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de β é definido expandindo a função escore \mathbf{U}_β em torno de um valor inicial $\beta^{(0)}$. Em seguida, é aplicado o método escore de Fisher para evitar a existência de uma matriz não positiva definida. Por fim, é possível chegar a um processo iterativo de mínimos quadrados ponderados da seguinte forma:

$$\beta^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)}.$$

de modo que $m = 0, 1, \dots$, e $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$. Nesse sentido, observa-se que \mathbf{z} assume o comportamento de uma variável dependente modificada e \mathbf{W} representa uma matriz de pesos que se modifica a cada passo do processo de iteração. Uma explanação completa do processo é apresentada em Paula (2013), por exemplo.

2.3 Função desvio

Compreende-se que o modelo saturado ($p = n$) é adequado para determinar a qualidade do ajuste de um determinado modelo sob investigação. Isso se dá através da inclusão de uma medida responsável por quantificar o valor da distância entre os valores ajustados $\hat{\beta}$ com esse modelo e dos correspondentes valores de y . Essa métrica de discrepância, entre o modelo saturado e o modelo sob investigação, baseia-se na estatística de teste da razão de verossimilhanças generalizada de Wilks (Wilks, 1938). Logo, a qualidade do ajuste de um MLG, denominada função desvio, pode ser definida como:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y})\}.$$

Alternativamente considerando a densidade (2.1), a equação acima pode ser reescrita como:

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\tilde{\theta}_i) - b(\hat{\theta}_i))\}, \quad (2.3)$$

de modo que $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$ representam, respectivamente, os estimadores de máxima verossimilhança para o modelo com p parâmetros e para o modelo saturado. Além do mais, é facilmente observado que o *desvio* é sempre maior ou igual a zero, decrescendo à medida que covariáveis irão sendo adicionadas ao modelo nulo. Além disso, o mesmo é utilizado para verificar a adequação do modelo sob investigação.

2.4 Testes de hipóteses

De acordo com McCulloch e Searle(2001) e Paula (2013), uma grande variedade de problemas de inferência ligados aos testes de hipóteses sobre o vetor de parâmetros β pode ser descrito da seguinte forma:

$$H_0 : \beta = \beta^0 \text{ versus } H_1 : \beta \neq \beta^0. \quad (2.4)$$

Assim, é possível afirmar a existência de essencialmente quatro estatísticas capazes de testar hipóteses do tipo descrito em (2.4). São elas:

1. de Wald, baseia-se no comportamento normal assintótico do estimador de verossimilhança $\hat{\beta}$.
2. **Teste de Wilks** ou *Teste de razão de verossimilhanças generalizada*, baseia-se na distribuição assintótica da razão de máximo das verossimilhanças de acordo com as hipóteses H_0 e $H_0 \cup H_1$. Além disso, tal estatística pode ser compreendida, no caso do MLG, como a diferença entre duas funções de desvio.
3. *Teste de Rao ou Teste Escore*, baseia-se nas atribuições assintóticas da função de escore.
4. *Teste Gradiente*, por apresentar uma maior flexibilidade, torna-se simples de ser calculada, de modo que não envolve nenhum cálculo matricial (LEMONTE; 2016).

2.5 Técnicas de diagnóstico

A análise de diagnóstico é uma etapa importante na avaliação da qualidade do ajuste de um modelo de regressão. Nesse momento, é possível realizar a identificação de discordâncias entre as suposições estabelecidas para o modelo, bem como o mapeamento de pontos (observações) influentes em determinados aspectos.

2.5.1 Pontos de alavanca

O intuito principal que está contido na definição de **ponto de alavanca** é a avaliação da influência de y_i sobre o próprio valor ajustado \hat{y}_i . Na estrutura definida do MLG para ϕ conhecido a matriz $\partial\hat{y}/\partial y^\top$ pode ser obtida como:

$$\widehat{GL} = \frac{\partial\hat{y}}{\partial y^\top} = \{\mathbf{D}_\beta(-\ddot{\mathbf{L}}_{\beta\beta})^{-1}\ddot{\mathbf{L}}_{\beta y}\}|\hat{\beta},$$

em que $\mathbf{D}_\beta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$, $\ddot{\mathbf{L}}_{\beta\beta} = \partial^2 L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$, $\ddot{\mathbf{L}}_{\beta y} = \partial^2 L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial y^\top$.

Com a substituição de $-\ddot{\mathbf{L}}_{\beta\beta}$ pelo seu valor esperado $\phi(\mathbf{X}^\top \mathbf{W} \mathbf{X})$, é possível obter de maneira aproximada:

$$\widehat{GL} = \hat{\mathbf{N}} \mathbf{X} (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \hat{\mathbf{N}}.$$

Finalmente, o elemento \widehat{GL}_{ii} pode ser escrito como:

$$\widehat{GL}_{ii} = \hat{\omega}_i x_i^\top (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} x_i,$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Para mais informações, recomenda-se a leitura Hoaglin e Welsch (1978), Cook e Weisberg (1982), Emerson, Hoaglin e Kempthorne (1984), St. Laurent e Cook (1992) e Wei, Hu e Fung (1998).

2.5.2 Resíduos

É possível considerar o resíduo ordinário da solução de mínimos quadrados da regressão linear ponderada contra \mathbf{X} . A partir disso, o resíduo padronizado pode ser definido a seguir:

$$t_{Si} = \frac{\phi^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

nesse caso, diz-se que h_{ii} é o i -ésimo elemento da diagonal principal da matriz \mathbf{H} .

De acordo com Paula (2013), os resíduos mais utilizados em modelos lineares generalizados são definidos a partir dos componentes da função desvio. Assim, sua forma padronizada é:

$$D_{Si} = \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{(1 - \hat{h}_{ii})}} = \frac{\phi^{1/2} d(y_i; \hat{\mu}_i)}{\sqrt{(1 - \hat{h}_{ii})}}.$$

2.5.3 Envelope Simulado

Desenvolvido por Atkinson (1985), o gráfico de probabilidade meio-normal com envelope simulado é uma outra técnica aplicada na avaliação do ajuste do modelo.

Nele, dispomos o i -ésimo valor ordenado dos resíduos padronizados r_i^* , contra o valor esperado da estatística de ordem, em valor absoluto, da normal padrão. O modelo pode ser descrito como:

$$E(|Z_{(i)}|) \approx \Phi^{-1} \left(\frac{i + N - 1/8}{2n + 1/2} \right).$$

Diz-se que, em particular, o gráfico é indicado para detecção simultânea de pontos aberrantes e/ou influentes. Como sugestão de aprofundamento, o leitor pode consultar Neter et al.(1996); Tan, Qu e Kutner (1997) e Venezuela (2003).

3 MODELO DE REGRESSÃO BETA

É comum em diversas situações de interesse a necessidade da modelagem de variáveis do tipo proporções ou taxas, como por exemplo: a proporção de casas que possuem serviços de TV a cabo, a porcentagem da renda gasta com combustível e alimentação, etc. Além disso, em diversos casos essas variáveis podem estar contidas no intervalo $(0,1)$. Kieschnick e McCullough (2003) apresentam essas variáveis classificadas em dois tipos. No primeiro, observa-se que os valores podem ser modelados por meio de um processo contínuo, enquanto que no segundo, é possível definir um intervalo fechado $[0,1]$ e seguem uma distribuição mista discreta-contínua. Ao analisar o primeiro tipo, é possível elencar, em linha com os autores, sete modelos: o modelo normal linear, o modelo logito, o modelo normal censurado, o modelo normal não linear, os modelos de regressão baseados na distribuição beta, a distribuição simplex e o modelo de quase-verossimilhança.

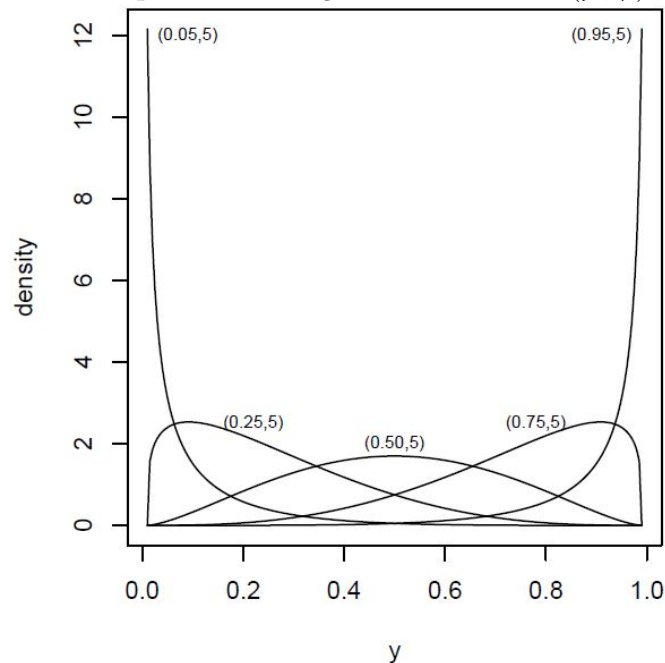
Além disso, em um estudo comparativo realizado por Kieschnick e McCullough (2003), foi possível indicar que, dentre os modelos acima citados, os modelos de regressão beta, simples e quase-verossimilhança são os mais adequados para modelagem de dados de proporções em que os mesmos apresentam respostas inferenciais semelhantes. Finalmente, o trabalho reforça que, dentre os três modelos, uma variação do Critério de Informação de Akaike (AIC) evidencia a proficiência do modelo de regressão beta.

O modelo de regressão baseado na distribuição beta, cuja contribuição presente em Ferrari e Cribari-Neto (2004) corresponde à reparametrização e ajuste a um formato similar a de um MLG, busca modelar situações em que a variável resposta é contínua, definida em um intervalo restrito $(0,1)$ e possui relação com outras variáveis por meio de uma estrutura de regressão. Além disso, o modelo proposto é fundamentado na suposição de que a variável resposta assume uma distribuição beta fazendo uso de uma parametrização da *lei beta* que é indexada pela média e um parâmetro de dispersão, ou seja, o modelo supõe que a distribuição da variável resposta é beta e sua média está relacionada a um preditor linear a partir de uma função de ligação. De acordo com essa definição, em diversos aspectos, a regressão beta é bastante semelhante aos modelos lineares generalizados (MLG's).

3.1 Distribuição Beta

Como dito anteriormente, a distribuição beta é muito flexível, assumindo uma grande variedade de formatos simétricos e assimétricos, incluindo os formatos de “J”, “U” e “J” invertido. Isso pode ser verificado pelas diversas densidades da distribuição apresentadas na Figura (1). A função densidade da distribuição beta, indexada pelos parâmetros $p > 0$

Figura 1 – Densidades beta para combinações diferentes de (μ, ϕ) .



Fonte: Ferrari e Cribari-Neto (2004)

e $q > 0$, é definida como:

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad (3.1)$$

em que $\Gamma(p)$ é definida por:

$$\Gamma(p) = \int_0^{\infty} y^{p-1} e^{-y} dy.$$

A média e a variância de y são, respectivamente, definidas por:

$$E(y) = p/p+q$$

e

$$\text{Var}(y) = \frac{pq}{(p+q)^2(p+q+1)}.$$

3.2 Definição do modelo de regressão beta

O modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004) apresenta uma reparametrização da distribuição beta, considerando a média da resposta e um parâmetro de dispersão. Essa reparametrização, em linha com Venezuela (2008), se dá devido ao fato de que a estrutura original (3.1) apresenta, apenas, parâmetros de forma. Assim, com a reparametrização é possível a obtenção de métricas de posição e precisão, representados por μ e ϕ , respectivamente. Diante disso, a (3.1) pode ser reescrita como:

$$f(y; \mu \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu \phi) \Gamma((1 - \mu) \phi)} y^{\mu \phi - 1} (1 - y)^{(1 - \mu) \phi - 1}. \quad (3.2)$$

Com isso, a média e a variância são dadas por:

$$E(y) = \mu$$

e

$$\text{Var}(y) = V(\mu) / 1 + \phi.$$

em que $V(\mu) = \mu(1 - \mu)$ é a função de variância, μ é a média da variável resposta e ϕ pode ser compreendido como uma métrica de precisão (dispersão) no sentido que, para μ fixado, quanto maior for o valor de ϕ , menor a variância de y .

Supondo que y_1, \dots, y_n sejam variáveis aleatórias independentes, em que cada y_t , $t = 1, \dots, n$, segue uma distribuição beta com parâmetros μ_t e ϕ , cuja densidade é definida de acordo com a (3.2). Define-se que o modelo de regressão beta é definido pela (3.2) e pelo componente sistemático

$$g(\mu_t) = \sum_{i=1}^k x_{ti} \beta_i = \eta_t, \quad (3.3)$$

em que $\eta_t = x_{ti}^T \beta_i$ é o preditor linear, β é um vetor de parâmetros desconhecidos a serem estimados, x_t^T é o vetor de variáveis explicativas associado ao t -ésimo indivíduo, assumindo-as fixas e conhecidas. A função $g(\cdot)$ é dita função de ligação (estrutura estritamente monótona e duplamente diferenciável).

Uma função amplamente utilizada é a logito, dado que é possível obter uma fácil interpretação dos parâmetros da regressão. De acordo com a equação acima

$$g(\mu_t) = \log \left(\frac{\mu_t}{1 - \mu_t} \right) = x_t^T \beta.$$

Assim,

$$\left(\frac{\mu_t}{1-\mu_t}\right) = \exp(x_t^\top \beta).$$

Além disso, pode-se escrever:

$$\mu_t = \frac{\exp(x_t^\top \beta)}{1 + \exp(x_t^\top \beta)}. \quad (3.4)$$

Nota-se que (3.4) é apresentada como a função inversa de $g(\mu_t)$. Quando utiliza-se a função de ligação logito os parâmetros da regressão podem ser compreendidos em termos de razão de chances (*odds ratio*). Com isso, assumindo que o valor da i -ésima variável regressora é acrescido em m unidades, ao passo que as demais variáveis dependentes do modelo em estudo permanecem inalteradas. Além disso, denotando μ^\dagger como o valor médio de y sob o novo valor das covariáveis, de modo que μ seja a média de y sob o valor original das covariáveis, dessa forma tem-se:

$$\frac{\mu^\dagger}{1-\mu^\dagger} = \exp(x_{t1}\beta_1 + \dots + (x_{ti} + m)\beta_i + \dots + x_{tk}\beta_k).$$

Logo, a razão de chances é dada por:

$$\exp(m\beta_i) = \frac{\mu^\dagger/(1-\mu^\dagger)}{\mu/(1-\mu)}.$$

3.3 Resíduos

Tendo em vista que os resíduos são a base primordial para a realização da análise de diagnóstico, um fase importante no processo de regressão para a validação adequada do modelo, a presente Seção busca apresentar os resíduos propostos por Espinheira et al.(2008) e Ferrari e Cribari-Neto (2004) tomando por base o modelo apresentado em (3.2).

3.3.1 Resíduo ordinário

Proposto por Ferrari e Cribari-Neto (2004), o resíduo ordinário padronizado pode ser expresso como:

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{\text{Var}}(y_t)}},$$

em que $\widehat{\text{Var}}(y_t) = \hat{\mu}(1-\hat{\mu})/(1+\hat{\phi})$, $\hat{\mu} = g^{-1}(x_t^\top \hat{\beta})$. Além disso, $\hat{\beta}$ e $\hat{\phi}$ são apresentados como os estimadores de máxima verossimilhança de β e ϕ , respectivamente.

3.3.2 Resíduo ponderado padronizado 1

Este resíduo foi estruturado tomando por base o método iterativo *scoring* de Fisher para a estimação de β , com ϕ é conhecido. A partir disso, dado o processo iterativo pode ser reescrito em termos de regressão de mínimos quadrados ponderados, Espinheira et al.(2008) apresentam a definição de resíduo ponderado como:

$$r_t^* = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\phi(v_t)}},$$

com $v_t = \psi'(\hat{\mu}_t\phi) + \psi'((1 - \hat{\mu}_t)\phi)$. Além disso, em linha com os autores supracitados, como $\text{Var}(y_t^*) = \psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi)$ o resíduo ponderado padronizado 1 é dado por:

$$r_t^p = \phi^{1/2} r_t^* = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{v_t}}.$$

3.3.3 Resíduo ponderado padronizado 2

O resíduo ponderado padronizado 2 é outra métrica desenvolvido por Espinheira et al.(2008), definido como:

$$r_t^p p = \frac{r_t^*}{\sqrt{\phi^{-1}(1 - h_{tt})}} = \frac{r_t^p}{\sqrt{(1 - h_{tt})}} = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{v(1 - h_{tt})}},$$

em que h_{tt} é compreendido como o t -ésimo elemento da diagonal de \mathbf{H} .

Com isso, deve-se considerar ϕ como fixo. Em termos práticos, para o cálculo de tais resíduos, deverá ser considerada a substituição de ϕ pela sua estimativa de máxima verossimilhança de $\hat{\phi}$.

Outro ponto de destaque é o fato de que o resíduo ponderado padronizado 2 apresenta-se como o mais eficiente na identificação de observações que possuem alta influência nas estimativas dos parâmetros. Em Espinheira et al.(2008) os autores mostram, por meio de um estudo com dados simulados, em que foi inserida uma observação atípica, que tal resíduo é capaz de apontar um valor claramente atípico, sendo, simultaneamente, um ponto *leverage* e *outlier*.

3.4 Métodos de diagnóstico

A importância da utilização de métricas de diagnóstico esta pautada no fato de que a aplicação da mesma em um modelo de regressão permite verificar possíveis distanciamentos entre as suposições feitas para o modelo em estudo, além de identificar observações extremas que possam interferir desproporcionalmente nos resultados do ajuste do modelo.

Dentro desse cenário, a presente Seção apresenta algumas métricas construídas tomando por base o método iterativo *scoring* de Fisher para a estimação dos parâmetros de regressão β , no que se refere à modelagem através do modelo de regressão beta.

Diante do exposto, o método iterativo *scoring* de Fisher para estimar β pode ser reexpresso como a solução de mínimos quadrados ponderados da forma

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\mathbf{X}^\top \mathbf{P}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{T}^{(m)} \mathbf{b}^{(m)}, \quad (3.5)$$

com $\mathbf{T} = \text{diag} 1/g'(\mu_1), \dots, 1/g'(\mu_n)$, $\mathbf{P} = \mathbf{W}$ e $\mathbf{b} = y^* - \mu^*$.

Com isso, (3.5) pode ser reescrito como o processo iterativo de mínimos quadrados ponderados apresentado a seguir:

$$\hat{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{P}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P}^{(m)} \mathbf{z}^{(m)},$$

em que $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{P}^{-1(m)} \mathbf{T}^{(m)} \mathbf{b}^{(m)}$, com $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^\top = \mathbf{X}\beta$. No processo de convergência, tem-se

$$\hat{\beta} = (\mathbf{X}^\top \hat{\mathbf{P}} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{P}} \mathbf{z}, \quad (3.6)$$

com $\mathbf{z} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{P}}^{-1} \hat{\mathbf{T}} \hat{\mathbf{b}}$, de modo que $\hat{\boldsymbol{\eta}}, \hat{\mathbf{P}}, \hat{\mathbf{T}}$ e $\hat{\mathbf{b}}$ são os valores de $\boldsymbol{\eta}, \mathbf{P}, \mathbf{T}$ e \mathbf{b} quando avaliados no estimado de máxima verossimilhança. Vale destacar que $\hat{\beta}$ apresentado em (3.6) pode ser compreendido como a estimativa de β fornecida pela regressão de uma variável dependente modificada, \mathbf{Z} , com uma matriz de especificação \mathbf{X} e uma matriz de pesos \mathbf{P} .

3.4.1 Ponto de alavanca

Tomando por base a equação (3.6), $\hat{\beta}$ pode ser definido como a solução de mínimos quadrados da regressão normal linear de $\hat{\mathbf{P}}^{1/2} \mathbf{z}$ possuindo como matriz de especificação $\hat{\mathbf{P}}^{1/2} \mathbf{X}$. Logo, pode-se concluir que o resíduo ordinário, apresentado como a diferença entre

os valores observados e ajustados, é dado por:

$$r_O = (I - H)\hat{P}^{1/2}z,$$

de modo que \mathbf{I} corresponde à matriz identidade e \mathbf{H} uma matriz diagonal descrita a seguir.

$$H = \hat{P}^{1/2}X(X^\top \hat{P}X)^{-1}X^\top \hat{P}^{1/2}.$$

Diante do exposto, admitindo que $\hat{P}^{1/2}z$ atua como vetor resposta, pode-se considerar \mathbf{H} como a matriz de projeção ortogonal, ou matriz chapéu. Além disso, é possível a utilização dos elementos da diagonal principal de \mathbf{H} , h_{tt} na identificação de pontos de alavanca. Tal processo é apresentado em Paula (2013) para modelos lineares generalizados(MLG's) e em Venezuela (2003), Venezuela, Botter e Sandoval (2007) e Paula (2013) no caso de MLG's com presença de medidas repetidas.

Ainda nesse cenário, um ponto de alavanca apresenta um perfil diferente dos demais em relação aos valores apresentados pelas variáveis explicativas. Nesse caso, na prática, a construção do gráfico de h_{tt} versus o índice das observações t , $t=1, \dots, n$ auxilia na identificação de possíveis pontos de alavanca. Toda via, é importante frisar que os valores de h_{tt} dependem da matriz de pesos, ou seja, observações com altos valores de h_{tt} nem sempre serão pontos de alavanca. Tal fato evidencia a necessidade de cautela na realização da análise de tais pontos.

3.4.2 *Distância de Cook*

Tendo em vista que um ponto influente é aquele que apresenta um peso desproporcional no valor das estimativas dos parâmetros de um dado modelo. Além disso, o mesmo apresenta um perfil diferente dos demais no que se refere ao valor da variável resposta e apresenta um valor elevado na matriz de projeção \mathbf{H} .

Com isso, Cook (1977) propôs uma métrica de influência de cada observação nas estimativas dos parâmetros. A mesma é conhecida como *distância de Cook* (DC) e pode ser definida a seguir:

$$DC_t = \frac{1}{k}(\hat{\beta} - \hat{\beta}_{(t)})^\top \mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X}(\hat{\beta} - \hat{\beta}_{(t)})$$

3.4.3 Gráfico de probabilidade meio-normal com envelope simulado

Desenvolvido por Atkinson (1985), o gráfico de probabilidade meio-normal com envelope simulado é uma outra técnica aplicada na avaliação do ajuste do modelo. Construído com base nos resíduos padronizados, o mesmo apresenta-se como uma ferramenta de diagnóstico bastante útil na avaliação do ajuste do modelo. Além disso, de acordo com Neter et. al (1996) tal gráfico pode utilizado mesmo quando os resíduos não apresentam uma distribuição normal. O modelo pode ser descrito como:

$$E(|Z_{(i)}|) \approx \Phi^{-1} \left(\frac{i + N - 1/8}{2n + 1/2} \right). \quad (3.7)$$

A seguir é apresentado, conforme apresentado em Ferrari e Cribari-Neto (2004), o algoritmo para a construção do gráfico de probabilidade meio-normal com envelope simulado:

ETAPAS PARA A ELABORAÇÃO DO GRÁFICO DE PROBABILIDADE MEIO-NORMAL COM ENVELOPE SIMULADO

- 1:** Para cada unidade experimental i ; $i = 1, \dots, t_i$, deverá ser simulado um vetor de respostas de tamanho t_i , observando o vetor de médias e a matriz de covariâncias ajustados aos dados originais.
- 2:** O mesmo modelo, utilizado para o ajuste dos dados originais, deverá ser aplicado ao ajuste das respostas simuladas no passo 1.
- 3:** Realizar o ajuste do modelo à amostra gerada e calcular os valores absolutos ordenados dos resíduos;
- 4:** Repetir os três primeiros passos mais 24 vezes, número de réplicas sugerido por Tan, Qu e Kutner (1997).
- 5:** Determinar o mínimo, a mediana e o máximo dos menores valores absolutos dos resíduos padronizados de todas as simulações.
- 6:** Organizar em um gráfico os valores mínimos, medianos e máximos obtidos nos passos supracitados versus os correspondentes valores esperados em (3.7). Após isso, associar os pontos referente aos valores mínimos, medianos e máximos.

A partir do processo descrito acima, a ocorrência de elevados desvios dos pontos em relação a mediana dos valores simulados ou a ocorrência de pontos próximos ou fora dos limites da banda de simulação são indicativos de que o modelo pode não ser adequado.

Ademais, caso existam, os pontos aberrantes devem aparecer no topo direito do gráfico separado dos demais pontos.

Mais detalhes sobre sua construção do gráfico de probabilidade meio-normal com envelope simulado são apresentados em Neter et. al (1996); Tan, Qu e Kutner (1997) e em Venezuela (2003) poderão ser vistos os algoritmos para gerar variáveis aleatórias correlacionadas das distribuições Binomial, Poisson e Gama.

3.5 Aplicação dos modelos de regressão beta através do uso do software R

A implementação do modelo de regressão beta, apresentado por Ferrari e Cribari-Neto (2004) e estendida por Simas, Barreto-Souza e Rocha (2010), é apresentada no `software R` por meio do pacote `betareg` (). Com isso, o pacote visa uma modelagem beta-distribuída com parametrização usando parâmetro de média e precisão, de modo que a média (μ) está ligada, como nos modelos lineares generalizados (GLMs), às respostas através de uma função de ligação e um preditor linear. Além disso, o parâmetro de precisão (ϕ) pode ser ligado a outro conjunto de regressores através de um segundo função de ligação, resultando em um modelo com dispersão variável.

No que tange à estimação, a mesma ocorre via função de máxima verossimilhança via `optim()` utilizando gradientes analíticos e (por padrão) iniciando valores de uma regressão linear auxiliar da resposta transformada. Além disso, a principal função e ajuste de modelo é `betareg` () que possui uma abordagem bastante simples para implementação da máxima verossimilhança em modelos de regressão no `software R`. Os comandos `formula` e `data` são utilizados para a definição do modelo e especificação dos dados, respectivamente. Os argumentos de `betareg` () são:

```
betareg(formula, data, subset, na.action, weights, offset, link = "logit",
link.phi = NULL, control = betareg.control(...), model = TRUE, y = TRUE, x
= FALSE, ...).
```

Recomenda-se que para mais informações e exemplos o leitor interessado consulte o manual de referencia do pacote `betareg` existente no `software R` (R Core Team, 2018).

Buscando apresentar uma aplicação do modelo de regressão beta e perpassando pelos assunto apresentados neste Capítulo, a seguir será apresentado um exemplo prático desenvolvido por Atkinson (1985) e com aplicação realizada por meio do pacote `betareg`.

3.5.1 Aplicação

A presente aplicação foi anteriormente apresentado em Atkinson (1985) e consiste na análise dos dados sobre gasolina de Prater (1956), de modo que a variável dependente (y) é compreendida como a proporção de petróleo convertida em gasolina após o processo de fracionamento de 32 observações. Além disso, as variáveis independentes são: tipo do petróleo e temperatura.

A partir disso, considera-se que as observações y_1, \dots, y_{32} são independentes e possuem distribuição beta com média μ_t , ($t= 1, \dots, 32$), com parâmetro de precisão ϕ desconhecido. O modelo é descrito da seguinte forma:

$$g(\mu_t) = \beta_0 + \beta_1 x_1 + \dots + \beta_{10} x_{10} \quad (3.8)$$

O ajuste do modelo de regressão especificado acima foi realizado considerando a função de ligação logito. Com isso, a partir da tabela a seguir, é apresentado os resultados da estimação. Assim, é possível afirmar que todas as variáveis são significativas para o modelo.

Tabela 3 – Valores estimados do modelo de regressão beta

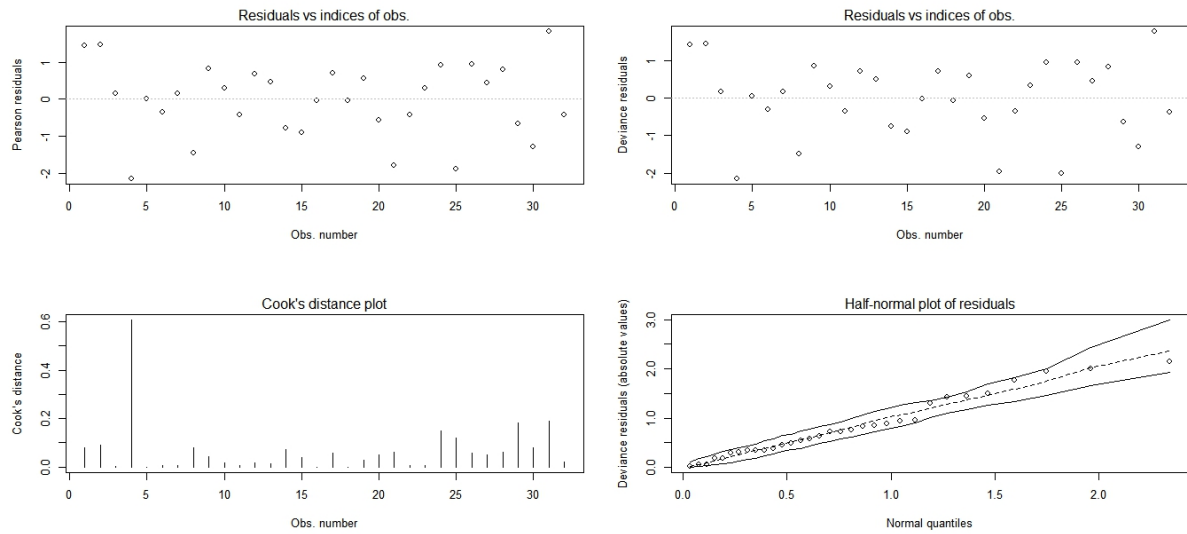
Parâmetro	Valor Estimado	Erro Padrão	valor - p
β_0	-6,16	0,18	<0,001
β_1	1,73	0,10	<0,001
β_2	1,32	0,12	<0,001
β_3	1,57	0,12	<0,001
β_4	1,06	0,10	<0,001
β_5	1,13	0,10	<0,001
β_6	1,04	0,11	<0,001
β_7	0,54	0,11	<0,001
β_8	0,50	0,11	<0,001
β_9	0,39	0,11	0,001*
β_{10}	0,01	0,00	<0,001
ϕ	440,30	110,00	<0,001

Fonte: Elaborado pelo autor

Seguindo a análise, é possível verificar as medidas de diagnóstico com o objetivo de identificar algum ponto de comportamento atípico.

A figura a seguir sugere que não há presença de pontos de alavanca. Contudo, apenas a observação 4 é apresentada como possível ponto influente. Além disso, gráfico de probabilidade meio-normal com envelope simulado (a 95%) evidencia que os pontos estão

Figura 2 – Gráficos de diagnóstico do modelo de regressão beta ajustado para os dados.



Fonte: Elaborado pelo autor

concentrados dentro da banda de confiança, sugerindo que o ajuste dos dados de acordo com o modelo beta é adequado.

Com a utilização da medida de variação percentual descrita a seguir, em linha com Miyashiro (2008), pode-se avaliar a capacidade da influência exercida pelas observações distintas das demais sobre as estimativas dos parâmetros de dispersão.

$$VP_{(s)} = \frac{\hat{\theta}_{-pontos(s)} - \hat{\theta}}{\hat{\theta}} * 100\%, \quad (3.9)$$

onde, $\hat{\theta}_{-pontos(s)}$ é apresentado como a estimativa de máxima verossimilhança de θ sem o(s) ponto(s) com comportamento(s) distinto(s) dos demais e $\hat{\theta}$ é a estimativa de θ com todos os pontos do modelo.

O reajuste do modelo sem a observação 4 pode ser visto na tabela a seguir. Dado a presença de 32 observações, implica-se que cada unidade observada deve influenciar em 3,13% nas estimativas dos parâmetros.

Tabela 4 – Valores estimados do modelo de regressão beta (sem a observação 4)

Parâmetro	Valor Estimado	Valor Estimado Sem o Ponto 4	VP
β_0	-6,16	-6,35	3,20%
β_1	1,73	1,89	9,21%
β_2	1,32	1,37	3,61%
β_3	1,57	1,63	3,36%
β_4	1,06	1,08	1,98%
β_5	1,13	1,15	1,57%
β_6	1,04	1,06	1,68%
β_7	0,54	0,57	3,95%
β_8	0,50	0,50	0,97%
β_9	0,39	0,39	-0,16%
β_{10}	0,01	0,012	4,55%
ϕ	440,30	577,80	31,23%

Fonte: Elaborado pelo autor

A partir disso, observa-se que não há grande alteração na estimação dos parâmetros e que a maior variação ocorre no parâmetro de precisão (ϕ). Assim, conclui-se que a eliminação do ponto 4 não muda as conclusões inferenciais.

No próximo capítulo do presente estudo serão introduzidos os conceitos referentes às Equações de Estimação Generalizadas, conforme propostas por Zeger e Liang (1986).

4 EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS

Tomando como base os MLGs, é possível afirmar que uma de suas principais suposições é a independência entre os sujeitos em observação. Um exemplo seria a construção de um modelo para o estudo do comportamento da mortalidade infantil assumindo a suposição que o conhecimento sobre ocorrência do óbito em uma criança não fornece nenhuma informação a respeito do estado de outra criança nesse estudo. Contudo, existem determinados tipos de estudos em que a hipótese de independência não se faz coerente, revelando, assim, que uma correlação entre os objetos em estudo pode ser considerada.

No exemplo anteriormente citado, caso a amostra contenha irmãos, é plausível supor que a hipótese de independência não seja adequada. Isso se dá, de acordo com Alexander et al.(1998), pelo fato de que crianças expostas às mesmas características maternas e a condições de gestação semelhantes, apresentam comportamentos mais parecidos do que os de gestações diferentes. Logo, de acordo com o exemplo apresentado, os sujeitos que não pertençam à mesma família seriam independentes, em oposição aqueles que são irmãos. Nesse sentido, a dependência de observações pode ocorrer em casos em que é possível identificar relações de agrupamento entre os elementos do objeto de estudo.

Além disso, verifica-se a possibilidade de ocorrência de correlação em pesquisas com dados que envolvam medidas repetidas, ou seja, estudos em que a estratégia de coleta seja pautada na observação de uma ou mais variáveis resposta de um grupo amostral de acordo com algum requisito avaliativo. Uma ilustração disso seria analisar o efeito de um repelente em um sujeito ao longo de um tratamento, com isso testes seriam realizados diariamente ao longo de dois meses em um grupo de voluntários. O grupo de medidas repetidas pode ser compreendido como ensaios em que o requisito de avaliação não pode ser aleatorizado, a exemplo, o tempo, gerando a ocorrência de dados longitudinais; estudos de corte, especialmente realizados em Bioestatística e dados em painel, comumente observados em estudos econométricos.

Diante do exposto, a literatura evidencia a existência de abordagens para para a análise de informações agrupadas. As principais são as Equações de Estimação Generalizadas (EEG's) e os Modelos de Efeitos Mistos. Trazendo o foco para as EEG's, o método foi proposto por Zeger e Liang (1986) com o objetivo de estimar parâmetros de regressão especialmente quando os dados estão correlacionados. Os autores basearam-se

nos MLG's, incluindo uma estrutura de correlação de “trabalho” (“working correlation matrix”) entre as observações para a obtenção de estimativas consistentes e não viciadas.

Em continuidade ao trabalho proposto inicialmente (conhecido como EEG1), Prentice e Zhao (1991) estenderam o modelo, permitindo a avaliação dos parâmetros de correlação por meio das equações de estimação. Essa contribuição passou a ser conhecida como EEG2. Song et al. (2004) apresentaram a modelagem do parâmetro de dispersão por meio das equações de estimação. Com isso, tem-se o surgimento da EEG3. Nesse sentido, as Equações de Estimação Generalizadas apresentam uma alta performance em sua estrutura funcional, visto que as mesmas ao incorporarem a estrutura de correlação de trabalho, são capazes de fornecer estimadores consistentes e assintoticamente normais para os parâmetros dos modelos em que tal procedimento estatístico é utilizado. Além do mais, as EEG's viabilizam uma flexibilidade na distribuição da variável resposta.

Para os leitores interessados na obtenção de mais detalhes, recomenda-se as seguintes estudos: Godambe (1960), Godambe (1991), Godambe e Kale (1991), Sen e Singer (1993), Jørgensen e Labouriau (1994), Artes (1997), Godambe (1997), Venezuela (2003), Artes e Botter (2005), Sen et al. (2009), Freitas (2018).

4.1 Funções de estimação

No que se refere às funções de estimação, diz-se que elas são estabelecidas através de uma relação funcional dos dados (\mathbf{y}) e dos parâmetros de interesse ($\boldsymbol{\theta}$). É relevante evidenciar que no estudo de tais funções o estabelecimento de premissas que garantam boas propriedades dos estimadores dos parâmetros envolvidos é um fator primordial.

Tomando por base a definição anteriormente citada, é possível afirmar que uma função $\boldsymbol{\psi}$ é uma função de estimação do vetor aleatório (\mathbf{y}) e dos dos parâmetros de interesse ($\boldsymbol{\theta}$) desde que para cada $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi}(\boldsymbol{\theta}; \mathbf{y}) = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n)^\top$ é uma variável aleatória, de forma que $\Theta \subseteq \mathbb{R}^p$ é o espaço paramétrico. Para essa definição é considerado que as situações em que Θ é pequeno e possui dimensão finita p . Partindo para uma amostra de n vetores aleatórios independentes $y_i = (y_{i1}, \dots, y_{ik})^\top$, $i = 1, \dots, n$ em que cada vetor esteja relacionado com uma função de estimação $\boldsymbol{\psi}_i$, logo uma função de estimação da amostra $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ pode ser definida como:

$$\boldsymbol{\Psi}_n(\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n \boldsymbol{\psi}_i(y_i; \boldsymbol{\theta}),$$

em que $\mathbf{y} = (y_1^\top, \dots, y_n^\top)^\top$. A função de estimação apresentada está restrita apenas às funções cujas raízes são estimadores dos parâmetros de interesse, ou seja,

$$\Psi_n(\mathbf{y}; \hat{\boldsymbol{\theta}}) = \mathbf{0}. \quad (4.1)$$

A partir disso, dispõe-se as seguintes definições sobre a função de estimação:

Definição 1: Admitindo y_i, \dots, y_n como uma amostra aleatória com $E(y_i) = \mu_i(\boldsymbol{\theta})$, com μ_i duplamente diferenciável com relação a $\boldsymbol{\theta}$ e $\text{Var}(y_i) = \sigma^2$, é possível definir que:

$$\Psi_n(\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [y_i - \mu_i(\boldsymbol{\theta})].$$

Diante disso, define-se que a função de estimação apresentada anteriormente é chamada de Equação de Estimação e suas raízes são os estimadores de mínimos quadrados de $\boldsymbol{\theta}$.

Definição 2: Define-se que $\Psi(\boldsymbol{\theta})$ e $\Phi(\boldsymbol{\theta})$ são funções de estimação equivalentes se:

$$\Psi(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})\Phi(\boldsymbol{\theta}),$$

em que $\mathbf{C}(\boldsymbol{\theta})$ é uma matriz quadrada de posto completo.

Dado que $\mathbf{C}(\boldsymbol{\theta})$ é de posto completo é possível afirmar que caso $\hat{\boldsymbol{\theta}}_n$ seja a raiz de uma função de estimação, ela será, também, raiz de todas as funções de estimação equivalentes. Vale ressaltar que uma função particular possui infinitas funções de estimação equivalentes.

Definição 3: Diz-se que a função $\Psi(\boldsymbol{\theta})$ como não viciada se:

$$E_{\boldsymbol{\theta}}[\Psi_n(\boldsymbol{\theta})] = \mathbf{0}, \forall \boldsymbol{\theta} \in \Theta.$$

Como consequência, se todas as funções de estimação ϕ_i forem não viciadas, logo Ψ_n também não será.

Definição 4: Seja Ψ_n uma função de estimação não viciada, sua matriz de variabilidade e sensibilidade, são dadas, respectivamente, por:

$$V_{\Psi}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\Psi_n(\boldsymbol{\theta})\Psi_n^\top(\boldsymbol{\theta})], \quad (4.2)$$

$$S_{\Psi}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial \boldsymbol{\theta}^{\top}} \Phi_n(\boldsymbol{\theta})\right]. \quad (4.3)$$

Dessa forma, é dito que as matrizes quadradas são de ordem p .

Definição 5: Admitindo $(\Omega, \mathcal{A}, \mathcal{P})$ como sendo um espaço de probabilidade, $\Omega \subset \mathbb{R}$ expresso como um espaço amostral sobre o qual define-se $\mathcal{P} = \{P_{\boldsymbol{\theta}}: \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, para algum $p \in \mathbb{N}$. Com isso, uma função de estimação $\Psi_n(\boldsymbol{\theta}) : \Omega \times \Theta \rightarrow \mathbb{R}^p$ é dita regular se $\forall \boldsymbol{\theta} \in \Theta$ e $i, j=1, \dots, p$,

1. $\Psi_n(\boldsymbol{\theta})$ é uma função de estimação não viciada;
2. A derivada $\frac{\partial \Psi_n(\boldsymbol{\theta})}{\partial \theta_i}$ existe e é contínua quase certamente $\forall \mathbf{y} \in \Omega$;
3. É possível permutar o sinal de integração e derivação de acordo com a seguinte forma:

$$\frac{\partial}{\partial \theta_i} \int_{\Omega} \Psi_n(\boldsymbol{\theta}; \mathbf{y}) dP_{\boldsymbol{\theta}} = \int_{\Omega} \frac{\partial}{\partial \theta_i} [\Psi_n(\boldsymbol{\theta}; \mathbf{y})] dP_{\boldsymbol{\theta}}.$$

Essa possibilidade se dá devido ao fato de que o $\Psi_n(\boldsymbol{\theta})$ pode ser integrável como função de \mathbf{y} para cada θ_i , dado a Propriedade 2 e assumindo que abrange uma função integrável;

4. $E_{\boldsymbol{\theta}}[\Psi_i(\boldsymbol{\theta})\Psi_j(\boldsymbol{\theta})] \in \mathbb{R}$ e $V_{\Psi}(\boldsymbol{\theta})$ é positiva definida;
5. $E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial \theta_i} \Psi_i(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_k} \Psi_j(\boldsymbol{\theta})\right] \in \mathbb{R}$ em que $i, k = 1, \dots, p$ e $S_{\Psi}(\boldsymbol{\theta})$ é não singular.

Definição 6: A matriz de informação de Godambe de $\boldsymbol{\theta}$ relacionada a uma função de estimação regular Ψ_n é definida por:

$$J_{\Psi}(\boldsymbol{\theta}) = S_{\Psi}^{\top}(\boldsymbol{\theta}) V_{\Psi}^{-1}(\boldsymbol{\theta}) S_{\Psi}(\boldsymbol{\theta}).$$

Por incorporar as informações sobre a variabilidade dos estimadores, a matriz de informação de Godambe possui a mesma função da matriz de informação de Fisher no contexto do uso da função score para a obtenção do estimador de máxima verossimilhança.

Definição 7: De acordo com Crowder (1987), uma classe de funções de estimação é dita ser aditiva ou linear, com $Q_i(\boldsymbol{\theta})$, $i=1, \dots, n$ definido como sendo matrizes não estocásticas e $u_i = u_i(\mathbf{y}_i; \boldsymbol{\theta})$ vetores com média zero mutualmente independentes, se :

$$\ell(u) = \left\{ \Psi_n \in \mathfrak{R} : \Psi_n(\boldsymbol{\theta}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}) u_i(\mathbf{y}_i; \boldsymbol{\theta}) \right\}.$$

Definição 8: A condição para que uma função de estimação regular seja considerada ótima é que as raízes da mesma possuam variancia assintótica mínima. De acordo com Crowder (1987), a função de estimação ótima da classe de funções de estimação lineares é dada por:

$$\Psi_n^*(\theta) = \sum_{i=1}^n Q_i^*(\theta) u_i(y_i; \theta), \quad (4.4)$$

de modo que

$$Q_i^*(\theta) = E \left(\frac{\partial u_i}{\partial \theta^\top} \right)^\top \text{Cov}(u_i)^{-1},$$

com

$$\text{Cov}(u_i) = \text{diag} \left\{ \mathbf{Var}(u_i)^{1/2} \right\} \mathbf{R}(u_i) \text{diag} \left\{ \mathbf{Var}(u_i)^{1/2} \right\}, \quad (4.5)$$

sendo $\mathbf{R}(u_i)$ a matriz de correlação autêntica de u_i , para $i = 1, \dots, n$.

Vale ressaltar que a Definição 2 também vale para garantir otimalidade de uma função de estimação. No que se refere às condições que garantem a normalidade assintótica dos estimadores obtidos a partir das funções de estimação regulares, é apresentado os teoremas abaixo:

Teorema 1: De acordo com Jørgensen e Labouriau (1994), assumindo que $\Psi : \chi \times \Theta \rightarrow \mathbb{R}^p$ seja um função de estimação regular e $\{\hat{\theta}_n\}_{n \geq 1}$ uma sequência de estimadores satisfazendo (4.1), admitindo que existe $\theta \in \Theta$ de forma que:

$$\hat{\theta} \xrightarrow{\mathbb{P}} \theta, \quad (4.6)$$

pode ser considerado assintoticamente normal, ou seja,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathbb{D}} \mathcal{N}(0, \bar{J}_\Psi^{-1}(\theta)), \quad (4.7)$$

em que

$$\bar{J}_\Psi(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ S_\Psi^\top(\theta) V_\Psi^{-1}(\theta) S_\Psi(\theta) \right\},$$

atua desempenhando o papel da matriz de informação de Godambe assintótica. No que se refere ao processo de convergência das equações (4.6) e (4.7), tem-se, respectivamente, um processo de convergência em probabilidade e convergência em distribuição.

Teorema 2: Admitindo a ocorrência das seguintes premissas:

1: $y_i, i=1, \dots, n$ são vetores aleatórios t_i -dimensionais independentes;

2: $\psi_i(\theta) = (\psi_{i1}, \dots, \psi_{ip})$, com $i=1, \dots, n$, são funções de estimação regulares;

3: $\Psi_n(\theta) = \sum_{i=1}^n \psi_i(\theta)$;

4: Para $\delta > 0$

$$E_{\theta} \left\{ \sup_{h: \|h\| \leq \delta} \left\| \frac{\partial}{\partial \theta^{\top}} \psi_i(\theta + h) - \frac{\partial}{\partial \theta^{\top}} \psi_i(\theta) \right\| \right\} \rightarrow \phi_{\delta},$$

a medida que $n \rightarrow \infty$, $\phi_{\delta} \rightarrow \infty$ quando $\delta \rightarrow \infty$ e $\phi_{\delta} \rightarrow 0$ quando $\delta \rightarrow 0$. Além disso, a convergência do processo é em probabilidade.

5: Quando $n \rightarrow \infty$ com convergência em probabilidade.

$$\frac{1}{n} \frac{\partial \Psi_n}{\partial \theta^{\top}}(\theta) \xrightarrow{\mathbb{P}} S_{\Psi}(\theta).$$

6: É possível demonstrar que com convergência em distribuição

$$\frac{1}{n} \sum_{i=1}^n \text{Cov}(\psi_i) \xrightarrow{\mathbb{D}} V(\theta)$$

é positiva definida.

7: Quando $n \rightarrow \infty$ com convergência em distribuição.

$$\frac{\Psi_n(\theta)}{\sqrt{n}} \xrightarrow{\mathbb{D}} \mathcal{N}_p(0, V_{\Psi}(\theta)).$$

8: Dado que $\hat{\theta}_n$ é a solução para $\Psi_n(z) = 0, z \in \Theta$ e sob condições que permitam a existência de uma sequencia de raizes de $\Psi_n(z)$, sendo limitada em probabilidade, ou restrita a um conjunto compacto quase certamente quando $n \rightarrow \infty$, pode-se mostrar que:

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta.$$

e

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathbb{D}} \mathcal{N}(0, \bar{J}_{\Psi}^{-1}(\theta)).$$

O Teorema 2 está provado em Jørgensen e Labouriau (1994, p. 144), por exemplo. Além disso, ressalta-se que tais premissas aqui apresentadas são generalizações das condições de regularidade de Frechet-Cramer-Rao (Sen et al., 2009). Sugere-se pra o leitor que busca um maior aprofundamento examinar o caso uniparamétrico em Godambe (1960).

4.2 Modelagem da média

Considerando a coleta de uma amostra aleatória de tamanho n e que a i -ésima unidade experimental, $i=1, \dots, n$, foi observada t_i vezes sob uma condição avaliativa, a exemplo o tempo, de modo que cada observação está relacionada a um valor da variável resposta y_{ij} . Além do mais, admite-se que cada observação sofre a influência de um conjunto de p covariáveis, revelando, assim, um controle ou substituição do controle local.

A partir do exposto, compreende-se que $x_{ij} = (x_{ij1}, \dots, x_{ijp})^\top$ seja o vetor que possui os valores das p covariáveis para o sujeito i no j -ésimo instante, $y_i = (y_{i1}, \dots, y_{it_i})^\top$ compreendido como um vetor ($t_i \times 1$) com as respostas observadas para o i -ésimo indivíduo e associado a ele $X_i = (x_{i1}, \dots, x_{it_i})^\top$ que se apresenta como uma matriz de especificação ($t_i \times p$). Com isso, restringindo a distribuição marginal de y_{ij} para uma distribuição pertencente a família exponencial linear, sua densidade marginal pode ser definida como:

$$f(y_{ij}|\theta_{ij}, \phi) = \exp \{ \phi [y_{ij}\theta_{ij} - b(\theta_{ij})] + c(y_{ij}, \phi) \} \mathbb{1}_{\mathcal{X}}(y_{ij}),$$

com

$$E(y_{ij}) = \mu_{ij} = b'(\theta_{ij}) \text{ e } \text{Var}(y_{ij}) = \phi^{-1} b''(\theta_{ij}),$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas; e ϕ^{-1} é interpretado como um parâmetro de dispersão conhecido. O modelo estabelecido para a média é dado por:

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta}.$$

4.3 Equações de Estimação Generalizadas

Propostas por Liang e Zeger (1986), as Equações de Estimação generalizadas são aplicadas em casos onde as observações de uma mesma unidade observacional são dependentes. Nesse sentido, o modelo proposto pelos autores apresentam uma função capaz de incorporar informações sobre a estrutura de dependência dos dados. Com isso, os fatores da função de estimação ótima são dados por:

$$\begin{aligned} E_{\boldsymbol{\beta}} \left(\frac{\partial u_i}{\partial \boldsymbol{\beta}^\top} \right)^\top &= E_{\boldsymbol{\beta}} \left[\frac{(y_i - \boldsymbol{\mu}_i)}{\partial \boldsymbol{\beta}^\top} \right]^\top = \\ &= -E_{\boldsymbol{\beta}} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \right)^\top = \\ &= -\mathbf{X}_i^\top \mathbf{G}_i \mathbf{A}_i = -\mathbf{D}_i^\top. \end{aligned}$$

e

$$\begin{aligned}\mathbf{Cov}_\beta(\mathbf{u}_i) &= \mathbf{Cov}(\mathbf{y}_i) = \mathit{diag} \left\{ \mathbf{Var}(y_{ij})^{1/2} \right\} \mathbf{R}_i \mathit{diag} \left\{ \mathbf{Var}(y_{ij})^{1/2} \right\} = \\ &= \phi^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2} = \Sigma_i\end{aligned}$$

A partir disso, tem-se que \mathbf{R}_i é dita como a matriz verdadeira de correlação das componentes de \mathbf{y}_i , $i = 1, \dots, n$ com todas as matrizes definidas para as Equações de Estimação Independentes.

Com base nisso, considerando a matriz verdadeira de correlação a equação de estimação de β é dada por:

$$\Psi_n^{G*}(\hat{\beta}_G^*) = \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \hat{\mu}_i) = 0. \quad (4.8)$$

Liang e Zeger (1986), tomando por base as condições gerais de regularidade, expressas no Teorema 2, conseguiram provar que $\hat{\beta}_G^*$, raiz de (4.8), é um estimador consistente de β .

Tomando por base a matriz de transformação de Godambe é possível relacionar que equação de estimação generalizada de β é dada por:

$$\mathbf{J}_G^*(\beta_G) = \sum_{i=1}^n \mathbf{D}_i^\top \Sigma_i^{-1} \mathbf{D}_i.$$

Consequentemente, um estimador (estimador “naive” ou “model-based”) consistente para a matriz de variâncias e covariâncias de $\hat{\beta}_G$ é:

$$(\hat{\mathbf{J}}_G^*)^{-1} = \left\{ \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\Sigma}_i^{-1} \hat{\mathbf{D}}_i \right\}^{-1}, \quad (4.9)$$

com todas as qualidades avaliadas em $\hat{\beta}_G$.

Outro conceito apresentado por Liang e Zeger (1986) importante é matriz de correlação de trabalho ou matriz de trabalho ($\mathbf{R}_i(\alpha)$). Esse conceito busca contornar o fato de que na prática, comumente, a função $\Psi_n^{G*}(\hat{\beta}_G^*)$ é pouco utilizada, pois a matriz verdadeira de correlação \mathbf{R}_i é desconhecida. Diante disso, é definido que $\mathbf{R}_i(\alpha)$ é uma matriz simétrica ($t_i \times t_i$) satisfazendo as condições para ser uma matriz de correlação, onde α é um vetor ($s \times 1$) que define completamente $\mathbf{R}_i(\alpha)$. O ponto de destaque desse processo é o fato de que a matriz trabalho não precisa ser necessariamente a verdadeira matriz de correlação das componentes dos \mathbf{y}_i 's, com isso, dado que ela reappresenta a correlação entre

as observações de um mesmo grupo, os valores de $\mathbf{R}_i(\boldsymbol{\alpha})$ estão contidos no intervalo $[-1; 1]$. Assim, as Equações de Estimação Generalizadas (EEG's) de $\boldsymbol{\beta}$ são expressas por:

$$\boldsymbol{\Psi}_n^G(\hat{\boldsymbol{\beta}}_G) = \boldsymbol{\Psi}_n^G \left[\hat{\boldsymbol{\beta}}_G, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}_G, \hat{\phi}) \right] = \sum_{i=1}^n \hat{\mathbf{D}}_i^\top \hat{\boldsymbol{\Omega}}_i^{-1} (y_i - \mu_i) = 0, \quad (4.10)$$

em que

$$\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\alpha}, \phi) = \phi^{-1} \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}, \quad (4.11)$$

denotando $\hat{\boldsymbol{\alpha}}$ como um estimador consistente de $\boldsymbol{\alpha}$.

Nesse ponto, deve-se ressaltar que as equações de estimação em (4.10) deixam de ser ótimas, podendo tornarem-se viesadas. Com isso, é preciso garantir que $\boldsymbol{\alpha}$ seja definido de modo que $\hat{\boldsymbol{\beta}}_G$ permaneça sendo um estimador consistente de $\boldsymbol{\beta}$, além de ser assintoticamente normal. Ainda nesse cenário é presuposto que quase sempre ϕ é desconhecido, requerendo um estimador para tal parâmetro. Finalmente, para que o estimador de $\boldsymbol{\beta}$ conserve suas propriedades de estimação ótima, os estimadores de $\boldsymbol{\alpha}$ e ϕ devem satisfazer algumas condições descritas no teorema abaixo:

Teorema 3: Supondo $\hat{\boldsymbol{\beta}}_G$ a raiz de (4.10), de acordo com as condições gerais de regularidade e assumindo que

- 1: $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$ é um estimador \sqrt{n} -consistente de $\boldsymbol{\alpha}$;
- 2: $\hat{\phi}(\boldsymbol{\beta})$ é um estimador \sqrt{n} -consistente de ϕ ;
- 3: $\left| \frac{\partial \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)}{\partial \phi} \right| \leq \mathbf{H}(\mathbf{y}, \boldsymbol{\beta})$, sendo $\mathbf{H}(\mathbf{y}, \boldsymbol{\beta})$ uma função limitada em probabilidade;

A partir disso, $\hat{\boldsymbol{\beta}}_G$ é um estimador consistente de $\boldsymbol{\beta}$ e

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \rightarrow \mathcal{N}_p(0, (\bar{\mathbf{J}}_G)^{-1}),$$

de modo que:

$$\bar{\mathbf{J}}_G(\boldsymbol{\beta}_G) = \frac{\mathbf{J}_G}{n} = \lim_{n \rightarrow \infty} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\},$$

com

$$\mathbf{S}_i = E_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} (y_i - \mu_i) \right] = \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} E_{\boldsymbol{\beta}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} (y_i - \mu_i) \right] = -\mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i$$

e

$$\begin{aligned} \mathbf{V}_i^G &= E_{\boldsymbol{\beta}} \left[\mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} (y_i - \mu_i) (y_i - \mu_i)^\top \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right] = \\ &= \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i E_{\boldsymbol{\beta}} \left[(y_i - \mu_i) (y_i - \mu_i)^\top \right] \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i = \\ &= \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{Cov}(\mathbf{y}_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i. \end{aligned}$$

A demonstração do resultado anteriormente apresentado se deve a Liang e Zeger (1986).

Ressalta-se que os resultados do Teorema 3 são válidos mesmo quando $\mathbf{R}(\boldsymbol{\alpha})$ não corresponde a verdadeira matriz de correlações de .

Proposta pelos autores anteriormente citados, a estimação consistente da matriz de covariâncias é escrita como:

$$\hat{\mathbf{J}}_G^{-1} = \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\Omega}_i^{-1} \hat{u}_i \hat{u}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right\} \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1} \quad (4.12)$$

A estimativa da expressão acima é obtida através da substituição de $\boldsymbol{\alpha}$; $\boldsymbol{\beta}$ e $\boldsymbol{\phi}$ pelo seus respectivos estimadores consistentes. Além disso, tal estimador recebe o nome de estimador robusto, empírico ou sanduíche. Um caso particular ocorre quando $\mathbf{R}_i(\boldsymbol{\alpha})$ for a matriz verdadeira de correlação , em que $\boldsymbol{\Omega}_i = \mathbf{Cov}(\mathbf{y}_i)$ e o estimador robusto coincidirá com o estimador “ naive ” definido em (4.9).

4.4 Testes de hipóteses

Tomando como referência na construção de estatísticas de teste para testes de hipóteses, são apresentados o teste de Wald e o teste de escore. Basicamente, tais testes são fundamentados para modelos que se baseiam em verossimilhança. A seguir serão apresentadas as adaptações das estatísticas do tipo Wald (Wald, 1943) e escore de Rao (Rao, 1948) para equações de estimação generalizadas os mesmos são definidos para as Equações de Estimação Generalizadas, em linha com Hardin e Hilbe (2003) e Venezuela (2003). Inicialmente, é necessário admitir que $\boldsymbol{\beta}$ é o vetor (p 1) de coeficientes de regressão, podendo ser expresso como o vetor aumentado $(\boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top)^\top$, de modo que $\boldsymbol{\gamma}$ possui os parâmetros de interesse e $\boldsymbol{\delta}$ os demais componentes. Os testes de hipóteses podem ser expressos da seguinte maneira: $\mathcal{H}_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ versus $\mathcal{H}_a : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$.

4.4.1 Teste de Wald

Por meio da utilização do estimador robusto, a estatística de Wald para EEG's é definida como:

$$Q_w = (\hat{\boldsymbol{\gamma}}_G - \boldsymbol{\gamma}_0)^\top \hat{\mathbf{J}}_{\boldsymbol{\gamma}_G}^{-1} (\hat{\boldsymbol{\gamma}}_G - \boldsymbol{\gamma}_0),$$

de modo que $\hat{\gamma}_G$ é o vetor ($q \times 1$) dos q primeiros componentes de $\hat{\theta}_G$ e $\hat{\mathbf{J}}_{\gamma_G}^{-1}$ a matriz ($q \times q$) do estimados robusto $\hat{\mathbf{J}}_G^{-1}$ definido em (4.12). Considerando que :

$$\sqrt{n}(\hat{\beta}_G - \beta_G) \rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{J}_G^{-1})$$

quando $n \rightarrow \infty$, tem-se, sob hipótese nula, que Q_w possui distribuição assintótica χ_q^2 (Qui-quadrado com q graus de liberdade).

4.4.2 Teste Escore

Partindo para o Teste Escore, o mesmo é dado por:

$$Q_s = \Psi^G(\gamma_0)^\top \hat{\mathbf{V}}_{\gamma_0}^{-1} \Psi^G(\gamma_0),$$

de modo que $\Psi^G(\gamma_0)$ é o subvetor ($q \times 1$) de $\Psi^G e_{\gamma_0}^{-1}$ a submatriz ($q \times q$) da matriz de variabilidade apresentada em (4.12), avaliados no vetor γ_0 . Analogamente ao teste de Wald generalizada, sob a hipótese nula, Q_s possui distribuição assintótica χ_q^2 .

Contudo, ressalta-se que a utilização do estimador robusto, poderá ocasionar alguns casos problemas de singularidade devido ao número de unidades observacionais ser pequeno. Como alternativa a isso, Rotnitzkye Jewell (1990) apresentam uma alternativa ao teste de Wald e ao teste de escore generalizado denominada testes de trabalho, com isso, tal abordagem é baseada na utilização do estimador *model-based* apresentado em (4.9).

4.5 Técnicas de diagnóstico para EEG's

Tendo em vista a relevância do processo de diagnósticos para a escolha adequada de um modelo, nesta Seção serão apresentadas as métricas de diagnóstico para a modelagem de EEG's apresentadas e desenvolvidas em Venezuela (2003), Venezuela et al.(2007) e Venezuela et al. (2011). Ademais, no que tange o tratamento de técnicas de diagnóstico para EEG's, reforça-se as contribuições dos trabalhos de Preisser e Qaqish (1996) - responsáveis pela apresentação de formas de detectar observações e/ou unidades experimentais influentes; Chang (2000) - autor do teste não-paramétrico para a avaliação da aleatoriedade dos resíduos e Pan (2001) - responsável pelo desenvolvimento de medidas capazes de escolherem a melhor matriz de correlação de trabalho, além selecionarem covariáveis baseados no AIC (critério de informação de Akaike).

4.5.1 Alavancagem

Hoaglin e Welsch (1978), Cook e Weisberg (1982), Emerson et al. (1984), St. Laurent e Cook (1992) e Wei et al. (1998) busca, como objetivo principal da aplicação do conceito de pontos de alavanca em modelos lineares normais, analisar a influência da i -ésima observação y_i sobre o próprio valor ajustado \hat{y}_i através de h_{ii} , os elementos da diagonal principal de \mathbf{H} , posto que h_{ii} corresponde à variação de \hat{y}_i quando y_i é acrescido de um infinitésimo. Partindo para a definição de alavancagem para MLG's, Pregibon (1981) propôs uma medida formulada capaz de fazer analogia entre a solução para β num MLG utilizando o método da máxima verossimilhança e a solução de mínimos quadrados de uma regressão linear ponderada. Dentro desse cenário, Venezuela et al. (2007) estenderam para EEG o modelo inicialmente apresentado por Pregibon (1981).

No que se refere à expansão do modelo, apresentado por Venezuela et al. (2007), é sugerido a utilização dos elementos da diagonal principal da matriz \mathbf{H} para a identificação de possíveis pontos de alavanca, assim como proposto por Hoaglin e Welsch (1978) para modelos normais lineares. Com isso, a ocorrência de um ponto de alavanca será fundamentada quando este for ponto remoto no subespaço gerado pelas colunas da matriz de especificação, ou seja, um valor alto de h_{ij} evidencia a influência do valor observado no correspondente valor ajustado.

Para a i -ésima unidade observacional pode ser caracterizada como unidade

observacional alavancada quando

$$\mathbf{h}_i = \frac{1}{t_i} \sum_{j=i}^t h_{ij} = \frac{\text{tr}(\mathbf{H}_i)}{t_i} \geq \frac{2p}{N}.$$

É importante frisar que os valores de h_{ij} dependem da matriz de pesos $\hat{\mathbf{W}}_i$. A identificação de detectar os possíveis pontos de alavanca através se dá por meio de um gráfico com os valores da diagonal principal da matriz \mathbf{H} , h_{ij} , com $i=1, \dots, n$ e $j = 1, \dots, t_i$, versus i indicando a ordem em que cada unidade observacional aparece no conjunto de dados. Ademais, caso o interesse de estudo seja verificar se a unidade observacional é um ponto de alavanca, então deverá ser feito um gráfico de h_{ij} versus os índice das unidades amostrais.

4.5.2 Análise de resíduos

Semelhante à matriz \mathbf{H} , é possível considerar o vetor de resíduos ordinários da solução de mínimos quadrados da regressão linear de z versus \mathbf{X} e os pesos de \mathbf{W} , com isso:

$$\mathbf{r}^* = \hat{\mathbf{W}}^{1/2}(z - \hat{\eta}) = \hat{\mathbf{W}}^{1/2}(\hat{\mathbf{A}}\hat{\mathbf{G}})^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}),$$

em que $\hat{\mathbf{A}} = \text{diag}(\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_n)$ e $\hat{\mathbf{G}} = \text{diag}(\hat{\mathbf{G}}_1, \dots, \hat{\mathbf{G}}_n)$, ambas com dimensão $(N \times N)$. Além disso, $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ e $\hat{\boldsymbol{\mu}} = (\hat{\boldsymbol{\mu}}_1^\top, \dots, \hat{\boldsymbol{\mu}}_n^\top)^\top$, ambas com dimensão $(N \times 1)$. A partir do exposto, assumindo que $\text{Cov}(z) = (\hat{\mathbf{G}}\hat{\mathbf{A}})^{-1} \text{Cov}(\mathbf{y})(\hat{\mathbf{A}}\hat{\mathbf{G}})^{-1} \cong \mathbf{W}^{-1}$, é possível mostrar que:

$$\begin{aligned} \text{Cov}(\mathbf{r}^*) &= \text{Cov}(\hat{\mathbf{W}}^{1/2}z - \hat{\mathbf{W}}^{1/2}\mathbf{X}\hat{\boldsymbol{\beta}}) = \\ &= \text{Cov}(\hat{\mathbf{W}}^{1/2}z - \hat{\mathbf{W}}^{1/2}\mathbf{X}(\mathbf{X}^\top \hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top \hat{\mathbf{W}}z) = \\ &= (\mathbf{I}_N - \mathbf{H})\hat{\mathbf{W}}^{1/2} \text{Cov}(z)\hat{\mathbf{W}}^{1/2}(\mathbf{I}_N - \mathbf{H}) \cong (\mathbf{I}_N - \mathbf{H}), \end{aligned}$$

de modo que \mathbf{I}_N representa a matriz identidade de dimensão $(N \times N)$ e \mathbf{H} a matriz de projeção $\mathbf{H}_i = \hat{\mathbf{W}}_i^{1/2}\mathbf{X}_i(\mathbf{X}_i^\top \hat{\mathbf{W}}_i\mathbf{X}_i)^{-1}\mathbf{X}_i^\top \hat{\mathbf{W}}_i^{1/2}$. Daí, dado que os elementos de \mathbf{r}^* podem possuir variâncias diferentes, tal fato dificulta compará-los entre si. Dessa forma, utiliza-se o resíduo padronizado associado à observação y_{ij} , que é dado por

$$r_{ij} = \frac{\mathbf{m}_{ij}^\top \mathbf{r}_i^*}{\sqrt{1 - h_{ij}}}, \quad (4.13)$$

em que \mathbf{m}_{ij} representa um vetor de tamanho t_i com a posição referente à observação y_{ij} possuindo o valor 1 com as demais posições contendo o valor zero. Além disso, h_{ij} é o j -ésimo elemento da diagonal principal de \mathbf{H}_i , com $i=1, \dots, n$ e $j = 1, \dots, t_i$.

4.5.3 *Envelope simulado*

Desenvolvido por Atkinson (1985), o gráfico de probabilidade meio-normal com envelope simulado é uma outra técnica aplicada na avaliação do ajuste do modelo. Construído com base nos resíduos padronizados, o mesmo apresenta-se como uma ferramenta de diagnóstico bastante útil na avaliação do ajuste do modelo. Além disso, de acordo com Neter et. al (1996) tal gráfico pode utilizado mesmo quando os resíduos não apresentam uma distribuição normal. O modelo pode ser descrito como:

$$E(|Z_{(i)}|) \approx \Phi^{-1} \left(\frac{i + N - 1/8}{2n + 1/2} \right). \quad (4.14)$$

Abaixo, em linha com Freitas (2018), é apresentado o algoritmo para a construção do gráfico de probabilidade meio-normal com envelope simulado:

ETAPAS PARA A ELABORAÇÃO DO GRÁFICO DE PROBABILIDADE MEIO-NORMAL COM ENVELOPE SIMULADO

- 1:** Para cada unidade experimental i ; $i = 1, \dots, t_i$, deverá ser simulado um vetor de respostas de tamanho t_i , observando o vetor de médias e a matriz de covariâncias ajustados aos dados originais.
- 2:** O mesmo modelo, utilizado para o ajuste dos dados originais, deverá ser aplicado ao ajuste das respostas simuladas no passo 1.
- 3:** Realizar o calculo dos resíduos padronizados utilizando (4.13). Após isso, ordenar os seus valores absolutos.
- 4:** Repetir os três primeiros passos mais 24 vezes, número de réplicas sugerido por Tan, Qu e Kutner (1997).
- 5:** Determinar o mínimo, a mediana e o máximo dos menores valores absolutos dos resíduos padronizados de todas as simulações.
- 6:** Repetir o passo anterior para os segundos menores valores absolutos dos resíduos padronizados das simulações, e analogamente, até os maiores valores absolutos dos resíduos padronizados das simulações. Ao final deste passo serão obtidos 3 vetores contendo os mínimos, as medianas e os máximos dos valores absolutos dos resíduos padronizados.
- 7:** Organizar em um gráfico os valores mínimos, medianos e máximos obtidos nos passos supracitados versus os correspondentes valores esperados em (4.14). Após isso, associar os pontos referente aos valores mínimos, medianos e máximos.

A partir do processo descrito anteriormente, a ocorrência de elevados desvios dos pontos em relação a mediana dos valores simulados ou a ocorrência de pontos próximos ou fora dos limites da banda de simulação são indicativos de que o modelo pode não ser adequado. Ademais, caso existam, os pontos aberrantes devem aparecer no topo direito do gráfico separado dos demais pontos.

Mais detalhes sobre sua construção do gráfico de probabilidade meio-normal com envelope simulado são apresentados em Neter et. al (1996); Tan, Qu e Kutner (1997) e em Venezuela (2003) poderão ser vistos os algoritmos para gerar variáveis aleatórias correlacionadas das distribuições Binomial, Poisson e Gama.

5 MODELOS DE REGRESSÃO BETA COM MEDIDAS REPETIDAS

O presente capítulo apresenta uma estrutura conceitual diretamente relacionada com os assuntos abordados nos Capítulos 2 e 3 do estudo aqui apresentado. Vale ressaltar que aqui será apresentado, de acordo com a proposta de Venezuela (2008), uma extensão das Equações de Estimação Generalizadas para os modelos de regressão beta com medidas repetidas. Nesse caso, tais equações apresentam como elemento primordial possibilidade da modelagem da média admitindo que há a presença de homogeneidade na dispersão, além de possibilitar a modelagem conjunta da média e da dispersão.

Além do exposto, tal capítulo busca explicar algumas métricas de diagnósticos para modelos que envolvem dados com medidas repetidas, bem como evidenciar os modelos de influência local para equações de estimação generalizadas fundamentadas por Cook (1986) e Cadigan e Farrell (2002). Por fim, será apresentada no Capítulo 5 uma aplicação com o objetivo de apresentar uma utilização prática das equações de estimação abordadas no presente estudo e validá-las por meio das técnicas de diagnósticos.

5.1 EEG's para modelos de regressão beta com presença de medidas repetidas

De acordo com Venezuela (2008) as equações de estimação generalizadas propostas aqui apresentadas foram desenvolvidas para os casos em que observa-se que a variável explicada possui medição contínua no intervalo unitário $(0,1)$ e há mais de uma observação numa mesma unidade experimental. Um exemplo disso, seria a observação da proporção de petróleo convertida em gasolina após o processo de fracionamento de um determinado lote ao longo de seis meses. A utilização de tais equações é embasada na suposição de que a variável resposta possui uma distribuição marginal do tipo beta e fazendo o uso da parametrização atribuída pela média e pelo parâmetro de precisão e apresentada na Seção (3.2).

Diante do exposto, busca-se definir a modelagem com base na perspectiva de que o parâmetro de posição é homogêneo. No que tange a ocorrência de modelos com uma possível heterogeneidade da precisão, aborda-se a modelagem conjunta da média e da precisão via EEG's.

5.1.1 Modelagem do parâmetro de posição

Considera-se que $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{it_i})^\top$ o vetor de respostas da i -ésima unidade observacional, com $i=1, \dots, n$ e que a densidade marginal da variável resposta possua distribuição beta, $y_{ij} \sim \mathcal{B}(\mu_{ij}, \phi)$, dada por:

$$f(y, \mu\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}. \quad (5.1)$$

Inicialmente, supõe-se que o parâmetro de precisão ϕ é conhecido e igual para todas as observações. Sem perda de generalidade, visando simplificar a notação, admite-se que $t_i = t, i = 1, \dots, n$. Com isso, admite-se que as médias μ_{ij} s são modeladas por

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta}, \quad (5.2)$$

sendo $g(\cdot)$ uma estrutura monótona e duplamente diferenciável, denominada por função de ligação, η_{ij} o preditor linear, \mathbf{x}_{ij} um vetor de covariáveis referente à j -ésima observação da i -ésima unidade observacional, com $i = 1, \dots, n$ e $j = 1, \dots, n$, além de $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top, p < n$, dito um vetor de parâmetros desconhecidos a serem estimados.

Para a realizar o uso das equações de estimação pautadas nos modelos de regressão beta com medidas repetidas, deve-se utilizar os conceitos apresentados em (4.4) sobre a definição de estimação linear ótima. Em linha com a definição apresentada, para que ocorra a adequação às propriedades das funções de estimação regulares, faz-se necessário que hajam vetores $\mathbf{b}_i = \mathbf{b}_i(\mathbf{y}_i; \boldsymbol{\beta}), i = 1, \dots, n$, com média zero, mutuamente independentes.

No que tange os modelos de regressão beta com medidas repetidas, primeiramente, define-se $\mathbf{b}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$, com $\mathbf{y}_i = (y_{i1}, \dots, y_{it})^\top$ e $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{it})^\top, i = 1, \dots, n$, em que os mesmos satisfazem as propriedades das funções de estimação regulares. Contudo, na presença de independência entre as observações da mesma unidade experimental, os vetores \mathbf{b}_i s não formam um classe $\ell(b)$ (ver definição (4.1)) que contém a função de escore proposta por Ferrari e Cribari-Neto (2004).

Uma forma alternativa para tratar isso, permitindo a criação das equações estimação capazes de generalizar tanto o caso em que há correlação entre as observações de uma mesma unidade experimental como em casos mais simples, onde há independência entre elas, Venezuela (2008) propõe:

$$\mathbf{b}_i = \mathbf{y}_i^* - \boldsymbol{\mu}_i^*,$$

com $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{it}^*)^\top$ e $\boldsymbol{\mu}_i^* = (\mu_{i1}^*, \dots, \mu_{it}^*)^\top$, $i = 1, \dots, n$. Tais vetores produzem uma classe $\ell(\mathbf{b})$ que, em caso de independência entre todas as observações, possui a função escore apresentada por Ferrari e Cribari, além de serem vetores de média zero, mutuamente dependentes e com as características das funções de estimação.

Admitindo a estrutura da densidade marginal de y_{ij} (5.1), o componente sistemático (5.2) e definindo $\mathbf{b}_i = \mathbf{y}_i^* - \boldsymbol{\mu}_i^*$, $i = 1, \dots, n$, dessa forma as equações de estimação generalizadas para modelos de regressão beta com medidas repetidas podem ser realizadas. Ademais, assumindo a ocorrência de dependência entre as observações da mesma unidade experimental i , os termos da função estimadora definida em (4.4) são dados por:

$$E \left(\frac{\partial \mathbf{b}_i}{\partial \boldsymbol{\beta}^\top} \right)^\top = -\mathbf{X}_i^\top \boldsymbol{\Lambda}_i$$

e

$$\text{Cov}(\mathbf{b}_i) = \text{Var}(\mathbf{y}_i^*)^{1/2} \mathbf{R}(\mathbf{y}_i^*) \text{Var}(\mathbf{y}_i^*)^{1/2} = \mathbf{A}_i^{1/2} \mathbf{R}(\mathbf{y}_i^*) \mathbf{A}_i^{1/2},$$

com isso, tem-se que $\mathbf{X}_i = (x_{i1}, \dots, x_{it})^\top$, $\boldsymbol{\Lambda}_i = \boldsymbol{\phi} \mathbf{G}_i \mathbf{A}_i$, $\mathbf{A}_i = \text{diag}(a_{i1}, \dots, a_{it})$ e $\mathbf{R}(\mathbf{y}_i^*)$ a matriz de correlação verdadeira de \mathbf{y}_i^* , com $\mathbf{G}_i = \text{diag}(\partial g^{-1}(\eta_{it})/\partial \eta_{i1}, \dots, \partial g^{-1}(\eta_{it})/\eta_{it})$ e $a_{ij} = \boldsymbol{\psi}'(\mu_{ij}\boldsymbol{\phi}) + \boldsymbol{\psi}'((1 - \mu_{ij})\boldsymbol{\phi})$, $i = 1, \dots, n$ e $j = 1, \dots, t$.

A função de estimação linear ótima de $\boldsymbol{\beta}$, considerando a matriz verdadeira de correlação de \mathbf{y}_i^* e $\boldsymbol{\phi}$ conhecido, equivale-se a:

$$\boldsymbol{\Psi}_1^o(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{X}_i^\top \boldsymbol{\Lambda}_i \text{Cov}(\mathbf{b}_i)^{-1} (\mathbf{y}_i^* - \boldsymbol{\mu}_i^*).$$

Conforme apresentado no Teorema 1, tem-se que $\hat{\boldsymbol{\beta}}$ (solução para $\boldsymbol{\Psi}_1^o(\boldsymbol{\beta}) = 0$) é um estimador consistente de $\boldsymbol{\beta}$. Além disso:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\text{D}} \mathcal{N}_p \left(0, \lim_{n \rightarrow \infty} n \left\{ \sum_{i=1}^n \mathbf{X}_i^\top \boldsymbol{\Lambda}_i \text{Cov}(\mathbf{b}_i)^{-1} \boldsymbol{\Lambda}_i \mathbf{X}_i \right\}^{-1} \right)$$

Em termos práticos, verifica-se que a matriz verdadeira de correlação é comumente desconhecida. Com isso, considerando a proposta de Liang e Zeger (1986), introduz-se o conceito da matriz de correlação de trabalho $\mathbf{R}(\boldsymbol{\alpha})$, uma matriz simétrica ($t \times t$) que satisfaz as condições para ser uma matriz de correlação, em que $\boldsymbol{\alpha}$ é um vetor ($s \times 1$) que atua caracterizando totalmente $\mathbf{R}(\boldsymbol{\alpha})$.

Finalmente, a função de estimação generalizada de $\boldsymbol{\beta}$ é definida a seguir:

$$\boldsymbol{\Psi}_1(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{X}_i^\top \boldsymbol{\Lambda}_i \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i^* - \boldsymbol{\mu}_i^*) = \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i \boldsymbol{\Lambda}_i^{-1} \mathbf{b}_i, \quad (5.3)$$

com $\mathbf{\Omega}_i = \mathbf{A}_i^{1/2} \mathbf{R}(\alpha)$ e $\mathbf{W}_i = \Lambda_i \mathbf{\Omega}_i^{-1} \Lambda_i$.

Nesse ponto, deve-se ressaltar que as equações de estimação em (5.3) deixam de ser ótimas, podendo tornarem-se viesadas. Com isso, é preciso garantir que $\hat{\alpha}$ seja definido de modo que α permaneça sendo um estimador consistente de β , além de ser assintoticamente normal. Ainda nesse cenário é pressuposto que quase sempre ϕ é desconhecido, requerendo um estimador para tal parâmetro. Finalmente, para que o estimador de β conserve suas propriedades de estimação ótima, os estimadores de α e ϕ devem satisfazer algumas condições descritas no teorema a seguir:

Teorema 3: Supondo $\hat{\beta}_G$ a raiz de (5.3), de acordo com as condições gerais de regularidade e assumindo que

- 1: $\hat{\alpha}(\beta, \phi)$ é um estimador \sqrt{n} -consistente de α dados β e ϕ ;
- 2: $\hat{\phi}(\beta)$ é um estimador \sqrt{n} -consistente de ϕ dado β ;
- 3: $\left| \frac{\partial \hat{\alpha}(\beta, \phi)}{\partial \phi} \right| \leq \mathbf{H}(\mathbf{y}, \beta)$, sendo $\mathbf{H}(\mathbf{y}, \beta)$ uma função limitada em probabilidade;

A partir disso, $\hat{\beta}_G$ é um estimador consistente de β e

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathbb{D}} \mathcal{N}_p(0, \mathbf{J}^{-1}),$$

de modo que $\mathbf{J} = \lim_{n \rightarrow \infty} \mathbf{J}_n/n$, com \mathbf{J} sendo a matriz de informação de Godambe de β associada à Ψ_1 definida como:

$$\mathbf{J}_n = \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\},$$

em que

$$\sum_{i=1}^n \mathbf{S}_i = E \left[\frac{\partial}{\partial \beta^\top} \Psi_1(\beta) \right] = - \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i \mathbf{X}_i$$

e

$$\sum_{i=1}^n \mathbf{V}_i = E \left[\Psi_i(\beta) \Psi_1^\top(\beta) \right] = \sum_{i=1}^n \mathbf{X}_i^\top \Lambda_i \mathbf{\Omega}_i^{-1} \text{Cov}(\mathbf{b}_i) \mathbf{\Omega}_i^{-1} \Lambda_i \mathbf{X}_i \text{Cov}(\mathbf{b}_i) \mathbf{\Omega}_i^{-1}$$

Em Artes (1997) é apresentada a demonstração do teorema acima, bem como evidenciado que os resultados assintóticos desse teorema são válidos mesmo quando $\mathbf{R}(\alpha)$ não é correspondente com a matriz verdadeira de correlação. Assim, a matriz de covariâncias de β é estimada de maneira consistente de acordo com a expressão a seguir:

$$\hat{\mathbf{J}}_n^{-1} = \left\{ \sum_{i=1}^n \mathbf{S}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{X}_i^\top \hat{\Lambda}_i \hat{\mathbf{\Omega}}_i^{-1} \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^\top \hat{\mathbf{\Omega}}_i^{-1} \hat{\Lambda}_i \mathbf{X}_i \right\} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\}^{-1}$$

A substituição de β , α e ϕ por suas respectivas estimativas consistente, acarreta na estimativa da expressão acima. Tal estimativa é denominada pela literatura de estimador robusto, empírico ou sanduíche. O estimador robusto se reduz a

$$\hat{\mathbf{J}}_n^{-1} = - \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1},$$

quando $\mathbf{R}(\alpha)$ representar a matriz verdadeira de correlação. Nesse caso, o estimador acima é chamado de estimador “naive” ou “model-based”.

5.1.2 Estimação de β , ϕ e α

O procedimento para o cálculo de β , ϕ e α relaciona o método *scoring* de Fisher para realizar a estimação de β com o método dos momentos para estimar ϕ e α . Estendendo a equação de estimação generalizada apresentada em (5.3) e tomando por base um valor inicial $\hat{\beta}^{(0)}$, o processo iterativo para estimar β é apresentado a seguir:

$$\begin{aligned} \hat{\beta}^{(m+1)} &= \hat{\beta}^{(m)} - \left\{ E \left[\frac{\partial}{\partial \beta^\top} \Psi_1(\hat{\beta}^{(m)}) \right] \right\}^{-1} \Psi_1(\hat{\beta}^{(m)}) = \\ &= \hat{\beta}^{(m)} + \left\{ \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \hat{\Lambda}_i^{-1} \hat{\mathbf{b}}_i \right] \right\}^{(m)} \end{aligned}$$

em que $m=1, \dots, 2$, corresponde ao número de interações. Além disso, o índice m no lado direito das equações evidencia a atualização dos matrizes e dos vetores de acordo com as estimativas de β , ϕ e α da m -ésima iteração.

Reescrevendo a equação anteriormente apresentada, levando em consideração o processo de mínimos quadrados ponderados, em que tal processo determina uma matriz de pesos \mathbf{W}_i e uma variável dependente modificada \mathbf{z}_i , obtêm-se :

$$\hat{\beta}^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_i \mathbf{z}_i \right] \right\}^{(m)}, \quad (5.4)$$

de modo que \mathbf{z}_i é definido como $\mathbf{z}_i = \hat{\eta}_i + \hat{\Lambda}_i^{-1} \hat{\mathbf{b}}_i$. Tal processo iterativo (5.4) será utilizado, especialmente, para a obtenção das métricas de diagnóstico.

Fazendo o uso do método dos momentos e da $\text{Var}(y_{ij} = v(\mu_{ij})/1 + \phi)$, o estimador obtido pelo método dos momentos, de acordo com Venezuela (2008) é dado por: observa-se

que a estimativa de ϕ obtida no m -ésimo passo do processo iterativo é descrito como:

$$\hat{\phi}^{(m)} = \left(\left\{ \sum_{i=1}^n \sum_{i=1}^{\top} \left(\frac{y_{ij} - \hat{\mu}_{ij}^{(m)}}{\sqrt{\hat{\mu}_{ij}^{(m)}(1 - \hat{\mu}_{ij}^{(m)})}} \right)^2 / (nt - p) \right\}^{-1} - 1 \right) \quad (5.5)$$

O processo de estimação de β permite que a estrutura de correlação entre as observações da mesma unidade experimental seja especificada das seguintes formas: identidade, padrão uniforme, auto-regressiva de ordem 1 e não estruturada, entre outras.

Diante do exposto, acompanhando a proposta apresentada por Artes (1997), são apresentados os estimadores para algumas estruturas utilizadas para definir $\mathbf{R}(\alpha)$ (matriz que reflete as correlações entre y_{ij}^* e y_{il}^*):

1: Na matriz de correlação uniforme, assume-se que $\text{Corr}(b_{ij}, b_{il}) = \alpha, \forall j \neq l$ e $1 \leq j, l \leq t$. Com isso, a estimativa de α no passo m é:

$$\hat{\alpha} = \left\{ \frac{\left(\sum_{i=1}^n \sum_{j>l}^{\top} \hat{b}_{ij} \hat{b}_{il} \right) \frac{nt}{\frac{1}{2}nt(t-1)}}{\left(\sum_{i=1}^n \sum_{j=1}^{\top} \hat{b}_{ij}^2 \right) \frac{2}{(t-1)}} \right\}^{(m)} = \left\{ \frac{\left(\sum_{i=1}^n \sum_{j>l}^{\top} \hat{b}_{ij} \hat{b}_{il} \right) \frac{2}{(t-1)}}{\left(\sum_{i=1}^n \sum_{j=1}^{\top} \hat{b}_{ij}^2 \right) \frac{2}{(t-1)}} \right\}^{(m)}.$$

2: Na matriz de correlação autoregressiva de primeira ordem, AR-1, especifica-se que $\text{Corr}(b_{ij}, b_{il}) = \alpha^{|j-l|}, \forall j \neq l$ e $1 \leq j, l \leq t$. Dessa forma, um estimador simples para α é:

$$\hat{\alpha} = \left\{ \frac{\left(\sum_{i=1}^n \sum_{j=1}^{t-1} \hat{b}_{ij} \hat{b}_{i,j+1} \right)}{\left(\sum_{i=1}^n \sum_{j=1}^{t-1} \hat{b}_{ij}^2 \sum_{i=1}^n \sum_{j=2}^t \hat{b}_{ij}^2 \right)^{1/2}} \right\}^{(m)}.$$

Um ponto de deficiência desse estimador se refere ao fato de que o mesmo utiliza apenas a informação sobre a dependência de observações consecutivas de b_{ij} .

3: Já, quando a matriz de correlação é do tipo não estruturada, $\alpha = (\alpha_{12}, \alpha_{13}, \dots, \alpha_{t-1,t})^{\top}$, apresenta-se como um vetor com $t(t-1)/2$ componentes, em que α_{jl} corresponde à correlação entre b_{ij} e b_{il} , com $i = 1, \dots, n$ e $j, l = 1, \dots, t$, com $j < l$. Assim, um estimador para α_{jl} é:

$$\hat{\alpha}_{jl}^{(m)} = \left\{ \frac{\left(\sum_{i=1}^n \hat{b}_{ij} \hat{b}_{il} \right)}{\left(\sum_{i=1}^n \hat{b}_{ij}^2 \right)^{1/2} \left(\sum_{i=1}^n \hat{b}_{il}^2 \right)^{1/2}} \right\}^{(m)}.$$

No que se refere à seleção da estrutura da matriz de correlação de trabalho, verifica-se que Liang e Zeger (1986) citam que para qualquer $\mathbf{R}(\alpha)$ utilizada há a garantida

de que $\hat{\beta}$ e $\hat{\mathbf{J}}_n^{-1}$ serão consistentes. Além disso, evidencia-se que a eficiência cresce a medida que a matriz de correlação de trabalho selecionada estiver próxima da matriz verdadeira de correlação. Em Wang e Carey (2003) é apresentado o estudo da eficiência relativa assintótica de $\hat{\beta}$ de acordo com a especificação da matriz de correlação de trabalho, do estimador de α e da matriz de planejamento. Pesquisas analíticas e numéricas desenvolvidas por tais autores evidenciam que o principal impacto gerado no processo de eficiência dos estimadores é oriundo da especificação da estrutura imposta à matriz de correlação de trabalho.

Como uma forma de sintetizar os processos acima, a seguir estão contidas as etapas que visam demonstrar como, na prática, os parâmetros β , ϕ e α devem ser estimados.

ETAPAS PARA A ESTIMAÇÃO DE β , ϕ E α

Etapa 1: Admitindo independência entre as observações da mesma unidade experimental, deve ser ajustado um modelo de regressão linear de $g(y)$ sobre \mathbf{X} via método dos mínimos quadrados ordinários, em que $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_n^\top)^\top$ e $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$. Vale ressaltar que ao assumir uma estrutura de independência, os parâmetros ϕ e α são eliminados do processo de estimação de β .

Etapa 2: Permanecendo com a suposição de independência entre as observações da mesma unidade, ou seja, admitindo que $\mathbf{R}(\alpha)$ seja igual a uma matriz identidade, sugere-se considerar as estimativas dos parâmetros de regressão da etapa anterior como os valores iniciais, $\hat{\beta}^{(0)}$, para a estimação de β e ϕ por meio dos processos descritos em (5.4) e (5.5), respectivamente. Tais estimativas são calculadas a cada passo do processo iterativo até que se atinja a convergência de β . Nessa etapa, são obtidas as estimativas de β e ϕ admitindo um modelo de regressão beta com independência entre as observações da mesma unidade experimental.

Etapa 3: Nesse passo é definida a estrutura de correlação de trabalho $\mathbf{R}(\alpha)$ para ser utilizada na modelagem dos dados. Caso a estrutura seja independente as estimativas de β e ϕ serão os valores obtidos na convergência do processo iterativo da Etapa 2. Caso contrário, considerando $\hat{\beta}^{(0)}$ e $\hat{\phi}^{(0)}$ as estimativas de β e ϕ definidos na etapa anterior, calcula-se os parâmetros de correlação ($\hat{\alpha}^{(0)}$). Dando continuidade, no passo seguinte do processo iterativo é calculado os parâmetros de regressão ($\hat{\beta}^{(m)}$), fazendo

uso de (5.4), o parâmetros de precisão ($\hat{\phi}^{(m)}$) e os parâmetros de correlação ($\hat{\alpha}^{(m)}$). Finalmente, o processo é repetido até que se chague à convergência de β .

5.1.3 Modelagem conjunta dos parâmetros de posição e dispersão

Tomando por base que $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{it_i})^\top$ seja o vetor ($t_i \times 1$) de respostas associado à i -ésima unidade experimental, com $i=1, \dots, n$. A partir disso, admitindo que a densidade marginal de y_{ij} possua distribuição beta, $y_{ij} \sim \mathcal{B}(\mu_{ij}, \phi_{ij})$, dada por:

$$f(y_{ij}; \mu_{ij}, \phi_{ij}) = \frac{\Gamma(\phi_{ij})}{\Gamma(\mu_{ij})\Gamma(\phi_{ij}-\mu_{ij})} y_{ij}^{\mu_{ij}-1} (1-y_{ij})^{\phi_{ij}-\mu_{ij}-1}, \quad (5.6)$$

Nesse ponto, evidencia-se a incorporação da heterogeneidade do parâmetro de precisão, em que cada unidade experimental y_{ij} possui sua própria precisão. Assim, ϕ_{ij} , $i=1, \dots, n$ e $j=1, \dots, t_i$. Tomando por base a estrutura do componente sistemático responsável por modelar a média, μ_{ij} , apresentada em (5.2) e que o componente sistemático que modela o parâmetro de precisão, ϕ_{ij} , como função de covariáveis \mathbf{q}_{ij} (que podem ser um subconjunto de \mathbf{x}_{ij}), seja dado por:

$$f(\phi_{ij}) = \delta_{ij} = \mathbf{q}_{ij}^\top \boldsymbol{\gamma}, \quad (5.7)$$

em que $f(\cdot)$ uma função monótona e duplamente diferenciável cuja inversa deve ser positiva, δ_{ij} o preditor linear e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$, $q < n$, um vetor de parâmetros desconhecidos a serem estimados, com $i=1, \dots, n$ e $j=1, \dots, t_i$.

Sem perda de generalidade, buscando facilitar a notação, tem-se que $t_i = t$, $i=1, \dots, n$. A estruturação das equações de estimação que modelam conjuntamente os parâmetros de posição e precisão, no que se refere aos modelos de regressão beta com medidas repetidas, leva em consideração a densidade marginal de y_{ij} apresentada em (5.6), os componentes sistemáticos (5.7) e (5.2) responsáveis pela modelagem dos parâmetros de posição e de precisão, respectivamente. Além disso, também é possível a modelagem dos vetores $\mathbf{b}_i = \mathbf{y}_i^* - \boldsymbol{\mu}_i^*$, $i= 1, \dots, n$.

A função de estimação ótima de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \text{ e } \boldsymbol{\gamma}^\top)^\top$, assumindo dependência entre as observações da mesma unidade experimental i , é equivalente a

$$\Psi_2^o(\boldsymbol{\theta}) = \sum_{i=1}^n \begin{pmatrix} \mathbf{X}_i^\top \mathbf{G}_i \boldsymbol{\Phi}_i \mathbf{A}_i \\ \mathbf{Q}_i^\top \mathbf{F}_i \mathbf{C}_i \end{pmatrix} \left(\mathbf{A}_i^1 / 2 \mathbf{R} (y_i^*) \mathbf{A}_i^1 / 2 \right)^{-1} (y_i^* - \boldsymbol{\mu}_i^*),$$

em que $\mathbf{X}_i = (x_{i1}, \dots, x_{it})^\top$, $\Phi_i = \text{diag}(\phi_{i1}, \dots, \phi_{it})$, $\mathbf{A}_i = \text{diag}(a_{i1}, \dots, a_{it})$, $\mathbf{Q}_i = (\mathbf{q}_{i1}, \dots, \mathbf{q}_{it})^\top$, $\mathbf{G}_i = \text{diag}(\partial g^{-1}(\eta_{i1})/\partial \eta_{i1}, \dots, \partial g^{-1}(\eta_{it})/\partial \eta_{it})$, $\mathbf{F}_i = \text{diag}(\partial f^{-1}(\delta_{i1})/\partial \delta_{i1}, \dots, \partial f^{-1}(\delta_{it})/\partial \delta_{it})$ e $\mathbf{C}_i = \text{diag}(c_{i1}, \dots, c_{it})$.

Além do mais, $\mathbf{R}(\mathbf{y}_i^*)$ a matriz verdadeira de correlação em função \mathbf{y}_i^* , em que $a_{ij} = \psi'(\mu_{ij}\phi_{ij}) + \psi((1-\mu_{ij})\phi_{ij})$ e $c_{ij} = \mu_{ij}\psi'(\mu_{ij}\phi_{ij}) - (1-\mu_{ij})\psi'((1-\mu_{ij})\phi_{ij})$, com $i=1, \dots, n$ e $j=1, \dots, t$. A demonstração aprofundada das expressões pode ser vista em Venezuela (2008). Em linha com o Teorema 1, é observado que $\hat{\theta}$, solução de $\Psi_2^o(\hat{\theta}) = 0$, é um estimador de θ . Ademais, com convergência em distribuição

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N_{p+q} \left(0, \lim_{n \rightarrow \infty} \left\{ \sum_{i=1}^n \begin{pmatrix} \mathbf{X}_i^\top \mathbf{G}_i \Phi_i \mathbf{A}_i \\ \mathbf{Q}_i^\top \mathbf{F}_i \mathbf{C}_i \end{pmatrix} \text{Cov}(\mathbf{b}_i)^{-1} \begin{pmatrix} \mathbf{X}_i^\top \mathbf{G}_i \Phi_i \mathbf{A}_i \\ \mathbf{Q}_i^\top \mathbf{F}_i \mathbf{C}_i \end{pmatrix}^\top \right\}^{-1} \right)$$

Realizando o mesmo procedimento da Seção 4.2.1, a matriz verdadeira de correlação de trabalho será substituída por uma matriz de correlação de trabalho, $\mathbf{R}(\alpha)$. Dessa forma, a função de estimação generalizada de θ é definida por:

$$\begin{aligned} \Psi_2(\theta) &= \begin{pmatrix} \Psi_2(\beta) \\ \Psi_2(\gamma) \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \mathbf{X}_i^\top \mathbf{G}_i \Phi_i \mathbf{A}_i \\ \mathbf{Q}_i^\top \mathbf{F}_i \mathbf{C}_i \end{pmatrix} \Omega_i^{-1} (y_i^* - \mu_i^*) = \\ &= \sum_{i=1}^n \mathbf{M}_i^\top \Lambda_i \Omega_i^{-1} \mathbf{b}_i = \sum_{i=1}^n \mathbf{M}_i \mathbf{W}_i \Lambda_i^{-\top} \mathbf{b}_i. \end{aligned} \quad (5.8)$$

De acordo com a expressão acima, $\Lambda_i^{-\top}$ é a inversa de Λ_i^\top , $\mathbf{M}_i = \begin{pmatrix} \mathbf{X}_i & 0 \\ 0 & \mathbf{Q}_i \end{pmatrix}$, $\Lambda_i = \begin{pmatrix} \mathbf{G}_i \Phi_i \mathbf{A}_i \\ \mathbf{F}_i \mathbf{C}_i \end{pmatrix}$, $\Omega_i = \mathbf{A}_i^{1/2} \mathbf{R}(\alpha) \mathbf{A}_i^{1/2}$ e $\mathbf{W}_i = \Lambda_i \Omega_i^{-1} \Lambda_i^\top$. Dado que Λ_i^\top é uma matriz de dimensão $t \times 2t$ com posto t , dessa forma $\Lambda_i^\top \Lambda_i^{-\top} = \mathbf{I}_t$, em que \mathbf{I}_t é uma matriz identidade de tamanho t (Graybill, 1976, Teorema 1.5.12). Vale ressaltar no presente caso apresentado, Λ_i não é definida como uma matriz assimétrica.

Tendo em vista que a Equação (5.8) deixa de ser ótima, o teorema a seguir apresenta as condições que o estimador de α possua para que satisfaça o estimador de θ seja consistente e assintoticamente normal.

Teorema 4: Admitindo que $\hat{\theta}$ seja a raiz de (5.8), de acordo com as condições gerais de regularidade, com $\hat{\alpha}\theta$ sendo um estimador de \sqrt{n} -consistente de α dado θ . Diante disso, $\hat{\theta}$ é um estimador consistente de θ e

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N_{p+q}(0, \mathbf{J}^{-1}),$$

de modo que $\mathbf{J} = \lim_{n \rightarrow \infty} \mathbf{J}_n/n$, com \mathbf{J} sendo a matriz de informação de Godambe de β associada à Ψ_2 definida como:

$$\mathbf{J}_n = \left\{ \sum_{i=1}^n \mathbf{S}_i \right\} \left\{ \sum_{i=1}^n \mathbf{V}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{S}_i \right\},$$

em que

$$\sum_{i=1}^n \mathbf{S}_i = E \left[\frac{\partial}{\partial \beta^\top} \Psi_2(\theta) \right] = - \sum_{i=1}^n \mathbf{M}_i^\top \mathbf{W}_i \mathbf{M}_i$$

e

$$\sum_{i=1}^n \mathbf{V}_i = E \left[\Psi_2(\theta) \Psi_2^\top(\theta) \right] = \sum_{i=1}^n \mathbf{M}_i^\top \Lambda_i \Omega_i^{-1} \text{Cov}(\mathbf{b}_i) \Omega_i^{-1} \Lambda_i \mathbf{M}_i \text{Cov}(\mathbf{b}_i) \Omega_i^{-1}$$

De maneira análoga a prova desse teorema pode ser encontrada em Artes (1997, Teorema 7, p.67). Ainda tomando por base a equação (5.8), para o ajuste conjunto dos parâmetros de posição e precisão, o estimador robusto para a matriz de covariâncias de $\hat{\theta}$ é:

$$\hat{\mathbf{J}}_n^{-1} = \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1} \left\{ \sum_{i=1}^n \mathbf{M}_i^\top \hat{\Lambda}_i \hat{\Omega}_i^{-1} \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^\top \hat{\Omega}_i^{-1} \hat{\Lambda}_i \mathbf{M}_i \right\} \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1}.$$

O estimador da expressão acima pode ser obtido por meio da substituição de θ e α por suas perspectivas estimativas consistentes. Além disso, o estimador “naive”, obtido quando $\mathbf{R}(\alpha)$ apresentar a matriz verdadeira de correlação dos y_i^* 's, é dado por:

$$\hat{\mathbf{J}}_n^{-1} = - \left\{ \sum_{i=1}^n \hat{\mathbf{S}}_i \right\}^{-1}.$$

5.1.4 Estimação de β , γ e α

O processo de iteração para o cálculo de $\hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top$ e $\hat{\alpha}$ une o método *scoring* de Fisher para a estimação dos parâmetros de regressão com o método dos momentos para a estimação dos parâmetros de correlação. Por meio da expansão da EEG apresentada em (5.9) quando situada em torno de um valor inicial, o processo de iteração para a estimação de θ segue a seguir.

$$\begin{aligned} \hat{\theta}^{(m+1)} &= \hat{\theta}^{(m)} - \left\{ E \left[\frac{\partial}{\partial \theta^\top} \Psi_2(\hat{\theta}^m) \right] \right\}^{-1} \Psi_2(\hat{\theta}^m) = \\ &= \hat{\theta}^{(m)} + \left\{ \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \mathbf{M}_i \right]^{-\top} \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \hat{\Lambda}_i^{-\top} \hat{\mathbf{b}}_i \right] \right\}^{(m)}, \end{aligned} \quad (5.9)$$

em que $m = 0, 1, 2, \dots$ o número de iterações. Reescrevendo a expressão acima na forma de um processo iterativo de método dos mínimos quadrados ponderados, tem-se que:

$$\hat{\theta}^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \mathbf{M}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \mathbf{z}_i \right] \right\}^{(m)}, \quad (5.10)$$

com $\mathbf{z}_i = \hat{\mathbf{v}} + \hat{\Lambda}_i^\top \hat{\mathbf{b}}_i$, em que $\mathbf{v}_i = (\boldsymbol{\eta}_i^\top, \boldsymbol{\delta}_i^\top)^\top$. De maneira prática, os passos para a estimação dos parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e $\boldsymbol{\alpha}$ são descritos a seguir.

ETAPAS PARA A ESTIMAÇÃO DE $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ E $\boldsymbol{\alpha}$

Etapa 1: Havendo a suposição de independência entre as observações da mesma unidade experimental, deve-se ajustar um modelo de regressão linear de $g(y)$ sobre \mathbf{X} e um de $f(\hat{\phi})$ sobre \mathbf{Q} , ambos por meio do método dos mínimos quadrados ordinários, em que $\boldsymbol{\phi} = (\phi_1^\top, \dots, \phi_n^\top)^\top$ e $\mathbf{Q} = (\mathbf{Q}_1^\top, \dots, \mathbf{Q}_n^\top)^\top$, com $\phi_i = (\phi_{i1}, \dots, \phi_{it})^\top$ e $i = 1, \dots, n$. Ademais, para a correção do parâmetro de precisão, o modelo sugerido por Ferrari e Cribari-Neto (2004) expresso por:

$$\hat{\phi}_{ij} = \frac{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}{\hat{\sigma}_{ij}^2},$$

de modo que,

$$\hat{\mu}_{ij} = g^{-1}(x_{ij}^\top \hat{\boldsymbol{\beta}}_{MQO}) \text{ e } \check{\sigma}_{ij}^2 = \check{\mathbf{e}}^\top \check{\mathbf{e}} [\check{\mathbf{G}}_i^{-2} j / (n - p)].$$

Com isso, $\hat{\boldsymbol{\beta}}_{MQO}$ é a estimativa gerada pelo ajuste do valor médio por meio do método dos mínimos quadrados ordinários, $\hat{\mathbf{e}} = g(y) - \mathbf{X} \hat{\boldsymbol{\beta}}_{MQO}$ o vetor referente aos resíduos ordinários de uma regressão linear com a variável resposta transformada e, por fim, $\hat{\mathbf{G}}_{ij}$ é o j -ésimo elemento da diagonal de $\hat{\mathbf{G}}_i$ quando avaliado em $\hat{\mu}_{ij}$.

Etapa 2: A estrutura da matriz de correlação de trabalho $\mathbf{R}(\boldsymbol{\alpha})$ deverá ser definida para que a mesma possa ser utilizada na modelagem dos dados. Diante disso, leva-se em conta as estimativas dos parâmetros de regressão da etapa anterior como medidas iniciais ($\boldsymbol{\beta}^{(0)}$ e $\boldsymbol{\gamma}^{(0)}$). Em seguida, obtêm-se a estimativa inicial de $\boldsymbol{\alpha}$ tomando por base os componentes do vetor \mathbf{b}_i . Em tal etapa, a estimação de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ é dada em processos iterativos separados. Tais processos são descritos a seguir.

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_{\beta i} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \hat{\mathbf{W}}_{\beta i} \mathbf{z}_{\beta i} \right] \right\}^{(m)},$$

com $\hat{\mathbf{W}}_{\beta_i} = \hat{\mathbf{G}}_i \hat{\mathbf{\Phi}}_i \hat{\mathbf{A}}_i \hat{\mathbf{\Omega}}_i^{-1} \hat{\mathbf{A}}_i \hat{\mathbf{\Phi}}_i \hat{\mathbf{G}}_i$ e $\mathbf{z}_{\beta_i} = \hat{\eta}_i + (\hat{\mathbf{A}}_i \hat{\mathbf{\Phi}}_i \hat{\mathbf{G}}_i)^{-1} \hat{\mathbf{b}}_i$.

$$\hat{\gamma}^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{Q}_i^\top \hat{\mathbf{W}}_{\gamma_i} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{Q}_i^\top \hat{\mathbf{W}}_{\gamma_i} \mathbf{z}_{\gamma_i} \right] \right\}^{(m)},$$

de modo que $\hat{\mathbf{W}}_{\gamma_i} = \hat{\mathbf{F}}_i \hat{\mathbf{C}}_i \hat{\mathbf{\Omega}}_i^{-1} \hat{\mathbf{C}}_i \hat{\mathbf{F}}_i$ e $\mathbf{z}_{\gamma_i} = \hat{\delta}_i + (\hat{\mathbf{C}}_i \hat{\mathbf{F}}_i)^{-1} \hat{\mathbf{b}}_i$. Vale ressaltar que os processos acima descritos são derivados da separação das equações de estimação de β e de γ , definidos em (5.8). Além do mais, as estimativas de β , γ e α devem ser calculadas a cada passo m até que se obtenha a convergência.

Etapa 3: Tomando as estimativas de β , γ e α , definidas na etapa anterior, como valores iniciais do processo iterativo definido em (5.9), deve-se repetir cada passo m do processo iterativo até a convergência de β e γ .

5.2 Técnicas de diagnósticos para dados com medidas repetidas

Tendo em vista a importância da etapa de análise de diagnóstico na avaliação da qualidade do ajuste de um modelo de regressão - dado que tal passo permite a identificação de possíveis distorções sobre as suposições realizadas para o modelo em questão, permitindo a identificação de pontos extremos que atuam interferindo de maneira desproporcional no resultado do ajuste dos dados - serão apresentadas algumas métricas de diagnóstico tomando por base as contribuições de Paula (2013) e Venezuela et al. (2007) de modo que, em linha com Venezuela (2008), serão apresentadas generalizações para os casos em que há abordagem de quaisquer equações de estimação apresentadas tanto para a modelagem do parâmetro de dispersão como para a modelagem conjunta dos parâmetros de posição e dispersão.

Reforça-se que as abordagens apresentadas suportam a presença de dados com medidas repetidas. Com base nisso, destacam-se, como modelos abordados no estudo, os elementos da diagonal da matriz de projeção, a distância de Cook e os resíduos do modelo ajustado. Finalmente, é necessário que para as métricas apresentadas a seguir tenham acurácia, a estimativa da matriz de correlação de trabalho - $\mathbf{R}(\hat{\alpha})$ - deverá ser próxima da verdadeira.

5.2.1 Pontos influentes, aberrantes e alavanca

Nesta parte do presente estudo serão apresentadas algumas métricas de diagnóstico construídas através de um processo de iterativo reponderado. Com isso, a estimação dos parâmetros de regressão θ poderá ser realizada tanto unicamente para a modelagem da média (quando $\theta = \beta$) quanto para a modelagem conjunta da média e da dispersão (quando $\theta = (\beta^\top, \gamma^\top)^\top$). De maneira análoga ao exposto acima, a interpretação dos parâmetros regressores (θ) está pautada em duas possibilidades, segundo Venezuela (2008):

Se o interesse for constituir medidas de diagnóstico que avaliem apenas o comportamento da modelagem da média, então os parâmetros de regressão θ são caracterizados pelos parâmetros de regressão β . Mas se o interesse for constituir medidas de diagnóstico que avaliem o comportamento conjunto dos ajustes da média e da dispersão dos dados, então os parâmetros de regressão são definidos por $\theta = (\beta^\top, \gamma^\top)^\top$.

A partir disso, as medidas de diagnóstico apresentadas possuem estruturas compreendidas como vetores ($N \times 1$) quando $\theta = \beta$ com $N = nt$, ou vetores ($2N \times 1$) na ocasião de $\theta = (\beta^\top, \gamma^\top)^\top$. De maneira genérica, o processo iterativo reponderado considerado para a estimação de θ é descrito a seguir:

$$\hat{\theta}^{(m+1)} = \left\{ \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \mathbf{M}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{M}_i^\top \hat{\mathbf{W}}_i \mathbf{z}_i \right] \right\}^{(m)}, \quad (5.11)$$

com $\mathbf{M}_i = \mathbf{X}_i$ quando $\theta = \beta$ ou $\mathbf{M}_i = \begin{pmatrix} \mathbf{X}_i & 0 \\ 0 & \mathbf{Q}_i \end{pmatrix}$ nos casos em que $\theta = (\beta^\top, \gamma^\top)^\top$, $\hat{\mathbf{W}}_i = \hat{\Lambda}_i \hat{\Omega}_i^{-1} \hat{\Lambda}_i$ e $\mathbf{z}_i = \hat{\mathbf{v}} + \hat{\Lambda}_i^\top \hat{\mathbf{b}}_i$, de modo que $\hat{\mathbf{v}} = \mathbf{M}_i \hat{\theta}$ com $i = 1, \dots, n$. No que se refere às dimensões de \mathbf{M}_i , \mathbf{W}_i e \mathbf{z}_i , tem-se $t' \times d$, $t' \times t'$ e $t' \times 1$ respectivamente, em que $t' = t$ e $d = p$ em casos em que os parâmetros θ são definidos unicamente pelos parâmetros β ou $t' = 2t$ e $d = p + q$ para os casos cuja relação é dada por $\theta = (\beta^\top, \gamma^\top)^\top$. Vale ressaltar que t é determinado pela quantidade de observações existentes em cada unidade experimental i , p é expresso como a dimensão de β e q representa a dimensão de γ , com $i = 1, \dots, n$. Partindo para a convergência do processo iterativo descrito em (5.11), de maneira simplificada, dado a retirada do somatório, a mesma é dada por:

$$\hat{\theta} = \left(\mathbf{M}^\top \hat{\mathbf{W}} \mathbf{M} \right)^{-1} \mathbf{M}^\top \hat{\mathbf{W}} \mathbf{M} \mathbf{z},$$

de modo que $\mathbf{M} = (\mathbf{M}_1^\top, \dots, \mathbf{M}_n^\top)^\top$, $\hat{\mathbf{W}} = \text{diag}(\hat{\mathbf{W}}_1, \dots, \hat{\mathbf{W}}_n)$ e $\mathbf{z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_n^\top)^\top$.

Dado o processo iterativo exposto, \mathbf{M} constitui uma matriz de covariáveis, $\hat{\mathbf{W}}$ é uma matriz de pesos e \mathbf{z} tem a função de uma variável dependente modificada. Ao longo do processo de iteração, $\hat{\mathbf{W}}$ e \mathbf{z} se alteram a cada passo do processo.

5.2.2 Ponto influente associado à distância de Cook

Tendo em vista quem um ponto influente pode ser compreendido como um ponto cujo perfil diverge dos demais no que se refere aos valores da variável resposta e, para além disso, apresenta um valor elevado em sua matriz de projeção \mathbf{H} . A distância de Cook (Cook, 1977) identifica pontos influentes por meio do efeito da exclusão em cada valor observado, ou seja, mede o distanciamento entre os valores estimados do vetor paramétrico utilizando-se de todas as observações () e sem o ponto observado como de possível influência $y_{ij}(\hat{\boldsymbol{\theta}}_{(ij)})$, $i=1,\dots,n$ e $j=1,\dots,t'$.

Além disso, como geralmente a obtenção da forma fechada para $(\hat{\boldsymbol{\theta}}_{(ij)})$ não é possível, aplica-se aproximação de um passo. Esse processo consiste em tomar a primeira iteração do processo iterativo pelo método *scoring* de Fisher (ver 5.11) no momento o mesmo é iniciado em $\hat{\boldsymbol{\theta}}$. No caso dos modelos de regressão com medidas repetidas, a aproximação estabelecida por Pregibon em 1981, pode ser escrita como:

$$\hat{\boldsymbol{\theta}}_{(ij)}^{(1)} = \hat{\boldsymbol{\theta}} - \frac{[\mathbf{M}^\top \hat{\mathbf{W}} \mathbf{M}]^{-1} [\mathbf{M}^\top \hat{\mathbf{W}}^{1/2} \mathbf{e}_{(ij)}] [\mathbf{e}_{(ij)}^\top \hat{\mathbf{W}}^{1/2} \hat{\boldsymbol{\Lambda}}^{-1} \hat{\mathbf{b}}]}{1 - h_{ij}}.$$

A análise gráfica do processo descrito acima consiste em fazer $(DC)_{ij}$ $i=1,\dots,n$ e $j=1,\dots,t'$, versus o índice i e indicar como um ponto influente como aquele cujo o valor é elevado quando comparado com os demais. No que se refere ao ajuste dos parâmetros, a distância de Cook para a unidade experimental i é dada por $(DC)_{\beta i} = \sum_j = 1^\top (DC)_{ij}/t$, com $i=1,\dots,n$. Já para processos que envolvem o ajuste do parâmetro $\boldsymbol{\gamma}$, a distância de Cook para a unidade experimental i é dada por $(DC)_{\gamma i} = \sum_j = t + 1^2 t (DC)_{ij}/t$ com $i=1,\dots,n$.

5.2.3 Ponto aberrante associado ao resíduo padronizado

O resíduo padronizado associado à observação y_{ij} é definido por:

$$(r_{PD})_{ij} = \frac{\mathbf{e}_{(ij)}^\top \hat{\mathbf{W}}^{1/2} (\mathbf{z} - \hat{\mathbf{v}})}{\sqrt{1 - h_{ij}}} = \frac{\mathbf{e}_{(ij)}^\top \hat{\mathbf{W}}^{1/2} \hat{\boldsymbol{\Lambda}}^{-1} \hat{\mathbf{b}}}{\sqrt{1 - h_{ij}}},$$

de modo que $\hat{\boldsymbol{\Lambda}} = \text{diag}(\hat{\boldsymbol{\Lambda}}_1, \dots, \hat{\boldsymbol{\Lambda}}_n)$, $\hat{\mathbf{b}} = \text{diag}(\hat{\mathbf{b}}_1^\top, \dots, \hat{\mathbf{b}}_n^\top)^\top$, $\mathbf{e}_{(ij)}$ é um vetor de tamanho adequado com valor 1 na posição da referida observação y_{ij} e 0 nas demais posições;

h_{ij} apresenta-se como o j -ésimo elemento da diagonal principal de \mathbf{H}_i , com $i=1,\dots,n$ e $j=1,\dots,t'$.

A identificação gráfica de um ponto aberrante associado ao resíduo padronizado consiste em utilizar o resíduo padronizado, $(r_{PD})_{ij}$, com $i=1,\dots,n$ e $j=1,\dots,t'$.

Espinheira et al. (2008) evidenciam o resíduo padronizado é o único capaz de mostrar claramente um ponto que foi gerado como uma observação aberrante quando se considera que todas as respostas são independentes e são modeladas por meio de uma distribuição beta com homogeneidade do parâmetro de dispersão.

5.2.4 Ponto de alavanca associado à matriz de projeção

Tomando por base a Equação (5.11), $\hat{\boldsymbol{\theta}}$ pode ser compreendido como a solução de mínimos quadrados da regressão normal linear de $\hat{\mathbf{W}}^{1/2}\mathbf{z}$ com a matriz de covariáveis dada por $\hat{\mathbf{W}}^{1/2}\mathbf{M}$, com isso o resíduo ordinário pode, compreendido como a diferença entre os valores observados e ajustados, pode ser expresso por:

$$r_O = \hat{\mathbf{W}}^{1/2}(\mathbf{z} - \hat{\mathbf{v}}) = (\mathbf{I} - \mathbf{H})\hat{\mathbf{W}}^{1/2}\mathbf{z},$$

com $\hat{\mathbf{v}} = (\hat{\mathbf{v}}_1^\top, \dots, \hat{\mathbf{v}}_n^\top)^\top$, \mathbf{I} a matriz identidade e \mathbf{H} uma matriz do tipo bloco diagonal dada por $\mathbf{H} = \text{diag}(\mathbf{H}_1, \dots, \mathbf{H}_n)$, de modo que:

$$\mathbf{H}_i = \hat{\mathbf{w}}_i^{1/2}\mathbf{M}_i(\mathbf{M}^\top\hat{\mathbf{w}}\mathbf{M})^{-1}\mathbf{M}_i^\top\hat{\mathbf{w}}_i^{1/2},$$

com dimensão $t' \times t'$, para todo $i=1,\dots,n$. Além disso, a matriz \mathbf{H} é simétrica e indepotente. Dado que $\mathbf{W}^{1/2}\mathbf{z}$ desempenha o papel do vetor resposta, \mathbf{H} é denominada de matriz de projeção ortogonal. A partir disso, evidencia-se a escolha pelos elementos da diagonal principal de \mathbf{H} para a identificação de pontos de alavanca como Pregibon (1981) realizou para os Modelos Lineares Generalizados e Venezuela et al. (2007) fizeram para os caso dos MLG's na presença de medidas repetidas.

Diante disso, caso a j -ésima observação da diagonal principal de \mathbf{H}_i , h_{ij} , possuir um valor alto em comparação com os valores das demais observações, com $i=1,\dots,n$ e $j=1,\dots,t$, a j -ésima observação da i -ésima unidade experimental poderá ser um ponto de alavanca quando realizado o ajuste dos parâmetros $\boldsymbol{\beta}$. A utilização do ferramental gráfico consiste na plotagem de h_{ij} versus i sendo capaz de indicar a ordem em que cada unidade experimental aparece na base de dados. Partindo para a análise da modelagem conjunta

da média e do parâmetro de dispersão, um ponto de alavanca no ajuste dos parâmetros de γ pode ser evidenciado quando h_{ij} possuir um valor alto em comparação com os valores das demais observações, com $i=1, \dots, n$ e $j=i+1, \dots, 2t$. Para tal, o procedimento gráfico é baseado na plotagem de h_{ij} versus i , em que $i=1, \dots, n$ e $j=i+1$.

5.3 Critérios de seleção de modelos e de matriz de correlação de trabalho

Tomando por base que o critério de informação de Akaike (AIC) permite a seleção do modelo mais parcimonioso entre os modelos sujeitos a um ajuste, revelando aquele que possui um bom ajuste com um número reduzido de parâmetros.

Segundo Venezuela (2008), a métrica proposta é considerada um estimador não viesado para a informação de Kullback-Leibler (Kullback e Leibler, 1951), em que a mesma é capaz de medir a discrepância entre um modelo concorrente e o verdadeiro modelo. Diante do exposto, observa-se que o modelo escolhido dentre aqueles que compõem o conjunto de modelos concorrentes será aquele que minimiza a medida AIC expressa a seguir:

$$AIC = -2\ell(\hat{\beta}) + 2p,$$

de modo que $\ell(\cdot)$ corresponde ao logaritmo da função de verossimilhança atribuída aos dados, $\hat{\beta}$ o estimador de máxima verossimilhança tomando por base o modelo concorrente e p a dimensão de β .

Dado que a fundamentação das equações de estimação generalizadas não são baseadas em verossimilhanças, observa-se a proposta de Pan (2001) que apresenta uma modificação na medida AIC sugerindo que ao admitir independência entre todas as observações, o modelo de equação de estimação apresentado por Liang e Zeger (1986) é análogo a função quasi-escore (MCCULLAGH e NELDER, 2013). Ainda, se restrito à família exponencial, a medida proposta por Pan (2001), torna-se equivalente à função escore expressa para tal família de distribuições.

Com isso, a métrica apresentada, denominada por *Quasilikelihood Information Criterion* (QIC), mostra-se útil tanto em processos de seleção de modelos quanto para a definição de uma matriz de correlação de trabalho. A partir do exposto, na presença de independência entre as observações e supondo homogeneidade da dispersão as equações de estimação apresentadas no presente estudo apresentam, em linha com Venezuela (2008),

equivalência com as suas respectivas funções escores.

Dito isto, a medida QIC pode ser definida, tomando por base o logaritmo da função de verossimilhança, como:

$$QIC = -2\ell(\hat{\beta}(\mathbf{R})) + 2tr(\hat{\mathbf{S}}_I \hat{\mathbf{J}}_{nR}^{-1}), \quad (5.12)$$

em que $\ell(\cdot)$ corresponde ao logaritmo da função de verossimilhança que constitui a função escore equivalente à equação de estimação quando a independência entre as observações é admitida, $\hat{\mathbf{S}}_I$ a matriz de sensibilidade quando estabelecido a estrutura de independência e $\hat{\mathbf{J}}_{nR}^{-1}$ o estimador empírico sob a estrutura de \mathbf{R} . Assim, os mesmos são avaliados em $\hat{\beta}(\mathbf{R})$ que corresponde à estimativa de β gerada por meio da matriz de correlação de trabalho \mathbf{R} . Finalmente, os parâmetros $\hat{\mathbf{S}}_I$ e $\hat{\mathbf{J}}_{nR}^{-1}$ serão assintoticamente análogos e $tr(\hat{\mathbf{S}}_I \hat{\mathbf{J}}_{nR}^{-1}) \approx 2$ quando estiverem corretas todas as especificações da modelagem via equações de estimação (Pan, 2001).

Considerando a métrica QIC definida em (5.12), a mesma pode ser reduzida para

$$QIC = -2\ell(\hat{\beta}(\mathbf{R})) + 2p, \quad (5.13)$$

a qual pode ser eventualmente utilizada no processo de seleção de covariáveis.

Destaca-se que as métricas apresentadas acima serão ditas verdadeiras quando o parâmetro de dispersão (ϕ^{-1}) for conhecido e único. Nos casos em que ϕ^{-1} for desconhecido, a avaliação da métrica QIC é dada pela utilização do maior valor estimado para o parâmetro de dispersão dentre os modelos candidatos e a métrica QIC_s utilizando a estimativa para ϕ^{-1} derivada do modelo ajustado com todas as covariáveis. Tal processo é realizado para que as medidas definidas em (5.12) e (5.13) possam alterar seus valores apenas em função da mudança da matriz de correlação de trabalho ou da inclusão ou exclusão de variáveis no modelo.

Tomando por base o estudo apresentado em Botter et.al (2006) e Venezuela (2008), o próximo capítulo apresentará uma aplicação com base em dados reais perpassando os assuntos abordados nos capítulos 4, 5 e 5.2.

5.4 Influência local para dados com medidas repetidas

O objetivo principal da utilização de influência local está centrado na necessidade de identificar, através de métricas apropriadas, a robustez do modelo à luz do efeito exercido

pelas pequenas perturbações no mesmo ou em seus dados. Se essas perturbações causarem efeitos desproporcionais em componentes específicos do modelo, então, há indícios de que ele está mal ajustado ou que existem afastamentos nas suposições do modelo proposto. Nesse sentido, a identificação das observações geradoras de discrepância pode auxiliar na definição de um modelo mais adequado, por sua vez robusto, aos dados.

E observado em Cook(1986) a proposta de um procedimento de avaliação para a influência desses pontos. No caso, busca-se examinar o que pequenas perturbações em componentes do modelo causam na superfície do afastamento pela verossimilhança (likelihood displacement), definido por

$$LD_{\omega} = 2 \{ \ell(\hat{\theta}) - \ell(\hat{\theta}_{\omega}) \},$$

em que $\omega = (\omega_1, \dots, \omega_N)^{\top}$, $\omega \in \Omega \subset \mathbb{R}^N$ é o vetor de perturbações (N) com $N = nt$, $\ell(\hat{\theta})$ e $\ell(\hat{\theta}|\omega)$ o logaritmo da função de verossimilhança para o modelo postulado e o logaritmo da função de verossimilhança para o modelo em perturbação, respectivamente.

De acordo com Cadigan e Farrell (2002), é possível descrever tal medida de influência local para um caso mais genérico, ou seja, é possível avaliar o afastamento de uma função de ajuste $\mathcal{F}(\theta)$, sendo esta duplamente diferenciável em θ e que possua como estimador de θ , representado por $\hat{\theta}$, a solução de

$$\Psi(\hat{\theta}) = \left[\frac{\partial \mathcal{F}(\theta)}{\partial \theta} \right] \Bigg|_{\theta=\hat{\theta}} = 0. \quad (5.14)$$

Com isso, a medida de afastamento pela função de ajuste (FD) é

$$FD_{\omega} = 2 \{ \mathcal{F}(\hat{\theta}) - \mathcal{F}(\hat{\theta}_{\omega}) \},$$

de modo que $\hat{\theta}_{\omega}$ é a estimativa que maximiza a função de ajuste com perturbação $\mathcal{F}(\theta|\omega)$. Vale ressaltar que nos casos em que a função de ajuste é definida pelo logaritmo da função de verossimilhança, observa-se que $FD_{\omega} = LD_{\omega}$. Dentro desse cenário, dado que \mathcal{F} pode alguma outra escolha, por exemplo, a função quase-verossimilhança, a proposta de Cadigan e Farrell mostra-se mais abrangente do que a proposta definida por Cook.

A seguir, em linha com Venezuela (2008), será apresentada a proposta de Cook(1986) sob a forma mais geral de Cadigan e Farrell (2002), definida como influência local generalizada.

5.5 Influência local generalizada

Admitido que o uso da medida $FD(\boldsymbol{\omega})$ é adequado para comparar $\boldsymbol{\theta}$ e $\boldsymbol{\theta}_{\boldsymbol{\omega}}$, no que se refere aos contornos de uma função qualquer de ajuste quando variado $\boldsymbol{\omega}$ em $\boldsymbol{\Omega}$. Contudo, realizar a avaliação de $FD(\boldsymbol{\omega})$ para todo $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ pode ser inviável posto que há uma infinidade de valores que $FD(\boldsymbol{\omega})$ pode vir a assumir. Diante disso, Cook(1986) propõe a análise do comportamento local de $FD(\boldsymbol{\omega})$ para qualquer valor de $\boldsymbol{\omega}$ pertencente a uma vizinhança $\boldsymbol{\omega}_0$, compreendido como um vetor de não perturbação definido anteriormente. Analogamente, tem-se que $\mathcal{F}(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \mathcal{F}(\boldsymbol{\theta}) \Rightarrow FD_{\boldsymbol{\omega}_0} = 0$.

Com isso, a proposta de Cook(1986) é estudar a curvatura normal (Bates e Watts, 1980) da linha projetada no gráfico $FD(\boldsymbol{\omega}_0 + \mathbf{a}\mathbf{d}) \times a$, em que $a \in \mathbb{R}$ e \mathbf{d} é apresentado como uma direção arbitrária de norma igual a um ($\|\mathbf{d}\|=1$). Em Cook(1986) é mostrado que a curvatura normal da direção \mathbf{d} é dada por

$$C_d = (\boldsymbol{\theta}) = 2|\mathbf{d}^\top \Delta^\top \ddot{\mathcal{F}} \Delta \mathbf{d}|,$$

de forma que $-\ddot{\mathcal{F}}$ é a matriz observada de Fisher

$$\ddot{\mathcal{F}} = \frac{\partial^2 \mathcal{F}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \frac{\partial \Psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}, \quad (5.15)$$

em que $\Psi(\cdot|\cdot)$ é definido como o vetor gradiente da função de ajuste de $F(\cdot|\cdot)$. Além disso, Δ é compreendido como a matriz

$$\Delta = \frac{\partial^2 \mathcal{F}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} = \frac{\partial \Psi(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top}, \quad (5.16)$$

com todas as unidades avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}_0$. Com isso, para identificar as observações que, na presença de pequenas perturbações, atuam realizando notável influência local em $FD(\boldsymbol{\omega}_0)$, deve-se estudar a direção do autovetor d_{\max} correspondente à linha projetada de maior curvatura C_{\max} que é obtida pelo maior autovalor da matriz

$$-\Delta^\top \ddot{\mathcal{F}}^{-1} \Delta. \quad (5.17)$$

Recomenda-se que o leitor interessado na temática consulte Lobato(2005) e Ospina (2007), para aprofunda a compreensão mais detalhada da ideia apresentada em Cook(1986).

Comumente, utiliza-se os seguintes gráficos de diagnóstico de influência local:

- Gráfico de índice *versus* d_{\max_i} ;
- Gráfico de índice *versus* curvatura normal padronizada C_i , com:

$$C_i = \frac{C_{hi}}{\sum_{j=1}^n C_{hj}},$$

de modo que \mathbf{h}_i é definido como um vetor unitário na direção de i -ésima posição. Em Poon e Poon (2002), outras formas para a padronização de C_h .

Outro aspecto a ser notado é o fato de que é possível avaliar a influência local apenas para um subvetor $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta}$, admitindo que tal vetor pode ser particionado de modo que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. A partir disso, a curvatura normal na direção \mathbf{d} é definida como:

$$C_d(\boldsymbol{\theta}_1) = 2 \left| \mathbf{d}^\top \Delta^\top (\ddot{\mathcal{F}}^{-1} - \ddot{\mathcal{F}}^{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2}) \Delta \mathbf{d} \right|,$$

de modo que,

$$\ddot{\mathcal{F}}^{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2} = \begin{bmatrix} 0 & 0 \\ 0 & \ddot{\mathcal{F}}_{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2}^{-1} \end{bmatrix},$$

em que $\ddot{\mathcal{F}}^{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2}$ é avaliada em $\hat{\boldsymbol{\theta}}$. Além disso, a partir do gráfico de índices *versus* d_{\max_i} da matriz $-\Delta^\top (\ddot{\mathcal{F}}^{-1} - \ddot{\mathcal{F}}^{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2}) \Delta$ é possível realizar a identificação de possíveis informações influentes em $\hat{\boldsymbol{\theta}}_1$. Analogamente, a curvatura normal para o subvetor $\boldsymbol{\theta}_2$ na direção \mathbf{d} é apresentada a seguir

$$C_d(\boldsymbol{\theta}_2) = 2 \left| \mathbf{d}^\top \Delta^\top (\ddot{\mathcal{F}}^{-1} - \ddot{\mathcal{F}}^{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1}) \Delta \mathbf{d} \right|,$$

com

$$\ddot{\mathcal{F}}^{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1} = \begin{bmatrix} \ddot{\mathcal{F}}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

novamente com $\ddot{\mathcal{F}}^{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1}$ avaliada em $\hat{\boldsymbol{\theta}}$. Além do mais, o gráfico de índices *versus* d_{\max_i} da matriz $-\Delta^\top (\ddot{\mathcal{F}}^{-1} - \ddot{\mathcal{F}}^{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1}) \Delta$ realiza a identificação de possíveis informações influentes em $\hat{\boldsymbol{\theta}}_2$.

5.6 Influência local para equações de estimação

Diate do exposto anteriormente na Seção 4.6, apresenta-se que medida de influência local proposta por Cadigan e Farrell (2002) pode ser constituída de a partir de

uma função genérica $\mathcal{F}(\boldsymbol{\theta})$ desde que exista e satisfaça (5.14). Dando continuidade com a Seção anterior, foi apresentado que tomando por base (5.14), a definição da medida de influência local pode ser realizada através do vetor gradiente $\boldsymbol{\Psi}(\boldsymbol{\theta})$ sem obrigatoriamente haver o conhecimento da função $\mathcal{F}(\boldsymbol{\theta})$ que o gera.

Assim, no que tange às equações de estimação generalizadas, é dado que há o desconhecimento da função de verossimilhança, bem como a função de ajuste responsável por gerar tal equação. Daí, em linha com Venezuela (2008), tal processo é garantido quando se assume que qualquer EEG (que utiliza a verdadeira matriz de correlação quando construída a partir de (4.4) ou que apresenta uma matriz de correlação de trabalho $\mathbf{R}(\boldsymbol{\alpha})$ conhecida, pois, dessa forma, é possível satisfazer as propriedades de quasi-verossimilhança citadas por McCullagh e Nelder (2013).

Nesse sentido, tais autores descrevem que uma função de quasi-escore com observações dependentes pode ser um vetor gradiente de quasi-verossimilhança, sob condição de que a derivada dessa função quasi-escore com relação à $\boldsymbol{\beta}$ seja uma matriz simétrica ou, analogamente, desde que as derivadas dos componentes de $\text{Cov}(y_i)^{-1}$ relacionados à $\boldsymbol{\mu}_i$ sejam iguais sob permutação de três índices. Em outros termos, admite-se a relação acima, desde que a expressão a seguir seja satisfeita.

$$\frac{\partial \text{Cov}(y_{ij}, y_{il})^{-1}}{\partial \mu_{ik}} = \frac{\partial \text{Cov}(y_{ij}, y_{ik})^{-1}}{\partial \mu_{il}} = \frac{\partial \text{Cov}(y_{il}, y_{ik})^{-1}}{\partial \mu_{ij}},$$

com $i = 1, \dots, n$ e $j, l, k = 1, \dots, t$. No presente caso, verifica-se que as derivadas dos componentes da matriz $\text{Cov}(\mathbf{b}_i)^{-1}$ definida em (4.5) no que se referem à $\boldsymbol{\mu}_i$ são ditas iguais sob a permutação dos três índices j, l e k , com $i = 1, \dots, n$ e $j, l, k = 1, \dots, t$. Destaca-se que tal processo ocorre quando há a correta utilização da verdadeira matriz de correlação $\mathbf{R}(\mathbf{b}_i)$ ou quando uma matriz de correlação de trabalho $\mathbf{R}(\boldsymbol{\alpha})$ conhecida é utilizada, visto que em ambos os casos não ocorre a dependência de $\boldsymbol{\mu}$.

Diante disso, é possível garantir que qualquer equação de estimação $\boldsymbol{\Psi}(\boldsymbol{\theta})$ que utiliza a verdadeira matriz de correlação ou matriz de correlação de trabalho conhecida é um vetor gradiente de uma função de ajuste \mathcal{F} . Ou seja, pode-se admitir que

$$\exists \mathcal{F}(\boldsymbol{\theta}) : \frac{\partial \mathcal{F}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\Psi}(\boldsymbol{\theta})$$

e

$$\boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}) = \mathbf{0},$$

de modo que $\hat{\theta}$ é o ponto de máximo da função de ajuste $\mathcal{F}(\theta)$. Com isso, pode-se haver a utilização da matriz Δ apresentada na Equação (5.16) para a obtenção do autovetor \mathbf{d}_{\max} a partir da equação (5.17).

Tendo em vista que a matriz $\ddot{\mathcal{F}}$ definida em (5.15) não pode ser obtida, é visto que Cadingan (1995) propõe a simplificação da mesma fazendo o uso de seu respectivo valor esperado. Diante do exposto, o valor esperado de $\ddot{\mathcal{F}}$ é definido pela matriz de sensibilidade apresentada na Equação (4.2).

Dessa forma, assumindo que $\Psi(\theta)$ é um vetor gradiente de uma função de ajuste \mathcal{F} , ainda que desconhecida e que (5.14) é satisfeita, Venezuela (2008) utiliza a ideia de Cook(1986) para a construção de uma medida de influência local para equação de estimação dada pelo autovetor \mathbf{d}_{\max} , correspondendo ao maior autovalor da matriz

$$-\Delta^\top \mathbf{S}^{-1} \Delta, \quad (5.18)$$

de modo que

$$\Delta = \frac{\partial \Psi(\theta | \omega)}{\partial \omega}$$

e

$$\mathbf{S} = E(\ddot{\mathcal{F}}) = E\left(\frac{\partial \Psi(\theta)}{\partial \theta^\top}\right)$$

são avaliadas em $\theta = \hat{\theta}$ e $\omega = \omega_0$. Reforça-se que pela adotada para a construção do presente estudo, quando se está trabalhando com modelagem da média sob homogeneidade do parâmetro de dispersão, o vetor de parâmetros θ representa apenas os parâmetros de regressão β . Diante disso, o gráfico de índice *versus* \mathbf{d}_{\max} pode revelar quais observações são influentes na estimação de β . No cenário em que há a ocorrência de um modelagem conjunta da média e da dispersão, \mathbf{d}_{\max} atua revelando quais as observações são influentes na estimação de $\theta = (\beta^\top, \gamma^\top)^\top$. Em tal situação, caso o interesse seja a identificação de observações influentes apenas na estimação dos parâmetros β ou γ , o conceito de partição, apresentado na Seção 6.1 poderá ser utilizado.

Dessa maneira, para a identificação de observações somente na estimação de β , a curvatura normal na direção d é dada por $C_d(\beta) = 2|d^\top \Delta^\top (\mathbf{S}^{-1} - \mathbf{S}^{\gamma\gamma}) \Delta d|$, em que

$$\mathbf{S}^{\gamma\gamma} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{S}_{\gamma\gamma}^{-1} \end{bmatrix},$$

com $\mathbf{S}_{\gamma\gamma}$ avaliada em $\hat{\theta}$.

Dessa forma, o gráfico de índices do maior autovetor de $-\Delta^\top(\mathbf{S}^{-1} - \mathbf{S}^\gamma)\Delta$ pode apontar as observações influentes, de acordo com o esquema de perturbação estudado, na estimação de β .

De forma semelhante, a curvatura normal para o vetor de parâmetro γ na direção d é dada por $C_d(\gamma) = 2|d^\top \Delta^\top(\mathbf{S}^{-1} - \mathbf{S}^{\beta\beta})\Delta d|$, em que

$$\mathbf{S}^{\beta\beta} = \begin{bmatrix} \mathbf{S}_{\beta\beta}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

com $\mathbf{S}_{\beta\beta}$ avaliada em $\hat{\theta}$.

Dessa forma, o gráfico de índices do maior autovetor de $-\Delta^\top(\mathbf{S}^{-1} - \mathbf{S}^{\beta\beta})\Delta$ pode apontar as observações influentes, de acordo com o esquema de perturbação estudado, na estimação de γ .

Tomando o exposto acima, serão apresentadas as medidas de influência local para alguns esquemas de perturbação na composição das equações de estimação generalizadas, como propostas por Venezuela (2008) para os modelos de regressão beta com medidas repetidas. De acordo com Venezuela (2008):

Essas medidas são desenvolvidas para os esquemas de perturbação ponderação de casos, da variável resposta, de uma covariável contínua da matriz de planejamento, no parâmetro de precisão e, finalmente, na matriz de correlação de trabalho.

Todos esses esquemas são considerados para a modelagem sob homogeneidade do parâmetro de dispersão, sendo os três primeiros, para a modelagem conjunta da média e do parâmetro de dispersão o qual admite dispersão variável.

Ainda sobre tais esquemas de perturbação, os mesmos atuam na avaliação do comportamento das medidas de influência local de diversas formas. Nesse caso, a ponderação de casos pode ser compreendida como uma perturbação na variância de cada unidade experimental, especialmente para o modelos normais lineares (Thomas e Cook, 1989). Perturbação na variável resposta pode ser vista como um meio alternativo de identificar *outliers* (Schwarzmann, 1991).

O esquema de perturbação individual das covariáveis auxilia na avaliação da influência de cada uma no processo estimativo, além de checar sua sensibilidade à valores altos. Porém, tal esquema apenas faz sentido caso a covariável seja de natureza contínua.

Perturbação no parâmetro de precisão atua indicando a sensibilidade do modelo em relação a suposição de homoscedasticidade. Por fim, a perturbação da matriz de correlação de trabalho pode indicar a necessidade da utilização de uma outra estrutura de matriz de correlação, por exemplo.

5.7 Esquemas de perturbação sob homogeneidade da dispersão

De acordo com Venezuela (2008), a equação de estimação generalizada para modelagem da média admitindo a suposição de homogeneidade do parâmetro de dispersão para o modelo de regressão beta pode ser expressa por:

$$\Psi_1(\beta) = \mathbf{X}^\top \Lambda \Omega^{-1} \mathbf{b} = \mathbf{X}^\top \mathbf{W} \Lambda^{-1} \mathbf{b}, \quad (5.19)$$

com $\Lambda = \mathbf{G}\mathbf{A}$ e $\mathbf{W} = \Lambda \Omega^{-1} \Lambda$. Além disso, a matriz de sensibilidade, definida em (4.3), é definida a seguir como:

$$\mathbf{S} = -\mathbf{X}^\top \mathbf{W} \mathbf{M}, \quad (5.20)$$

de modo que $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_t^\top)^\top$, $\Lambda = \text{diag}(\Lambda_1, \dots, \Lambda_n)$, $\Omega = \text{diag}(\Omega_1, \dots, \Omega_n)$, $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$ e $\mathbf{b} = \text{diag}(\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$.

Diante do exposto, as medidas de influência local, descritas a seguir para alguns esquemas de perturbação com suposição de heterogeneidade da dispersão, com definições apresentadas por meio de (5.43) e (5.44), avaliadas de acordo com as estimativas do modelo pressuposto $(\hat{\beta}^\top, \hat{\phi}^\top)^\top$ e em ω_0 .

5.7.1 Ponderação de casos

Considerando o seguinte esquema de perturbação

$$\Psi_1(|\omega) = \mathbf{X}^\top \mathbf{W} \Lambda^{-1} \text{diag}(\omega) \mathbf{b}, \quad (5.21)$$

em que $\omega = (\omega_1^\top, \dots, \omega_n^\top)^\top$, com $\omega_i = (\omega_{i1}, \dots, \omega_{in})^\top$, $i=1, \dots, n$. Nesse caso, ω_0 é um vetor que atua representando a ausência de perturbação. Com isso ω_0 assume $\omega_{ij} = 1$, com $i=1, \dots, n$ e $j=1, \dots, t$.

Para o modelo de perturbação definido em (5.21), facilmente é possível definir que $\Delta = \mathbf{X}^\top \mathbf{W} \Lambda^{-1} \text{diag}(\mathbf{b})$. Desse modo, a matriz apresentada em (5.17) é dada a seguir

por

$$\text{diag}(\mathbf{b})\Lambda^{-1}\mathbf{W}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}\Lambda^{-1}\text{diag}(\mathbf{b}), \quad (5.22)$$

de modo que se encontre avaliada em $\boldsymbol{\omega}_0$ e $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\phi}})^\top$.

Finalmente, de acordo com Venezuela (2008), quando temos que a equação (5.21) é definida a partir da equação de estimação para o modelo de regressão beta (5.3) e com todas as observações independentes entre si, a matriz Δ definida para a ponderação de casos corresponde à matriz apresentada por Ospina (2007, Seção 3.4.1).

5.7.2 Perturbação da variável resposta

Tomando por base um esquema aditivo de perturbação da variável resposta y_{ij} é dado por

$$y_{\omega ij} = y_{ij} + \omega_{ij}\sqrt{\text{Var}(y_{ij})}, \quad (5.23)$$

com $\omega_{ij} = 0$ equivalente à $\boldsymbol{\omega}_0 = \mathbf{0}$, ou seja, atua representando a ausência de perturbação.

Partindo para o estudo da equação apresentada em (5.19), é possível identificar que a única parte dependente da variável resposta está no vetor \mathbf{b} . De tal modo, deve-se considerar \mathbf{b}_ω o vetor \mathbf{b} com perturbação na variável resposta. A partir disso, a equação de estimação perturbada pode ser escrita como

$$\Psi_1(\boldsymbol{\beta}|\boldsymbol{\omega}) = \mathbf{X}^\top\mathbf{W}\Lambda^{-1}\mathbf{b}_\omega, \quad (5.24)$$

em que $\mathbf{b}_\omega = (\mathbf{b}_{\omega 1}^\top, \dots, \mathbf{b}_{\omega n}^\top)^\top$, com $\mathbf{b}_{\omega i} = (\mathbf{b}_{\omega i 1}^\top, \dots, \mathbf{b}_{\omega i t}^\top)^\top$, $i = 1, \dots, n$.

Diante do exposto, a matriz apresentada em (5.18) fica definida por $\Delta = \mathbf{X}^\top\mathbf{W}\Lambda^{-1}\mathcal{B}$. Além disso, $\mathcal{B} = \partial\mathbf{b}_\omega/\partial\boldsymbol{\omega}^\top$.

Com base na equação (5.17), a medida de influência local com perturbação da variável resposta pode ser obtida por meio da matriz a seguir

$$\mathcal{B}^\top\Lambda^{-1}\mathbf{W}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}\Lambda^{-1}\mathcal{B}, \quad (5.25)$$

avaliada em $\boldsymbol{\omega}_0$ e $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\phi}})^\top$.

No contexto da Equação (5.3), os vetores de $\mathbf{b}'s$ são definidos de diferentes formas, tal fato ocasiona medidas de influência local distintas para cada caso. Diante de

tal fato, para o contexto do modelo de regressão beta, na equação (5.3) o vetor perturbado na variável resposta referente à i -ésima unidade experimental $\mathbf{b}_{\omega i}$ é dado por

$$\mathbf{b}_{\omega i} = \mathbf{y}_{\omega i}^* - \boldsymbol{\mu}_i^*, \quad (5.26)$$

de modo que $\mathbf{y}_{\omega i}^* = (y_{\omega i 1}^*, \dots, y_{\omega i t}^*)^\top$, com $y_{\omega i j}^* = \log\left(\frac{y_{\omega i j}}{1 - y_{\omega i j}}\right)$, $i = 1, \dots, n$ e $j = 1, \dots, t$.

5.7.3 Perturbação individual das covariáveis

Adotando a existência de um esquema aditivo de perturbação na k -ésima coluna da matriz de covariáveis \mathbf{X} , com $\mathbf{x}_k = (x_{11k}, x_{12k}, \dots, x_{ntk})$, em linha com a proposta sugerida por Thomas de Cook, (1990). Diante do exposto, o vetor perturbado $\mathbf{x}_{\omega k}$ tem cada componente expresso por

$$x_{\omega i j k} = x_{i j k} + \omega_{i j} s_{x_k}, \quad (5.27)$$

com s_{x_k} representando um fator de escala obtido por meio do desvio padrão de \mathbf{x}_k , com $i = 1, \dots, n$ e $j = 1, \dots, t$. Buscando exemplificar, se $k \neq 2$ e $k \neq p$, o modelo dado em (3.3) de acordo com o esquema descrito acima em (5.27) é

$$g(\boldsymbol{\mu}_{\omega i j}) = \boldsymbol{\eta}_{\omega i j} = \beta_1 + x_{i j 2} \beta_2 + \dots + x_{\omega i j k} \beta_k + \dots + x_{i j p} \beta_p. \quad (5.28)$$

Em tal caso, a perturbação $\boldsymbol{\omega}_0 = \mathbf{0}$.

Como visto na Equação (5.19), verifica-se que todos os seus componentes são modificados quando há a perturbação em uma covariável. Assim, Venezuela (2008) propõe a seguinte equação de estimação de perturbação

$$\Psi_1(\boldsymbol{\beta} | \boldsymbol{\omega}) = \mathbf{X}^\top \boldsymbol{\Lambda}_\omega \boldsymbol{\Omega}_\omega^{-1} \mathbf{b}_\omega, \quad (5.29)$$

com $\boldsymbol{\omega}$ indicando que as matrizes $\mathbf{X}, \boldsymbol{\Lambda}$ e $\boldsymbol{\Omega}$, bem como o vetor \mathbf{b} são dependentes, de algum modo, da perturbação apresentada em (5.27).

Consequentemente, de acordo com Herville (1997) a derivada de $\Psi_1(\boldsymbol{\beta} | \boldsymbol{\omega})$ com relação à $\boldsymbol{\omega}^\top$ pode ser expressa a seguir como

$$\Delta = \mathbf{X}_\omega^\top \boldsymbol{\Lambda}_\omega \left[\boldsymbol{\Omega}_\omega^{-1} \frac{\partial \mathbf{b}_\omega}{\partial \boldsymbol{\omega}^\top} + \frac{\partial \boldsymbol{\Omega}_\omega^{-1}}{\partial \boldsymbol{\omega}^\top} \text{diag}(\mathbf{b}_\omega) \right] + \left[\mathbf{X}_\omega^\top \frac{\partial \boldsymbol{\Lambda}_\omega}{\partial \boldsymbol{\omega}^\top} + \frac{\partial \mathbf{X}_\omega^\top}{\partial \boldsymbol{\omega}^\top} \text{diag}(\mathbf{b}_\omega) \right] \boldsymbol{\Omega}_\omega^{-1} \text{diag}(\mathbf{b}_\omega), \quad (5.30)$$

de modo que a derivada de \mathbf{X}_ω^\top com relação à ω^\top é uma matriz $p \times N$ de zeros exceto pela k -ésima linha que é formada pela constante s_{x_k} , com $N = nt$, além disso

$$\frac{\partial \Omega_\omega^{-1}}{\partial \omega^\top} = \Omega_\omega^{-1} \frac{\partial \Omega_\omega^{-1}}{\partial \omega^\top} \Omega_\omega^{-1}. \quad (5.31)$$

As derivadas de Λ_ω , Ω_ω^{-1} e \mathbf{b}_ω estão todas com relação à ω^\top e são definidas de acordo com $\Psi_1(\beta)$ definida de acordo com a equação de estimação apresentada em (5.3).

Como exposto anteriormente, a equação de estimação definida em (5.3) representa a EEG para os modelos de regressão beta e sua derivada é representada a seguir por

$$\frac{\partial \Lambda_\omega}{\partial \omega^\top} = \left(\frac{\partial \Lambda_{\omega 1}}{\partial \omega_1^\top}, \dots, \frac{\partial \Lambda}{\partial \omega_n^\top} \right). \quad (5.32)$$

Com isso, para a equação descrita em (5.3), tem-se

$$\begin{aligned} \Lambda_{\omega i} &= \phi \mathbf{G}_i \mathbf{A}_{\omega i}, \\ \Omega_{\omega i} &= \mathbf{A}_{\omega i}^{1/2} \mathbf{R}(\alpha) \mathbf{A}_{\omega i}^{1/2} \text{ e} \\ \mathbf{b}_{\omega i} &= y_i^* - \mu_{\omega i}^*, \end{aligned} \quad (5.33)$$

com $\mathbf{G}_{\omega i} = \text{diag}(\partial g^{-1}(\eta_{\omega i 1})/\partial \eta_{\omega i 1}, \dots, \partial g^{-1}(\eta_{\omega i t})/\partial \eta_{\omega i t})$, $\mathbf{A}_{\omega i} = \text{diag}(a_{\omega i 1}, \dots, a_{\omega i t})$ e $\mu_{\omega i}^* = (\mu_{\omega i 1}^*, \dots, \mu_{\omega i t}^*)^\top$, de modo que $a_{\omega i j} = \phi'(\mu_{ij}\phi) + \phi'((1 - \mu_{ij})\phi)$ e $\mu_{\omega i j}^* = \phi(\mu_{ij}\phi) - \phi((1 - \mu_{ij})\phi)$, com $i = 1, \dots, n$ e $j = 1, \dots, t$.

Aponta-se que nos casos em que todas as observações forem dependentes entre si, a matriz Δ definida para perturbação individual das covariáveis proposta por Venezuela (2008) deverá coincidir com a medida de influência apresentada por Ospina (2007, Seção 3.4.4).

5.7.4 Perturbação do parâmetro de precisão

Considerando uma estrutura de perturbação do parâmetro de precisão, de modo que esse não seja constante ao longo das observações e nem entre os indivíduos, tal parâmetro pode ser compreendido como

$$\phi_{\omega i j} = \frac{\phi}{\omega_{ij}}, \quad (5.34)$$

com $i = 1, \dots, n$ e $j = 1, \dots, t$. EM tal caso de perturbação, $\omega = 1$. Partindo da equação (5.19), obtêm-se a equação de estimação perturbada a seguir

$$\Psi_1(\beta|\omega) = \mathbf{X}^\top \Lambda_\omega \Omega_\omega^{-1} \mathbf{b}_\omega, \quad (5.35)$$

de modo que o índice ω evidencia que as matrizes Λ e Ω e o vetor \mathbf{b} dependem, de alguma maneira, da perturbação definida acima em (5.34). Dessa forma, segundo Harville (1997), a matriz apresentada em (5.16) pode ser expressa a seguir por

$$\Delta = \mathbf{X}^\top \Lambda_\omega \Omega_\omega \frac{\partial \mathbf{b}_\omega}{\partial \omega^\top} + \mathbf{X}^\top \Lambda_{\partial \Omega_\omega}^{-1} \partial \omega^\top \text{diag}(\mathbf{b}_\omega) + \mathbf{X}^\top \frac{\partial \Lambda_\omega}{\partial \omega^\top} \Omega_\omega^{-1} \text{diag}(\mathbf{b}_\omega), \quad (5.36)$$

em que as derivadas de Λ_ω , Ω_ω^{-1} e \mathbf{b}_ω estão todas com relação à ω^\top é definida em (5.3).

Finalmente, a equação $\Psi_1(\beta|\omega)$, a partir de (5.3) pode ser definida para os modelos de regressão beta com medidas repetidas.

Com isso, em linha com a equação (5.3), tem-se que

$$\begin{aligned} \Lambda_{\omega i} &= \mathbf{G}_i \Phi_{\omega i} \mathbf{A}_{\omega i}, \\ \Omega_{\omega i} &= \mathbf{A}_{\omega i}^{1/2} \mathbf{R}(\alpha) \mathbf{A}_{\omega i}^{1/2} \text{ e} \\ \mathbf{b}_{\omega i} &= y_i * -\mu_{\omega i}^*, \end{aligned} \quad (5.37)$$

com $\Phi_{\omega i} = \text{diag}(\Phi_{\omega i1}, \dots, \Phi_{\omega it})$, $\mathbf{A}_{\omega i} = \text{diag}(a_{\omega i1}, \dots, a_{\omega it})$ e $\mu_{\omega i}^* = (\mu_{\omega i1}^*, \dots, \mu_{\omega it}^*)^\top$, de modo que $a_{\omega ij} = \phi'(\mu_{ij} \phi_{\omega ij}) + \phi'((1 - \mu_{ij}) \phi_{\omega ij})$ e $\mu_{\omega ij}^* = \phi(\mu_{ij} \phi_{\omega ij}) - \phi((1 - \mu_{ij}) \phi_{\omega ij})$, com $i = 1, \dots, n$ e $j = 1, \dots, t$.

Deve-se notar que nos casos em que todas as observações forem dependentes entre si, a medida de influência local para perturbação individual das covariáveis proposta por Venezuela (2008) deverá coincidir com a medida de influência apresentada por Ospina (2007, Seção 3.4.4).

5.7.5 Perturbação na matriz de correlação de trabalho $R(\alpha)$

A princípio, supondo que $\mathbf{R}(\alpha)$ seja uma matriz de correlação de trabalho definida numa forma geral dada pelo vetor de correlações $\alpha = (\alpha_{11}, \alpha_{22}, \dots, \alpha_{tt})^\top$, de modo que $\alpha_{jj} = 1$ e $\alpha_{jj'} = \alpha_{j'j}$, com $j \neq j'$ e $j, j' = 1, \dots, t$. De uma maneira simplificada, caso a matriz de correlação de trabalho for a padrão uniforme, então $\alpha_{jj'} = \alpha$; caso seja a auto-regressiva de ordem 1 (AR-1), então $\alpha_{jj'} = \alpha^{|j-j'|}$, para todo $j \neq j'$ e $j, j' = 1, \dots, t$. Reforça-se que a matriz de correlação de trabalho para a mdelagem dos dados é sempre a mesma para todas as unidades observacionais $i, i = 1, \dots, n$.

Em Venezuela (2008) e Venezuela et al. (2011) é observado que os autores propõem um possível esquema de perturbação no vetor de correlações (α) de forma admitir

que esse não seja o mesmo entre as unidades experimentais e nem ao longo das observações. Tal esquema de perturbação é definido como

$$\alpha_{\omega i(jj')} = \frac{\alpha_{jj'}}{\omega_{i(jj')}}, \quad (5.38)$$

de modo que $\alpha_{jj'}=1$ quando $j = j'$ e $\alpha_{jj'} = \alpha_{j'j}$ quando $j \neq j'$, com $i = 1, \dots, n$ e $j, j' = 1, \dots, t$. Para esse tipo de perturbação, $\omega_0 = 1$.

Partindo para a equação de estimação apresentada em (5.19), perturbada de acordo com o esquema (5.38) é definida por

$$\Psi_1(\beta|\omega) = \mathbf{X}^\top \Lambda \Omega_\omega^{-1} \mathbf{b}, \quad (5.39)$$

alternando unicamente a matriz Ω com tal perturbação. Diante disso, a matriz definida em (5.16) é expressa por

$$\Delta = \mathbf{X}^\top \Lambda \frac{\partial \Omega_\omega^{-1}}{\partial \omega^\top} \text{diag}(\mathbf{b}), \quad (5.40)$$

sendo a derivada de Ω_ω^{-1} dada em (5.32) com

$$\frac{\partial \Omega_\omega}{\partial \omega^\top} = \text{Var}(\mathbf{b})^{1/2} \frac{\partial \mathbf{R}(\alpha_\omega)}{\partial \omega^\top} \text{Var}(\mathbf{b})^{1/2}, \quad (5.41)$$

em que $\partial \alpha_{\omega i(jj')}/\partial \omega_{i(jj')} = -\alpha_{jj'}/\omega_{i(jj')}^2$, sendo qualquer estrutura da matriz de correlação de trabalho. Assim, a derivada de Ω_ω com relação à ω^\top avaliada no vetor com ausência de perturbação é dada por $-\Omega$.

Tomando por base a equação (5.18), a medida de influência local com perturbação na matriz de correlação de trabalho é extraída da matriz a seguir

$$\text{diag}(\mathbf{b}) \Omega^{-1} \Lambda^\top \mathbf{X} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \Lambda \Omega^{-1} \text{diag}(\mathbf{b}), \quad (5.42)$$

de modo a ser avaliada em ω_0 e em $(\hat{\beta}^\top, \hat{\phi})^\top$.

Por fim, a matriz definida acima (5.42) é a mesma em que obtêm-se a medida de influência local para ponderação de casos, apresentada em (5.22). Tendo em vista isso, evidencia-se que a ponderação de casos, além de ser compreendida como uma perturbação na variância de cada observação, a mesma também pode ser vista como uma perturbação na matriz de correlação de trabalho.

5.8 Esquemas de perturbação sob heterogeneidade da dispersão

De acordo com Venezuela (2008), a equação de estimação generalizada para modelagem da média admitindo a suposição de homogeneidade do parâmetro de dispersão para o modelo de regressão beta pode ser expressa por:

$$\Psi_1(\beta) = \mathbf{M}^\top \Lambda \Omega^{-1} \mathbf{b} = \mathbf{M}^\top \mathbf{W} \Lambda^{-1} \mathbf{b}, \quad (5.43)$$

com $\Lambda = \mathbf{G}\mathbf{A}$ e $\mathbf{W} = \Lambda \Omega^{-1} \Lambda$. Além disso, a matriz de sensibilidade, definida em (4.3), é escrita a seguir como:

$$\mathbf{S} = -\mathbf{M}^\top \mathbf{W} \mathbf{M}, \quad (5.44)$$

com isso $\mathbf{M} = (\mathbf{M}_1^\top, \dots, \mathbf{M}_t^\top)^\top$, $\Lambda = \text{diag}(\Lambda_1, \dots, \Lambda_n)$, $\Omega = \text{diag}(\Omega_1, \dots, \Omega_n)$, $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$ e $\mathbf{b} = \text{diag}(\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$. Deve-se ressaltar que a matriz Λ definida para modelos de regressão beta não é uma matriz quadrada, com isso é importante especificar sua matriz transposta quando conveniente.

Diante do exposto, as medidas de influência local, descritas a seguir para alguns esquemas de perturbação com suposição de heterogeneidade da dispersão, com definições apresentadas por meio de (5.43) e (5.44), avaliadas de acordo com as estimativas do modelo pressuposto $\hat{\theta}^\top$ e em ω_0 .

5.8.1 Ponderação de casos

Considerando o seguinte esquema de perturbação

$$\Psi_2(\theta|\omega) = \mathbf{M}^\top \mathbf{W} \Lambda^{-\top} \text{diag}(\omega) \mathbf{b}, \quad (5.45)$$

em que $\omega = (\omega_1^\top, \dots, \omega_t^\top)^\top$, com $\omega_i = (\omega_{i1}, \dots, \omega_{it})^\top$, $i=1, \dots, n$. Nesse caso, ω_0 é um vetor de uns e representa ausência de perturbação.

Diante disso, a matriz Δ , equivalente a derivada $\Psi_2(\theta|\omega)$, definida em (??), com relação à ω^\top é dada por $\Delta = \mathbf{M}^\top \mathbf{W} \Lambda^{-\top} \mathbf{B}$, de modo que para o modelo de regressão beta $\mathbf{B} = \text{diag}(\mathbf{b})$.

Finalmente, a matriz definida em (5.17) é definida por

$$\mathbf{B} \Lambda^{-1} \mathbf{W} \top \mathbf{M} (\mathbf{M}^\top \mathbf{W} \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{M} \Lambda^{-\top} \mathbf{B}, \quad (5.46)$$

que deve ser avaliada em $\hat{\theta}$ e em ω_0 .

5.8.2 Perturbação da variável resposta

Admitindo \mathbf{b}_ω e \mathbf{b} com perturbação na variável resposta de acordo com a estrutura apresentada em (5.23), a equação de estimação perturbada fica escrita como

$$\Psi_2(\boldsymbol{\theta}|\boldsymbol{\omega}) = \mathbf{M}^\top \mathbf{W} \boldsymbol{\Lambda}^{-\top} \mathbf{b}_\omega, \quad (5.47)$$

de modo que $\mathbf{b}_\omega = (\mathbf{b}_{\omega 1}^\top, \dots, \mathbf{b}_{\omega n}^\top)^\top$, com $\mathbf{b}_{\omega i} = (\mathbf{b}_{\omega i 1}, \dots, \mathbf{b}_{\omega i t})^\top$, $i=1, \dots, n$.

Diante disso, em concordância com a equação (5.17), a medida de influencia local com perturbação da variável resposta é obtida por meio da matriz

$$\mathcal{B}^\top \boldsymbol{\Lambda}^{-1} \mathbf{W}^\top \mathbf{M} (\mathbf{M}^\top \mathbf{W} \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{W} \boldsymbol{\Lambda}^\top \mathcal{B}, \quad (5.48)$$

avaliada em $\hat{\boldsymbol{\theta}}$ e em $\boldsymbol{\omega}_0$, com $\mathcal{B} = \partial \mathbf{b}_\omega / \partial \boldsymbol{\omega}^\top$.

No contexto dos modelos de regressão beta, o vetor perturbado na variável resposta referente à i -ésima unidade experimental $\mathbf{b}_{\omega i}$ é dado por meio da expressão (5.26). Com isso, igualmente para o caso de modelagem conjunta da média e da dispersão s_{ij} passa a ser definido como $s_{ij} = \sqrt{\mathbf{v}(\boldsymbol{\mu})_{ij} / (1 + \phi_{ij})}$, com $i=1, \dots, n$ e $j=1, \dots, t$.

6 APLICAÇÃO

O estudo anteriormente aplicado em Botter et.al (2006) e Venezuela (2008) apresenta como objetivo principal a comparação das performances motoras entre recém nascidos Pré-termo, também denominado prematuro, em que o nascimento ocorre até 36 semanas de gestação; e *A termo*, cujo nascimento está compreendido após 36 semanas de gestação. A relevância do estudo está evidenciada na necessidade do acompanhamento e definição do comportamento motor de tais nascidos, pois, em vista das intercorrências neonatais e da ação da gravidade no desenvolvimento da postura, equilíbrio e locomoção, o nascimento prematuro é caracterizado qualitativamente distinto do nascimento a termo (RESTIFFE; 2007).

Diante do exposto, no que se refere ao registro do desenvolvimento dos nascidos, foi considerada a *idade corrigida* (em dias) dos mesmos, ou seja, a idade corrigida corresponde a diferença entre a idade cronológica (equivalente ao tempo de vida desde o nascimento até o momento presente em que se deu o estudo) e o indicador de prematuridade (referente ao número de semanas que faltaram entre o nascimento pré-termo e as 40 semanas de idade gestacional).

A observação dos recém-nascidos se deu de 2 a 13 vezes, como intuito de avaliar o quantitativo de posições (escore) que o indivíduo era capaz de executar e a idade em que ocorreu a execução de cada posição. Com isso, o valor do escore foi registrado podendo apresentar variação entre 0 e 58. Para o estudo, a variável resposta proposta foi definida como o *percentual de de escore atingido com relação à pontuação máxima (58pontos)*². A seguir, na Figura 3, é possível a verificação do comportamento da variável resposta com base na avaliação dos dois grupos.

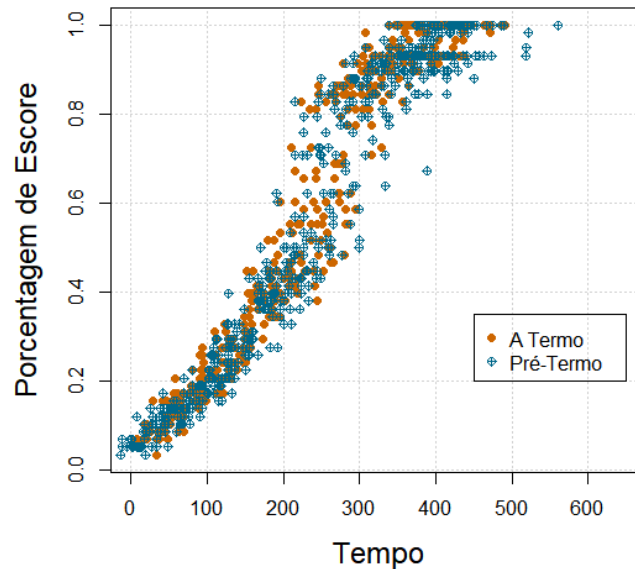
Com isso, a partir da configuração da variável resposta, e com base na estrutura dos dados é ajustado um modelo de regressão beta com medidas repetidas, função de ligação do tipo complemento log-log, estrutura de correlação AR-1 e homogeneidade no parâmetro de dispersão. A estrutura de modelagem da média é dada por

$$\log[-\log(1 - \mu_{ij})] = \beta_0 + \beta_1(x_{1ij}) + \beta_2(x_{2ij}) + \beta_3(x_{1ij} * x_{2ij}), \quad (6.1)$$

de modo que x_{1ij} refere-se ao tipo de nascimento, ou seja, $x_{1ij} = 0$ se “A Termo” ou $x_{1ij} = 1$ se “Pré-Termo” e x_{2ij} representa a idade corrigida em dias. Além do mais, ressalta-se que para

os valores registrados da variável resposta iguais a 0 e 1, considerou-se, respectivamente, iguais a 0,001 e 0,999.

Figura 3 – Gráficos de dispersão do percentual de escore versus a idade corrigida, para os grupos “A termo” e “Pré-Termo”.



Fonte: Elaborado pelo autor

A modelagem da média se deu por meio da utilização do **software R** de modo a permitir uma boa performance na estimação dos parâmetros do modelo, cujo os quais estão apresentados na Tabela 5, bem como a identificação da significância da relação entre as variáveis grupo (x_1) e tempo (x_2) na explicação do aumento médio do escore alcançado, tomando por como base a pontuação máxima com um nível descritivo $p=0.001$.

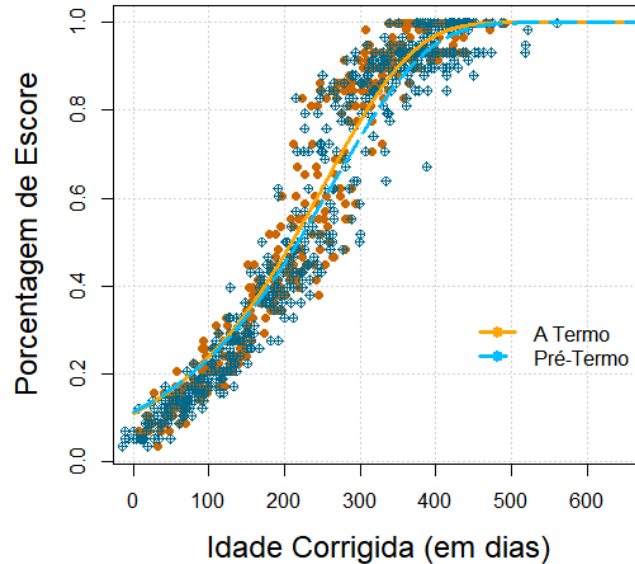
Tabela 5 – Estimativas e erros padrões para os parâmetros da modelagem sob homogeneidade da dispersão

Parâmetro	Estimativa	Erro Padrão		Nível
		Robusto	Naive	Descrivo
β_0 Intercepto	-2,187	0,058	0,050	< 0,001
β_1 x_1	0.065	0,073	0,066	0,373
β_2 x_2	0,009	< 0,001	< 0,001	< 0,001
β_3 $x_1 * x_2$	-0,001	< 0,001	< 0,001	< 0,001
ϕ Precisão	13,923			
α Correlação	0,251			

Fonte: Elaborado pelo autor

Ademais, as curvas ajustadas para o modelo encontra-se a seguir na Figura 4.

Figura 4 – Curvas ajustadas para o modelo de regressão beta sob homogeneidade da dispersão.

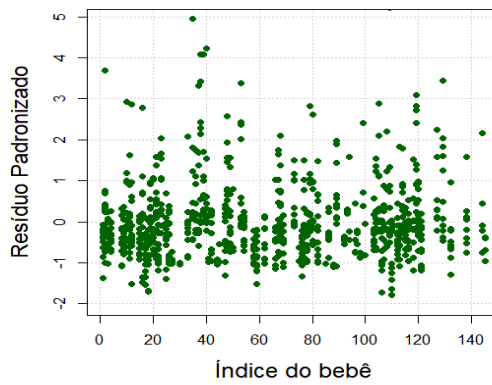
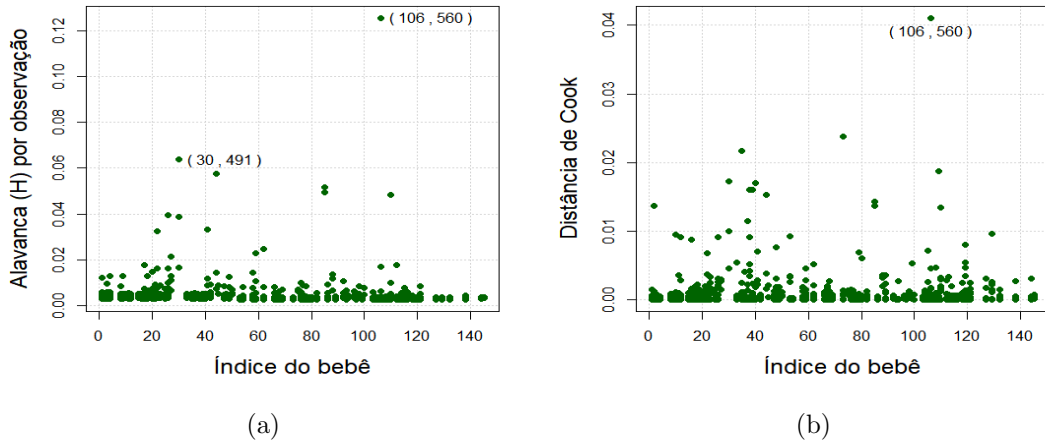


Fonte: Elaborado pelo autor

No que se refere à análise de diagnóstico do modelo supracitado, utilizou-se a medida de alavancagem baseada na matriz de projeção, a distância de Cook e o resíduo padronizado, presentes na Figura 5. A análise gráfica permitiu identificar que a observação atribuída ao recém-nascido 30, cuja idade corrigida foi de 491 dias, (30,491), apresentou-se como um possível ponto de alavanca. Ademais, um outro ponto classificável como alavanca e influente, foi a observação do recém-nascido 106, cuja idade corrigida é foi de 560 dias, (106,560).

Seguindo a análise por meio da aplicação do modelos de influência local, com base nos esquemas de perturbação (ponderação dos casos, variável resposta, covariável da idade corrigida (x_2) e precisão), foi possível evidenciar a existência do indicativo de heterogeneidade do parâmetro de precisão. Tal fato é justificado pela ausência da dispersão homogênea dos valores de d_{max} entre os bebês de índices de 1 a 50. Tais gráficos encontram-se dispostos na Figura 6.

Figura 5 – Gráficos de diagnóstico do modelo de regressão beta sob homogeneidade da dispersão.



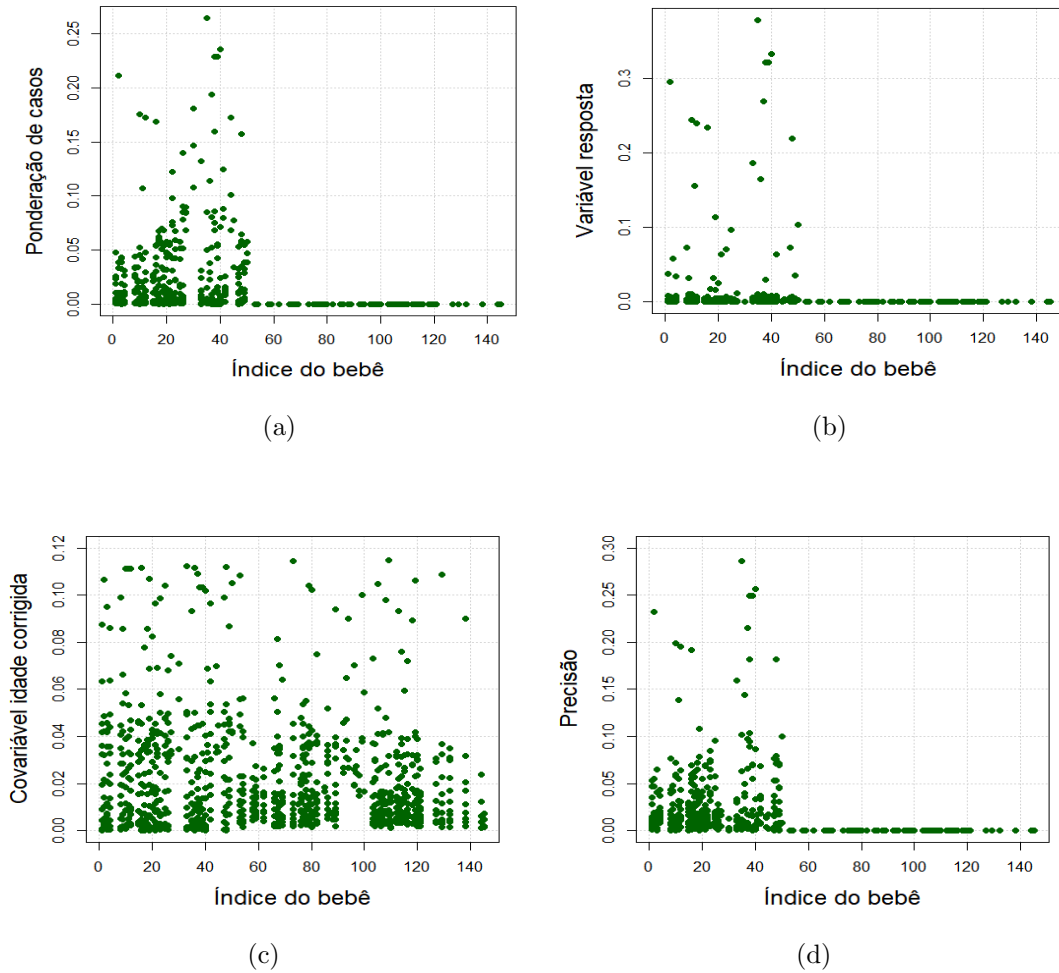
Fonte: Elaborado pelo autor

Diante do exposto, dado a evidência fornecida pela aplicação do modelos de influência local, buscou-se eliminar a possível presença da heterogeneidade da precisa. Para tal, foi realizado uma modelagem conjunta da média sob o modelo apresentado em (6.1) e da precisão do modelo, apresentada a seguir.

$$\log(\phi_{ij}) = \gamma_0 + \gamma_1(x_{1ij}).$$

Dito isto, é possível verificar, por meio da Tabela 6, a partir da estimativa dos parâmetros que as deduções para os parâmetros de β são as mesmas realizadas para a modelagem sob homogeneidade da dispersão anteriormente apresentadas.

Figura 6 – Gráficos de diagnóstico do modelo de regressão beta sob homogeneidade da dispersão.



Fonte: Elaborado pelo autor

Com relação à curva ajustada para o modelo, a mesma encontra-se na Figura 7. Outro aspecto a ser destacado é a redução do resíduo padronizado. Porém, foi identificado a inversão da influência, passando a ser maior no grupo Pré-Termo, bem como o fato de que na nova modelagem ainda se fez presente a atuação de influências distintas, oriundas dos dois grupos, na estimação dos parâmetros regressores.

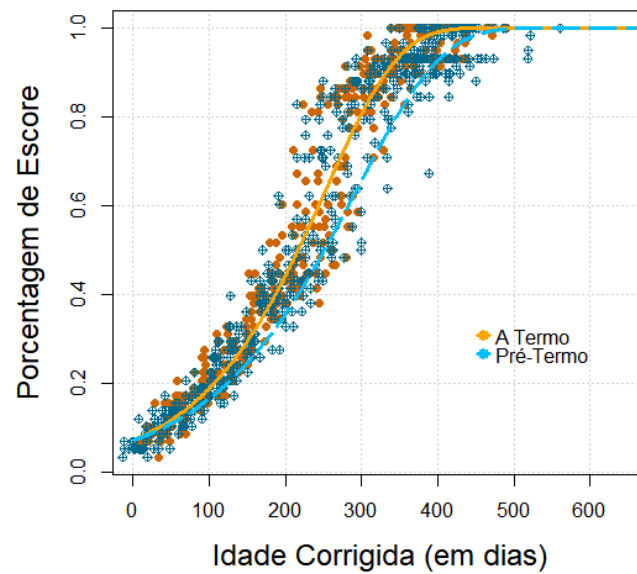
As Figuras 8 e 9 apresentam os gráficos de diagnóstico do ajuste de β e γ para o modelo de regressão beta sob heterogeneidade da dispersão, respectivamente.

Tabela 6 – Estimativas e erros padrões para os parâmetros da modelagem sob heterogeneidade da dispersão

Parâmetro	Estimativa	Erro Padrão		Nível
		Robusto	Naive	Descrivo
β_0 Intercepto	-2,596	0,054	0,034	< 0,001
β_1 x_1	0,029	0,077	0,040	0,704
β_2 x_2	0,001	< 0,001	< 0,001	< 0,001
β_3 x_3	-0,001	< 0,001	< 0,001	< 0,001
γ_0 Intercepto	4,348	0,022	0,007	< 0,001
γ_1 x_1	0.424	0.029	0,008	< 0,001
α Correlação	0,816			

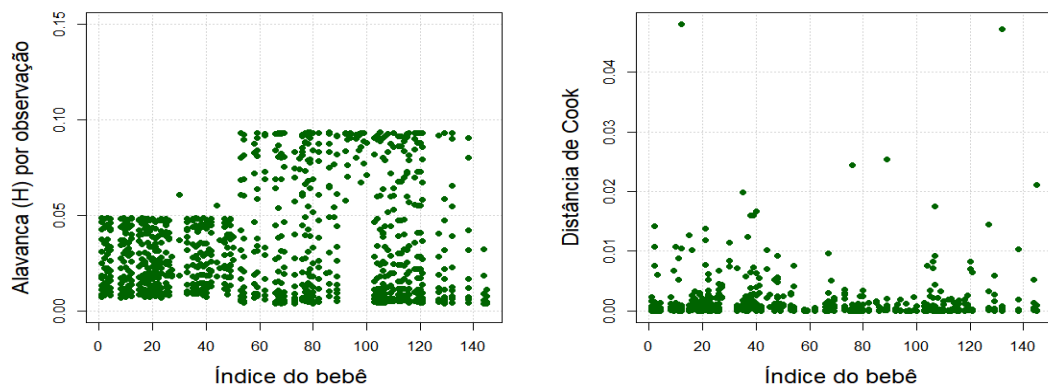
Fonte: Elaborado pelo autor

Figura 7 – Curva ajustadas para o modelo de regressão beta sob heterogeneidade da dispersão.



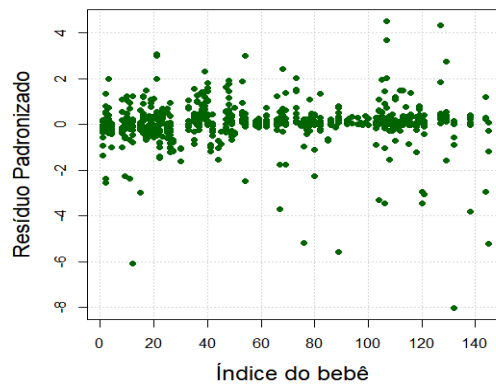
Fonte: Elaborado pelo autor

Figura 8 – Gráficos de diagnóstico do ajuste de β para o modelo de regressão beta sob heterogeneidade da dispersão.



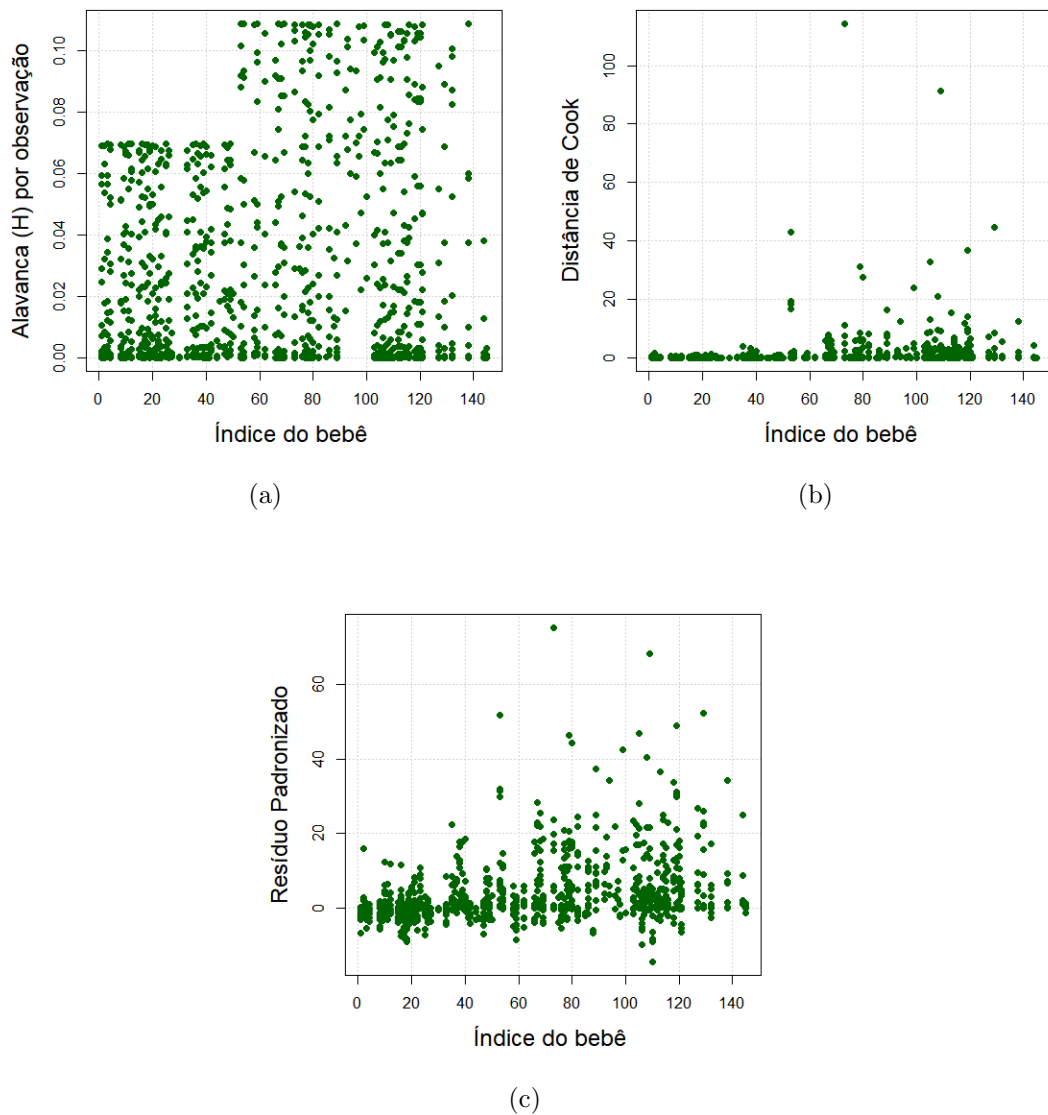
(a)

(b)



(c)

Figura 9 – Gráficos de diagnóstico do ajuste de γ para o modelo de regressão beta sob heterogeneidade da dispersão.



Fonte: Elaborado pelo autor

7 CONSIDERAÇÕES FINAIS

O presente trabalho apresentou a proposta metodológica dos Modelos de Regressão Beta com Medidas Repetidas Via Teoria das Equações de Estimação Generalizadas apresentada por Venezuela (2008), além de apresentar o desenvolvimento das Equações de Estimação Generalizadas desenvolvidas por Liang e Zeger (1986) e dos Modelos de Regressão Beta apresentadas por Ferarri e Cribari-Neto (2004), instrumentos fundamentais na construção da metodologia apresentada.

Além disso, em linha com Venezuela (2008) o estudo também apresenta a construção teórica que dá suporte às técnicas de diagnóstico para dados com medidas repetidas, além do estudo sobre influenciador local para dados com medidas repetidas.

No que se refere ao processo de aplicação do estudo que tomou por base dados reais sobre a performance motora de recém-nascidos prematuros e não-prematuros. Assim, tomando por base a modelagem apresentada no Capítulo anterior, destacam-se como principais achados a boa performance do modelo, a identificação de pontos aberrantes e influentes, além da evidência de que a matriz de correlação de trabalho considerada em tal estudo deveria ser distinta para cada grupo. Tal afirmativa pode ser evidenciada pelo fato de que a utilização da métrica de influência local com esquema de perturbação fornece uma interpretação direta, podendo assim ser interpretada como uma perturbação na matriz de trabalho.

Outro ponto de destaque, corroborando com a teoria, foi a presença da heterogeneidade do parâmetro de precisão pode ser justificada pelo fato de que na base de dados, verificou-se que os indivíduos com índices até 50 são pertencentes ao grupo A Termo (não-prematuro) enquanto que os demais índices são provenientes do grupo Pré-Termo (prematuro).

Diante o exposto, conclui-se que a pesquisa alcançou os objetivos apresentados, tornando possível o estudo, a compreensão dos temas que permeiam a temática abordada e a aplicação dos mesmos.

Por fim, como sugestão de trabalhos futuros, sugere-se o estudo e modelagem das equações de estimação generalizadas para os modelos de regressão simplex, propostas em Song e Tan (2000) e Song et al.(2004) e amplamente discutida em Venezuela (2008).

REFERÊNCIAS

- ALEXANDER G.R.; KOGAN M.; MARTIN J.; PAPIERNIK E. What are the fetal growth patterns of singletons, twins, and triplets in the United States? **Clin Obstet Gynecol**, n. 41, p. 115–25, 1998.
- ARTES, R. **Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão**. São Paulo: IME-USP. Tese de Doutorado, 1997.
- Artes, R.; Botter, D. A. **Funções de Estimação em Modelos de Regressão**. São Paulo: Insper - Instituto de Ensino e Pesquisa - IME – USP. 2005.
- ATKINSON, A. C. **Plots, Transformations and Regressions: An Introduction to Graphical Methods of Diagnostic Regression Analysis**. Oxford: Clarendon Press, 1985. (Oxford Statistical Science Series).
- BIRCH, M.W. Maximum likelihood in three-way contingency tables. **Journal of the Royal Statistical Society**, B52, p. 220- 233, 1963.
- BLISS, C. I. The calculation of the dosage-mortality curve. **Annals of Applied Biology**, v.22, p. 134-167, 1935.
- CADINGAN, N. G. Local influence in structural equation models. **Structural Equation Modeling: A Multidisciplinary Journal** , Routledge, v. 2, n. 1, p. 13-30, 1995.
- CADIGAN, N. G.; FARRELL, P. J. Generalized local influence with applications to fish stok cohort analysis. **Journal of the Royal Statistical Society Series C**, v. 51, p. 469-483, 2002.
- COOK,K.D. Detection of influential observations in linear regressions. **Technometrics**, Taylor Francis, v. 19, n. 1, p. 15-18, 1977
- COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Wiley], v. 48, n. 2, p. 133–169, 1986.

- COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. New York: Chapman and Hall, 1982. (Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/37076>).
- CHANG, Y.-C. Residuals analysis of the generalized linear models for longitudinal data. **Statistics in Medicine**, v.19, p. 77-93, 2000.
- CROWDER, M. On linear and quadratic estimating functions. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 74, n. 3, p. 591–597, 1987.
- DOBSON, A. J.; BARNETT, A. G. **An introduction to generalized linear models**. 3 ed. USA: CRC Press, 2008. 307 p.
- EMERSON, J. D.; HOAGLIN, D. C.; KEMPTHORNE, P. J. Leverage in least squares additive-plus-multiplicative fits for two-way tables. **Journal of the American Statistical Association**, [American Statistical Association, Taylor Francis, Ltd.], v. 79, n. 386, p. 329–335, 1984.
- ESPINHEIRA P.L; FERRARI S.L.P; CRIBARI-NETO F. On beta regression residuals. **Journal of Applied Statistics**, v. 35, n. 4, p. 407–419, 2008.
- EVERITT, B. S.; HORTHON, T. **Statistical analyses using r**. New York: CRC Press, 2010.
- FEIGL, P. E ZELLEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, v. 1, n. 21, p. 826-838, 1965.
- FERRARI, S.L.P.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. **Journal of Applied Statistics** , v.7, n. 31, p. 799-815, 2004.
- FISHER, R.A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society**, n. 222, p. 309-368, 1922.
- FREITAS, J. V. B. **Modelagem de dados com medidas repetidas via equações de estimação generalizadas**. Universidade Federal do Ceará. Monografia (Graduação em Estatística), 2018.
- GLASSER, M. Exponential survival with covariance. **Journal of the American Statistical Association**, n. 62, p. 561-568, 1967.

GODAMBE, V. P. An optimum property of regular maximum likelihood estimation. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 31, n. 4, p. 1208–1211, 1960.

GODAMBE, V. P. (Ed.). **Estimating Functions**. Oxford: Oxford University Press, 1991.

GODAMBE, V.; KALE, B. (Ed.). **Estimating functions: an overview**. Oxford: Estimating Functions. (Ed. V.P. Godambe), Oxford University Press, 1991. 1–20 p.

GODAMBE, V. P. Estimating functions: A synthesis of least squares and maximum likelihood methods. **Lecture Notes-Monograph Series**, Institute of Mathematical Statistics, v. 32, p. 5–15, 1997.

GRAYBILL, F.A. **Theory and Application of the Linear Model**. California: Wadsworth Publishing Company, 1976.

HARDIN, J. W.; HILBE, J. M. **Generalized Estimating Equations**. Chapman and Hall/CRC, 2003.

HARVILLE, D. A. **Matrix Algebra Form a Statistician's Perspective**. New York: Taylor Francis Group, 1997.

HOAGLIN, D. C.; WELSCH, R. E. The hat matrix in regression and anova. **The American Statistician**, Taylor Francis, v. 32, n. 1, p. 17–22, 1978.

JØRGENSEN, B.; LABOURIAU, R.S. **Exponential Families and Theoretical Inference**. Lecture Notes, Department of Statistics, University of British Columbia, 1994.

KIESCHNICK, R.; MCCULLOGH, B.D. Regression analysis of variates observed on (0,1): percentages, proportions and fraction, **Statistical Modelling**, n. 3, p. 193-213, 2003.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 22, n. 1, p. 79–86, 1951.

LEMONTE, A.J. **The Gradient Test: Another Likelihood-Based Test**. Academic Press: London, 2016.

- LIANG, K.-Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 73, n. 1, p. 13–22, 1986.
- LOBATO, D. **Influência Local em Modelos de Regressão**. Campina Grande: UFCG. Dissertação de Mestrado, 2005.
- MCCULLAGH P.; NELDER J. **Generalized linear models**. London: 2nd edn, Chapman and Hall, 2013.
- MCCULLOCH, C. E.; SEARLE, S. R. **Linear and Generalized Linear Mixed Models**. New York, Wiley, 2001.
- MIYASHIRO, E.S. **Modelos de regressão beta e simplex para análise de proporções**. Dissertação de mestrado em Estatística. USP: São Paulo, 2008.
- MYERS, R. H.; MONTGOMERY, D. C.; VINING, G. G.; ROBINSON, T. J. **Generalized Linear Models with Applications in Engineering and the Sciences**. New York: 2 ed. , John Wiley Sons, inc., Hoboken, New Jersey, 2010.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society A**, n. 135, p. 370–84, 1972.
- NETER, J.; KUTNER, M. H.; NASCHSTHEIM, C. J.; WASSERMAN, W. **Applied Linear Statistical Models**. Chicago: [s.n.], 1996.
- OSPINA, P.L.E. **Regressão beta**. Tese de doutorado. IME-USP: São Paulo, 2007.
- PAN, W. Akaike's information criterion in generalized estimating equations. **Biometrics**, , v. 57, n. 1, p. 120–125, 2001.
- PAULA, G.A. **Modelos de regressão com apoio computacional**. São Paulo: Departamento de Estatística. Universidade de São Paulo, 2013.
- POON, W.-Y.; POON, Y. S. Conformal normal curvature and assessment of local influence. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 61, n. 1, p. 51–61, 2002
- PRATER, N.H. Estimate gasoline yields from crude, **Petroleum Refiner** n. 35, p. 236-238, 1956.

PREGIBON, D. Logistic regression diagnostics. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 9, n. 4, p. 705–724, 1981.

PREISSER, J. S.; QAQISH, B. F. Deletion diagnostics for generalised estimating equations. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 83, n. 3, p. 551–562, 1996.

PRENTICE, R. L.; ZHAO, L. P. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. **Biometrics**, [Wiley, International Biometric Society], v. 47, n. 3, p. 825–839, 1991.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <<https://www.r-project.org/>> .

RAO, R. C. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. **Mathematical Proceedings of the Cambridge Philosophical Society**, Cambridge University Press, v. 44, n. 1, p. 50–57, 1948.

RAO, M. N.; RAO, C. R. Linked cross-sectional study for determining norms and growth rates: A pilot survey on indian school-going boys. **Sankhya: The Indian Journal of Statistics, Series B (1960-2002)**, Springer, v. 28, n. 3/4, p. 237–258, 1966.

RESTIFFE, A. P. **O desenvolvimento motor de recém-nascidos pré-termo e a termo até a aquisição da marcha segundo Alberta Infant Motor Scale: um estudo de coorte**. São Paulo: USP. Tese Doutorado, 2007.

ROTNITZKY, A.; JEWELL, N. P. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 77, n. 3, p. 485–497, 1990.

SCHWARZMANN, B. A connection between local-influence analysis and residual diagnostics. **Technometrics**, Taylor Francis, v. 33, n. 1, p. 103–104, 1991.

SEN, P. K.; SINGER, J. M. **Large Sample Methods in Statistics: an introduction with applications**. Boca Raton: [s.n.], 1993.

- SEN, P. K.; SINGER, J. M.; LIMA, A. C. Pedroso de. **From Finite Sample to Asymptotic Methods in Statistics**. Cambridge: Cambridge University Press, 2009. (Cambridge Series in Statistical and Probabilistic Mathematics).
- SIMAS, A.B.; BARRETO-SOUZA, W.; ROCHA, A.V. Improved Estimators for a General Class of Beta Regression Models. **Computational Statistics and Data Analysis**, n. 54, p. 348-366, 2010.
- ST. LAURENT, R. T.; COOK, R. D. Leverage and superleverage in nonlinear regression. **Journal of the American Statistical Association**, n. 87, p. 985-990, 1992.
- TAN, M.; QU, Y.; H.KUTNER, M. Model diagnostics for marginal regression analysis of correlated binary data. **Communications in Statistics - Simulation and Computation**, Taylor Francis, v. 26, n. 2, p. 539–558, 1997.
- THOMAS, W.; COOK, R. D. Assessing influence on regression coefficients in generalized linear models. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 76, n. 4, p. 741–749, 1989.
- TURKMAN, M.A.A.; SILVA, G.L. **Modelos Lineares Generalizados: da teoria à prática**. Lisboa, 2000.
- VENEZUELA, M.K. **Modelos Lineares Generalizados para Análise de Dados com Medidas Repetidas**. São Paulo: IME - USP. Dissertação de Mestrado, 2003.
- VENEZUELA, M. K.; BOTTER, D. A.; SANDOVAL, M. C. Diagnostic techniques in generalized estimating equations. **Journal of Statistical Computation and Simulation**, Taylor Francis, v. 77, n. 10, p. 879–888, 2007.
- VENEZUELA, M. K. **Equação de estimação generalizada e influência local para modelos de regressão beta com medidas repetidas**. São Paulo: IME - USP. Tese de Doutorado, 2008.
- VENEZUELA, M. K.; SANDOVAL, M. C.; BOTTER, D. A. Local influence in estimating equations. **Computational Statistics Data Analysis**, v. 55, n. 4, p. 1867 – 1883, 2011.

WANG, Y. -G; CAREY, V. Working correlation structure misspecification, estimation and covariance desing: Implications for generalised estimating equations performance. **Biometrika**, n.90, p. 29-41, 2003.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. **Transactions of the American Mathematical Society**, American Mathematical Society, v. 54, n. 3, p. 426-482, 1943.

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

WILKS, S.S. (1938) . The Large Distribution of the Likelihood Ratio for Testing Composite Hypotheses. **Annals of Mathematical Statistics**, n. 9, p. 60-62, 1938.

ZEGER, S.L; LIANG, K.-Y. Longitudinal data analysis for discrete and continuous outcomes, **Biometrics**, n. 42, p. 121-130, 1986.

ZIPPIN, C.; ARMITAGE, P. Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. **Biometrics**, n. 22, p. 665-672, 1966.